informs
https://pubsonline.informs.org/journal/stsy

STOCHASTIC SYSTEMS
Vol. 14, No. 3, September 2024, pp. 337–361
ISSN 1946-5238 (online)

Distributionally Robust Observable Strategic Queues

Yijie Wang,[a,*] Madhushini Narayana Prasad,[b] Grani A. Hanasusanto,[c] John J. Hasenbein[b]

[a] School of Economics and Management, Tongji University, Shanghai 200092, China; [b] Graduate Program in Operations Research and Industrial Engineering, The University of Texas at Austin, Austin, Texas 78712; [c] Department of Industrial and Enterprise Systems Engineering, University of Illinois Urbana-Champaign, Urbana, Illinois 61801

*Corresponding author

Contact: yijiewang514@gmail.com, https://orcid.org/0000-0002-5705-892X (YW); madhushini@utexas.edu, https://orcid.org/0000-0001-5259-3395 (MNP); gah@illinois.edu, https://orcid.org/0000-0003-4900-2958 (GAH); has@me.utexas.edu, https://orcid.org/0000-0002-6729-0968 (JJH)

Received: April 14, 2022
Revised: July 30, 2023
Accepted: April 12, 2024
Published Online in Articles in Advance: May 16, 2024

https://doi.org/10.1287/stsy.2022.0009

Abstract. This paper presents an extension of Naor's analysis on the join-or-balk problem in observable $M/M/1$ queues. Although all other Markovian assumptions still hold, we explore this problem assuming uncertain arrival rates under the distributionally robust settings. We first study the problem with the classical moment ambiguity set, where the support, mean, and mean-absolute deviation of the underlying distribution are known. Next, we extend the model to the data-driven setting, where decision makers only have access to a finite set of samples. We develop three optimal joining threshold strategies from the perspectives of an individual customer, a social optimizer, and a revenue maximizer such that their respective worst-case expected benefit rates are maximized. Finally, we compare our findings with Naor's original results and the traditional sample average approximation scheme.

Funding: This research was supported by the National Science Foundation [Grants 2342505 and 2343869].

Keywords: distributionally robust optimization • economic queues • Naor's model • parameter uncertainty

1. Introduction

Imposing tolls to regulate queueing systems was first studied by Naor (1969). He considers a single-server first come, first served (FCFS) queue with stationary Poisson arrivals at a known rate λ. Service times are independent and identically and exponentially distributed with the rate μ. Customers are assumed to be risk neutral and homogenous from an economic perspective. Each customer receives a reward of \$$R$ upon service completion and incurs a cost of \$$C$ per unit of time spent in the system (including in service). In the observable model, every arriving customer inspects the queue length and decides whether to join (reneging is not allowed) or balk (i.e., not join the queue). This strategic decision making is the key factor differentiating this model from the classic $M/M/1$ queueing model.

Naor (1969) derives an optimal threshold strategy n. The customer joins the queue if and only if the system length is less than n. He computes this threshold value under three different control strategies: (1) *individual optimization* (n_e) where the customers act in isolation, aiming to maximize their own expected net benefit rate; (2) *social optimization* (n_s) where the objective is to maximize the long-run rate at which customers accrue net benefit; and (3) *revenue maximization* (n_r) where the agency imposes a toll on the customers joining the queue with the goal of maximizing its own revenue. The most important result by Naor (1969) is the relation $n_r \le n_s \le n_e$, which implies that the customers tend to join the system at a higher rate when left to themselves than is socially optimal. This is because customers do not consider the negative externalities they impose on customers who arrive later. The result also implies that the revenue-maximizing firms allow fewer customers to join their system than the socially optimal case.

Many authors have expanded on the seminal work by Naor (1969); a detailed review of these game-theoretic models is presented in a recent book by Hassin and Haviv (2003). Some of the other recent works (Burnetas and Economou 2007, Economou and Kanta 2008, Guo and Hassin 2011) involve deriving threshold strategies in a classic Naor setting with server shutdowns. Although Economou and Kanta (2008) study the system with server breakdowns and repairs, Burnetas and Economou (2007) analyze the system where the server shuts off when idle and incurs a setup time to resume. A slight variant of this model is given by Guo and Hassin (2011), where the server resumes only when the queue length exceeds a given critical length. Also, Guo and Zipkin (2007) explore the effects of three different levels of delay information and identify the specific cases that do and do not require such information to improve the performance. Haviv and Oz (2016) review the properties of several existing regulation schemes and devise a new mechanism where customers are given priority based on the queue length. Afèche and Ata (2013) study the observable $M/M/1$ queue with heterogenous customers, with some patient and some impatient of given proportions.

All the aforementioned works explore the Naor (1969) model by assuming deterministic arrival or service rates. However, in many real-world scenarios, customers may behave differently during different periods. Hence, there is merit in building a model that performs well under uncertainty. A possible approach is to consider distributional uncertainty on the customers' interarrival times (Bandi et al. 2015). Unfortunately, such granular data are usually inaccessible or simply not stored in practice. In addition, even if the granular data are collected, the problem is still theoretically challenging as the $M/M/1$ structure no longer holds. In this case, calculating the long-run expected social benefit or revenue rates would be difficult, and there is no such study in the context of strategic queues. To avoid these shortcomings, some recent studies propose to address uncertainty by taking the arrival or service rate as a random variable (Liu and Hasenbein 2019, Hassin et al. 2023). For example, they assume that the customer arrival rate is a random variable for each weekday, and the queue manager seeks a strategy that maximizes the long-run social benefits or revenue. This modeling assumption is an expedient approach to account for customer arrival variation while considering the issues of data accessibility and model complexity. Compared with granular data, historical arrival rates are much easier to obtain; most restaurants can collect the historical daily arrival rates by referring to the recorded sales quantity in their accounting books, but only a tiny portion of restaurants keep records of the exact arrival time of each customer. In addition, by assuming that the arrival rate is a fixed random variable within each time slot, the model preserves the $M/M/1$ structure, which enables the use of many elegant results from the paper of Naor (1969).

Papers with related assumptions as in our model include Debo and Veeraraghavan (2014), who consider a system where the arriving customers cannot completely observe the service rate and value. They assume that the server belongs to one of two known types and that the service rate and prior probability for each type are known. Liu and Hasenbein (2019) study a stochastic extension of the Naor (1969) model by relaxing the assumption of a certain arrival rate. They assume that the arrival rate is drawn from a probability distribution that is known to the decision maker. Chen and Hasenbein (2020) further extend the stochastic model to the unobservable setting. They show that the social optimizer induces a lower expected arrival rate than the revenue maximizer in this setting. Hassin et al. (2023) also investigate the unobservable stochastic model from the perspective of strategic customers and demonstrate that the model exhibits a rate-biased arrivals see time averages property. Despite their conceptual appeal, all these works require that the arrival or service rate distribution is known precisely to decision makers, which may not be realistic in practice. In this paper, we extend the classical Naor model for observable systems by relaxing these assumptions, where we assume the arrival rate is uncertain and governed by an unknown underlying distribution, whereas the service rate is deterministic.

To this end, we consider an alternate modeling paradigm called the *distributionally robust optimization* (DRO) (Scarf 1957, Žáčková 1966, Shapiro and Kleywegt 2002). Unlike the traditional stochastic optimization model, DRO acknowledges the lack of full distributional information on the random arrival rate. Instead, the decision maker is assumed to have access to partial information, such as the moments and structural properties of the arrival rate distribution, or some limited historical observations. In this setting, the objective is to derive optimal threshold strategies that maximize the worst-case expected benefit rate, where the worst case is taken over an *ambiguity set* of all distributions consistent with the available information about the true distribution. Such max-min problems have been studied since the seminal work by Scarf (1957), but they have only received more attention with the advent of modern robust optimization techniques (Bertsimas and Sim 2004, Ben-Tal et al. 2009). Since then, a substantial body of literature has been devoted to studying well-known optimization problems under uncertainty in a distributionally robust setting; see Delage and Ye (2010), Li et al. (2014), Wiesemann et al. (2014), Hanasusanto et al. (2015), Shafieezadeh-Abadeh et al. (2015), and Ardestani-Jaafari and Delage (2021). Nevertheless, the distributionally robust framework has not been considered in the context of the classical Naor observable strategic queue model. The paper fills this gap in the literature.

We first study the distributionally robust queue model with a mean-absolute deviation (MAD) ambiguity set (Postek et al. 2018), where partial information about the distribution mean and MAD are known. Next, we extend our model to the data-driven setting, where queue system managers only have access to a finite number of independent and identically distributed training samples collected from historical observations. We construct a data-driven mean-absolute deviation (DD-MAD) ambiguity set that mitigates estimation errors from the empirical moment estimators. The resulting distributionally robust model with a data-driven ambiguity set admits a semidefinite programming (SDP) reformulation for the social optimization problem and a linear programming reformulation for the revenue maximization problem. To properly determine the robustness parameters, we establish a new distribution-free confidence interval for the empirical MAD. Although such confidence intervals exist for the empirical mean and variance (Delage and Ye 2010), to the best of our knowledge, none are available for the empirical MAD. Herrey (1965) derives the confidence interval for the empirical MAD under normal distribution data, whereas other works mostly focus on median-absolute deviation; see Bonett and Seier (2003), Abu-Shawiesh et al. (2018), and Arachchige and Prendergast (2019). Using this result, we further derive finite-sample guarantees for the data-driven MAD model, in which optimal values provide high-confidence lower bounds on the expected social benefit or revenue rate. We also benchmark our data-driven MAD ambiguity set with the popular Wasserstein ambiguity set (Pflug and Wozabal 2007, Esfahani and Kuhn 2018, Esfahani et al. 2018, Gao and Kleywegt 2023), which is widely used in the data-driven setting as it can offer attractive finite-sample guarantees. The results demonstrate that our proposed data-driven MAD model shares a similar guarantee as the Wasserstein model while generating a significantly more tractable reformulation.

The main contributions of this paper can be summarized as follows.

1. We propose a new model to tackle the uncertain arrival rate in the Naor strategic queue problem using the emerging DRO framework. The model does not impose any specific distributional assumption; instead, it optimizes in view of the worst-case distribution within a prescribed ambiguity set. Benefitting from this robustification framework, the model alleviates the overfitting issue and yields attractive out-of-sample performance.

2. We prove that the revenue rate function is concave, whereas the social benefit rate function is either concave or unimodal under some mild prerequisites. We then show that these properties enable a closed-form solution for the worst-case expectation problem with an MAD ambiguity set. For the general cases, we derive an SDP reformulation for the social optimization problem and a linear programming reformulation for the revenue optimization problem.

3. We extend the distributionally robust model to the data-driven setting, where queue system managers only have access to a finite set of historical observations. To mitigate the adverse effect of the estimation errors from the empirical MAD, we robustify the ambiguity set by adding an extra layer of robustness to the empirical mean and MAD estimators. The data-driven MAD model admits an SDP reformulation for the social optimization problem and a linear programming reformulation for the revenue maximization problem. We then establish a distribution-free confidence interval for the empirical MAD and derive finite-sample guarantees for the distributionally robust model with a data-driven MAD ambiguity set. Compared with the Wasserstein ambiguity set, the data-driven MAD ambiguity set admits a more efficient reformulation of fixed complexity, where the number of constraints does not scale with the sample size.

The remainder of the paper is structured as follows. In Section 2, we propose the distributionally robust queue model and analyze the relationship between different thresholds under the distributionally robust setting. Section 3 presents tractable reformulations for the worst-case expectation problem with a classical MAD ambiguity set. Section 4 explores the distributionally robust model with a data-driven MAD ambiguity set and derives theoretical finite-sample guarantees. Finally, the out-of-sample performances of our distributionally robust models are assessed empirically in Section 5.

1.1. Notations

The set of all probability measures supported on Ξ is written as $\mathcal{P}_0(\Xi) := \{\mu \in \mathcal{M}_+ : \int_\Xi \mu(d\xi) = 1\}$, where $\mathcal{M}_+$ denotes the set of nonnegative Borel measures. All random variables are designated by tilde signs (e.g., $\tilde{\rho}$), whereas their realizations are denoted without tildes (e.g., ρ). We denote by $\mathbb{E}_\mathbb{P}[c(\tilde{\rho})]$ the expectation of a cost function with respect to the random variable $\tilde{\rho}$ under distribution $\mathbb{P}$. We define $\lfloor n \rfloor$ to be the largest integer less than or equal to n and $\|\boldsymbol{x}\|_p$ to be the p-norm of a vector $\boldsymbol{x}$. For any set Ξ, we let $\text{int}(\Xi)$ denote its interior. The cone of $k \times k$ positive semidefinite matrices is denoted by $\mathbb{S}_+^k$.

2. Distributionally Robust Strategic Queues Model

The extension of the Naor (1969) seminal queue model to the stochastic optimization setting with an uncertain arrival rate was first proposed by Liu and Hasenbein (2019), who consider an $M/M/1$ queue system with a random arrival rate $\tilde{\lambda} \sim \mathbb{P}^\star$ and a deterministic service rate μ. The queue system operates under a first come, first served discipline, and the true distribution of the uncertain arrival rate $\tilde{\lambda}$ is known by the system manager.

Because the service rate μ is deterministic, without loss of generality, we consider the traffic intensity $\tilde{\rho} := \frac{\tilde{\lambda}}{\mu}$ as the uncertain parameter throughout the remainder of the paper. The stochastic model aims to find an optimal threshold that maximizes the expected benefit rate: that is,

$$\max_{n \in \mathbb{Z}_+} \mathbb{E}_{\mathbb{P}^\star}[c_n(\tilde{\rho})].$$

Here, $c_n(\tilde{\rho})$ is a general return function, which can be replaced with the social benefit rate function or the revenue rate function depending on the system manager's objective.

In practice, the true distribution $\mathbb{P}^\star$ is never available to the system manager and typically has to be estimated using the empirical distribution generated from the historical observations. Although the empirical-based methods may work well on the observed data set, they often fail to achieve an acceptable out-of-sample performance because they do not consider any possible disturbances from the limited historical observations.

In this paper, we endeavor to address this fundamental shortcoming using ideas of DRO. The DRO approach does not impose any single distribution on the uncertain arrival rate. Instead, it constructs an ambiguity set $\mathcal{P}$ containing all plausible probability distributions that are consistent with the partial information as well as historical observations. In this setting, the objective is to derive an optimal threshold strategy $\hat{n}$ that maximizes the worst-case expected benefit rate, where the worst case is taken over all distributions from within this ambiguity set: that is,

$$\max_{n \in \mathbb{Z}_+} \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[c_n(\tilde{\rho})]. \tag{1}$$

Because the model optimizes the expected benefit rate in view of the worst-case distribution, it mitigates overfitting to the observed samples and helps improve the performance in out-of-sample circumstances.

In this paper, we study the distributionally robust model from the perspective of an individual customer, a social optimizer, and a revenue maximizer. We first derive the results that hold for any generic ambiguity set $\mathcal{P}$.

2.1. Individual Optimization

We determine a pure threshold strategy in which each arriving customer decides to join or not join the queue based on the observed queue length, independent of the strategy adopted by other customers. A newly arrived customer makes a decision (to join or not join) based on the net gain $R - (i+1)C/\mu$, where i is the number of people currently in the queue, and will join the queue if it is nonnegative. Note that net gain is deterministic because it is independent of the random arrival rate. Thus, the optimal joining threshold for any arriving customer is given by

$$\hat{n}_e = \left\lfloor \frac{R\mu}{C} \right\rfloor. \tag{2}$$

This result coincides with the original result of Naor (1969) (i.e., $\hat{n}_e = n_e$) because the net gain of a newly arrived customer only depends on the current queue length and the service rate, which are all deterministic. On the other hand, as an individual optimizer, the customer can ignore the rates of future arrivals because they will not affect the time to service.

2.2. Social Optimization

We next analyze the distributionally robust threshold for a social optimizer. The social benefit rate for a realization of the traffic intensity ρ and a fixed threshold n is given by

$$f_n(\rho) := \begin{cases} R\mu \dfrac{\rho(1-\rho^n)}{1-\rho^{n+1}} - C\left(\dfrac{\rho}{1-\rho} - \dfrac{(n+1)\rho^{n+1}}{1-\rho^{n+1}}\right) & \text{if } \rho \neq 1 \\ R\mu \dfrac{n}{n+1} - C\dfrac{n}{2} & \text{if } \rho = 1. \end{cases} \tag{3}$$

One can verify that $\lim_{\rho \to 1} R\mu \frac{\rho(1-\rho^n)}{1-\rho^{n+1}} - C\left(\frac{\rho}{1-\rho} - \frac{(n+1)\rho^{n+1}}{1-\rho^{n+1}}\right) = R\mu \frac{n}{n+1} - C\frac{n}{2}$, which indicates that the function $f_n(\rho)$ is continuous in ρ. The distributionally robust model determines an optimal threshold $\hat{n}_s$ that maximizes the worst-case expected social benefit rate $Z_s(n)$: that is, $\hat{n}_s \in \arg\max_{n \in \mathbb{Z}_+} Z_s(n)$, where

$$Z_s(n) := \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[f_n(\tilde{\rho})]. \tag{4}$$

We first investigate the relationship between the optimal thresholds $\hat{n}_e$ and $\hat{n}_s$.

Proposition 1. *There exists an optimal threshold of the social optimizer less than or equal to the optimal threshold of an individual customer: that is,* $\exists \hat{n}_s \, s.t. \, \hat{n}_s \leq \hat{n}_e$.

Proposition 1 enables decision makers to search for the best threshold from $\{1, \ldots, \hat{n}_e\}$. We remark that the participation of customers is not affected by the distributionally robust setting because the queue adopts the FCFS discipline, so subsequent arrivals are immaterial after the customer has joined the queue. Thus, the socially optimal threshold is always achievable by reducing the individual threshold from n_e to $\hat{n}_s$.

2.3. Revenue Optimization

We now consider a profit-maximizing firm that aims to maximize its expected revenue rate by imposing a toll t on every joining customer. In this setting, customers base their joining decision on this imposed toll t and evaluate the service completion only by $R - t$. Recall that customers join the queue if and only if the expected net gain is nonnegative. Therefore, determining an optimal toll t is equivalent to choosing a queue-length threshold n that maximizes the expected revenue rate, where $n = \lfloor \frac{(R-t)\mu}{C} \rfloor$.

The revenue rate for a realization of the traffic intensity and a fixed threshold n is given by

$$r_n(\rho) := \begin{cases} (R\mu - Cn)\dfrac{\rho(1-\rho^n)}{1-\rho^{n+1}} & \text{if } \rho \neq 1 \\[2ex] (R\mu - Cn)\dfrac{n}{n+1} & \text{if } \rho = 1. \end{cases} \tag{5}$$

One can show that $\lim_{\rho \to 1} \frac{\rho(1-\rho^n)}{1-\rho^{n+1}} = \frac{n}{n+1}$, which indicates that $f_n(\rho)$ is continuous. The distributionally robust model determines an optimal threshold $\hat{n}_r$ that maximizes the worst-case expected revenue rate $Z_r(n)$: that is, $\hat{n}_r \in \arg\max_{n \in \mathbb{Z}_+} Z_r(n)$, where

$$Z_r(n) := \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[r_n(\tilde{\rho})]. \tag{6}$$

Similarly, we first investigate the relationship between the optimal thresholds $\hat{n}_e$ and $\hat{n}_r$.

Proposition 2. *There exists an optimal threshold of the revenue maximizer less than or equal to the optimal threshold of an individual customer: that is,* $\exists \hat{n}_r \, s.t. \, \hat{n}_r \leq \hat{n}_e$.

So far, we have presented the generic distributionally robust observable queue models for an individual customer, a social optimizer, and a revenue maximizer. However, we have not specified the ambiguity set for the social and revenue optimization problems. In the following sections, we will investigate different types of ambiguity sets and derive their tractable reformulations.

3. Distributionally Robust Strategic Queues with an MAD Ambiguity Set

We study the DRO model with an MAD ambiguity set. Suppose the support $[a, b]$, mean m, and MAD d of the random parameter $\tilde{\rho}$ are known to the decision makers. Then, we can construct an ambiguity set containing all possible distributions that are consistent with the partial information, defined as

$$\mathcal{P} := \{\mathbb{P} \in \mathcal{P}_0([a, b]) : \mathbb{E}_{\mathbb{P}}[\tilde{\rho}] = m, \mathbb{E}_{\mathbb{P}}[|\tilde{\rho} - m|] = d\}. \tag{7}$$

We develop efficient solution schemes to find the optimal threshold strategies for a social optimizer and a revenue maximizer, given by $\hat{n}_s$ and $\hat{n}_r$, respectively, such that the worst-case expected benefit rates are maximized. In order to derive tractable reformulations for the distributionally robust models, we assume $m \in (a, b)$ and $d \in (0, \overline{d})$, where $\overline{d} := \frac{2(m-a)(b-m)}{b-a}$ is the largest possible mean-absolute deviation attained by any distribution with the given support and mean.

3.1. Social Optimization

To determine an optimal joining threshold for a social optimizer, we compute the worst-case expected social benefit rate $Z_s(n)$ for every $n \in \mathbb{Z}_+$ satisfying $1 \leq n \leq n_e$ and choose an $\hat{n}_s$ such that $\hat{n}_s \in \arg\max_{n \in \mathbb{Z}_+} Z_s(n)$. To this end, we show how to compute the worst-case expected social benefit rate for a fixed n. Suppose the distribution mean and MAD of $\tilde{\rho}$ are precisely known; then, the worst-case expected social benefit rate is given by the optimal value of the moment problem

$$\begin{aligned} Z_s(n) = \inf_{\nu \in \mathcal{M}_+} \quad & \int_\Xi f_n(\rho)\nu(\mathrm{d}\rho) \\ \text{s.t.} \quad & \int_\Xi |\rho - m|\, \nu(\mathrm{d}\rho) = d \\ & \int_\Xi \rho\, \nu(\mathrm{d}\rho) = m \\ & \int_\Xi \nu(\mathrm{d}\rho) = 1, \end{aligned} \tag{8}$$

where $\Xi := [a,b]$ is the support of $\tilde{\rho}$. The third constraint of (8) restricts the nonnegative measure ν to be a probability distribution, whereas the first and second constraints impose that the distribution's MAD and mean are equal to d and m, respectively. The objective of the problem is to find a feasible distribution that minimizes the expected social benefit rate.

The semi-infinite linear optimization Problem (8) is hard to solve because it searches for the best decision from an infinite-dimensional space of probability measures. To derive a tractable reformulation, we focus on the dual problem. We first define $F(\rho) := \alpha|\rho - m| + \beta\rho + \gamma$ and derive the dual problem as

$$\begin{aligned} \sup_{\alpha,\beta,\gamma \in \mathbb{R}} \quad & \alpha d + \beta m + \gamma \\ \text{s.t.} \quad & F(\rho) \le f_n(\rho) \qquad \forall \rho \in [a,b]. \end{aligned} \tag{9}$$

Notice that $F(\rho)$ is a two-piece piecewise affine function majorized by $f_n(\rho)$. We know that if $f_n(\rho)$ is a piecewise affine function or a concave function, the semi-infinite constraint will reduce to a linear constraint because we only need to check the satisfaction of the constraint at points $\rho = a, m$, and b. However, the social benefit rate function is neither concave nor piecewise affine, making the problem difficult. To solve this optimization problem, we first investigate the properties of the social benefit rate function $f_n(\rho)$. For clarity of exposition, we relegate some of the proofs to Appendix B.

Lemma 1. *The social benefit rate function $f_n(\rho)$ has the following properties if $\frac{R\mu}{C} \ge n+1$.*
1. *$f_n(\rho)$ is strictly concave for $\rho \in [0,1]$.*
2. *$f_n(\rho)$ is either concave increasing or unimodal for $\rho \in [0,\infty)$.*
3. *The sign of the second derivative $f_n''(\rho)$ changes at most once over $[0,\infty)$.*

From Lemma 1, we know that the social benefit rate function has some appealing properties. Specifically, the function is either concave increasing or unimodal on the nonnegative axis, and when it is unimodal, the function changes from a concave function to a convex function at some point. The next lemma further asserts that the complementary slackness property holds for the primal and dual problems, which will later help us determine the worst-case distribution.

Lemma 2. *The optimal values of the primal-dual pair (8) and (9) coincide, and their optimal solutions $\nu^\star$ and $(\alpha^\star, \beta^\star, \gamma^\star)$, respectively, satisfy the complementary slackness condition*

$$(f_n(\rho) - \alpha^\star|\rho - m| - \beta^\star\rho - \gamma^\star)\nu^\star(\mathrm{d}\rho) = 0 \qquad \forall \rho \in [a,b].$$

The proofs of the lemmas are relegated to Appendix B. Combining Lemmas 1 and 2, we are ready to show that Problem (8) can be solved analytically under certain conditions. Specifically, we divide this problem into three cases and derive an explicit expression of the worst-case distribution for each case.

Proposition 3. *Assume $m \in [0,1]$ and $\frac{R\mu}{C} \ge n+1$. Let $(\rho_t, f_n(\rho_t))$ be the tangent point on f_n for the line that passes through $(m, f_n(m))$. For any $n \ge 1$, we have one of the following three cases.*

1. *If $f_n(b) + f_n'(b)(m-b) \ge f_n(m)$, then the extremal distribution that solves (4) is a three-point distribution supported on $\rho_1 = a$, $\rho_2 = m$, $\rho_3 = b$, with corresponding probabilities*

$$p_1 = \frac{d}{2(m-a)},\ p_2 = 1 - \frac{d}{2(m-a)} - \frac{d}{2(b-m)},\ p_3 = \frac{d}{2(b-m)}.$$

2. *If* $f_n(b) + f_n'(b)(m-b) < f_n(m)$ *and* $d < d_0 := \frac{2(m-a)(\rho_t - m)}{\rho_t - a}$, *then the extremal distribution is a three-point distribution supported on* $\rho_1 = a$, $\rho_2 = m$, $\rho_3 = \rho_t$, *with probabilities*

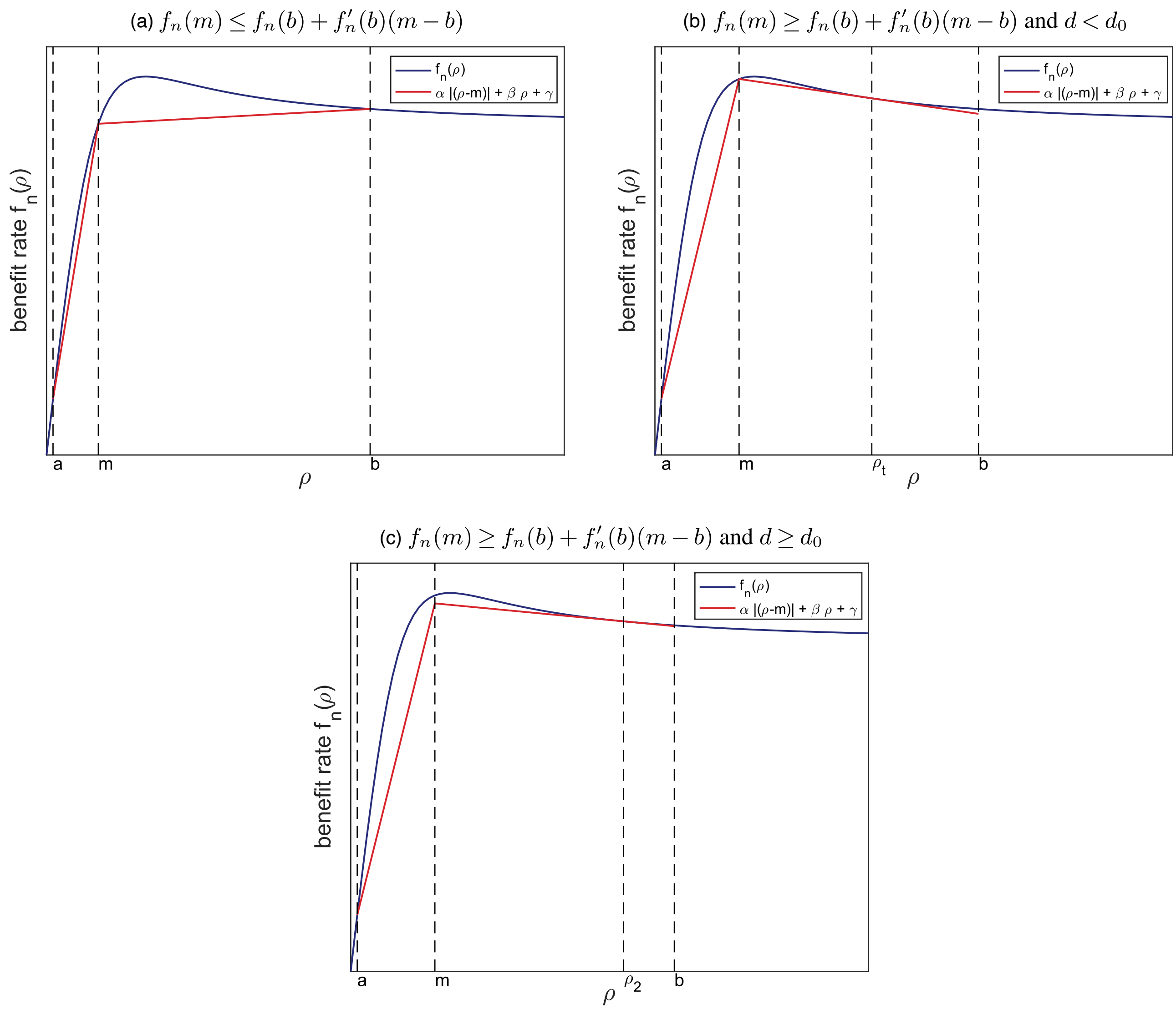
$$p_1 = \frac{d}{2(m-a)},\ p_2 = 1 - \frac{d}{2(m-a)} - \frac{d}{2(\rho_t - m)},\ p_3 = \frac{d}{2(\rho_t - m)}.$$

3. *If* $f_n(b) + f_n'(b)(m-b) < f_n(m)$ *and* $d \geq d_0 := \frac{2(m-a)(\rho_t - m)}{\rho_t - a}$, *then the extremal distribution is a two-point distribution supported on* $\rho_1 = a$, $\rho_2 = \frac{ad + 2m(a-m)}{d + 2(a-m)}$, *with probabilities*

$$p_1 = \frac{d}{2(m-a)},\ p_2 = 1 - \frac{d}{2(m-a)}.$$

Figure 1 depicts the optimal two-piece piecewise affine function described in Proposition 3. We remark that the tangent point $(\rho_t, f_n(\rho_t))$ in Figure 1(b) can be determined efficiently by the bisection method. Specifically, we set $[l, u] = [m, b]$ as the initial search interval for the algorithm. In each iteration, we compute the derivative at the

Figure 1. Visualization of the Three Cases



Notes. In panel (a), the optimal piecewise affine function is determined by points $(a, f_n(a))$, $(m, f_n(m))$, and $(b, f_n(b))$. In panel (b), the parameters satisfy $f_n(m) \geq f_n(b) + f_n'(b)(m-b)$ and $d < d_0$. Thus, the optimal two-piece piecewise affine function touches $f_n(\rho)$ at $(a, f_n(a))$, $(m, f_n(m))$, and $(\rho_t, f_n(\rho_t))$, where $(\rho_t, f_n(\rho_t)$ is the tangent point. In panel (c), $f_n(m) \geq f_n(b) + f_n'(b)(m-b)$ still holds, whereas $d \geq d_0$. In this case, the extremal distribution degenerates to a two-point distribution. (a) $f_n(m) \leq f_n(b) + f_n'(b)(m-b)$. (b) $f_n(m) \geq f_n(b) + f_n'(b)(m-b)$ and $d < d_0$. (c) $f_n(m) \geq f_n(b) + f_n'(b)(m-b)$ and $d \geq d_0$.

midpoint $\rho = \frac{u+l}{2}$, and we check whether it is the tangent point by calculating the difference between $f_n(m)$ and $f_n'\left(\frac{u+l}{2}\right)\left(m - \frac{u+l}{2}\right) + f_n\left(\frac{u+l}{2}\right)$. If the difference is small enough, we terminate the algorithm; otherwise, we set $u = \frac{u+l}{2}$ if the difference is positive or set $l = \frac{u+l}{2}$ if the difference is negative, and then, we go back to the first step with the updated interval $[l, u]$.

We remark that the use of the MAD ambiguity set and its geometric interpretation is motivated by a recent work by van Eekelen et al. (2022), who analyze the worst-case performance of the $GI/G/1$ queue (Bhat 2008) under mean-dispersion constraints. The authors demonstrate that measuring the dispersion by MAD, instead of variance, significantly simplifies the analysis and enables a closed-form solution for the extremal distribution whenever the loss function is convex. Unfortunately, our problem is different as the social benefit rate function is neither convex nor concave. Nevertheless, by establishing some useful properties of the social benefit rate function and exploiting its geometric interpretation in the dual Problem (9), we are able to explicitly express the extremal distribution when $m \le 1$ and $\frac{R\mu}{C} \ge n+1$. Using this result, we can compute the worst-case expected social benefit rate $Z_s(n)$ efficiently.

Theorem 1. *Assume $m \in [0,1]$ and $\frac{R\mu}{C} \ge n+1$. Let $(\rho_t, f_n(\rho_t))$ be the tangent point on $f_n(\rho)$ for the line that passes through $(m, f_n(m))$. For any $n \ge 1$, we have the following three cases.*

1. *If $f_n(b) + f_n'(b)(m-b) \ge f_n(m)$, then*

$$Z_s(n) = \frac{d}{2(m-a)} f_n(a) + \left(1 - \frac{d}{2(m-a)} - \frac{d}{2(b-m)}\right) f_n(m) + \frac{d}{2(b-m)} f_n(b).$$

2. *If $f_n(b) + f_n'(b)(m-b) < f_n(m)$ and $d < d_0 := \frac{2(m-a)(\rho_t - m)}{\rho_t - a}$, then*

$$Z_s(n) = \frac{d}{2(m-a)} f_n(a) + \left(1 - \frac{d}{2(m-a)} - \frac{d}{2(\rho_t - m)}\right) f_n(m) + \frac{d}{2(\rho_t - m)} f_n(\rho_t).$$

3. *If $f_n(b) + f_n'(b)(m-b) < f_n(m)$ and $d \ge d_0 := \frac{2(m-a)(\rho_t - m)}{\rho_t - a}$, then*

$$Z_s(n) = \frac{d}{2(m-a)} f_n(a) + \left(1 - \frac{d}{2(m-a)}\right) f_n\left(\frac{ad + 2m(a-m)}{d + 2(a-m)}\right).$$

Theorem 1 enables us to solve the worst-case expectation problem analytically under certain conditions. However, for the more general case, we are unable to solve it in a closed form. In the following theorem, we show that the worst-case expectation problem admits a semidefinite programming reformulation that can be solved in polynomial time using standard off-the-shelf solvers, such as SDPT3 (Toh et al. 1999) and MOSEK (ApS 2022).

Theorem 2. *For any $n \ge 1$, the worst-case expected social benefit rate $Z_s(n)$ coincides with the optimal value of the following semidefinite program*:

$$\begin{array}{lll}
\sup & \alpha d + \beta m + \gamma & \\
\text{s.t.} & \alpha, \beta, \gamma \in \mathbb{R}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^{n+3}, \boldsymbol{X}, \boldsymbol{X}' \in \mathbb{S}_+^{n+3} & \\
& y_1 = R\mu - C - y_0 + y_{n+3},\ y_2 = -R\mu - y_{n+3}, & \\
& y_3, \ldots, y_n = 0,\ y_{n+1} = -R\mu + C(n+1) - y_0, & \\
& y_{n+2} = R\mu - Cn + y_0 - y_{n+3}, & \\
& y_0 = \alpha m + \gamma, y_{n+3} = -\alpha + \beta & \\
& \displaystyle\sum_{i+j=2l-1} x_{ij} = 0 & \forall l \in [n+4] \\
& \displaystyle\sum_{i+j=2l} x_{ij} = \sum_{q=0}^{l} \sum_{r=q}^{n+3+q-l} y_r \binom{r}{q} \binom{n+3-r}{l-q} a^{r-q} m^q & \forall l \in [n+4] \cup \{0\} \\
& z_1 = R\mu - C - z_0 + z_{n+3},\ z_2 = -R\mu - z_{n+3}, & \\
& z_3, \ldots, z_n = 0,\ z_{n+1} = -R\mu + C(n+1) - z_0, & \\
& z_{n+2} = R\mu - Cn + z_0 - z_{n+3} & \\
& z_0 = -\alpha m + \gamma, z_{n+3} = \alpha + \beta & \\
& \displaystyle\sum_{i+j=2l-1} x'_{ij} = 0 & \forall l \in [n+4] \\
& \displaystyle\sum_{i+j=2l} x'_{ij} = \sum_{q=0}^{l} \sum_{r=q}^{n+3+q-l} y'_r \binom{r}{q} \binom{n+3-r}{l-q} m^{r-q} b^q & \forall l \in [n+4] \cup \{0\}.
\end{array} \tag{10}$$

The proof of this theorem relies on the following lemma, which expresses a univariate polynomial inequality in terms of semidefinite constraints.

Lemma 3 (Bertsimas and Popescu 2005, Proposition 3.1(f)). *The polynomial $g(\rho) = \sum_{r=0}^{k} y_r \rho^r$ satisfies $g(\rho) \ge 0$ for all $\rho \in [a, b]$ if and only if there exists a positive semidefinite matrix $X = [x_{ij}]_{i,j=0,\dots,k} \in \mathbb{S}_+^{k+1}$ such that*

$$0 = \sum_{i,j:i+j=2l-1} x_{ij} \qquad \forall l = 1,\dots,k$$

$$\sum_{q=0}^{l} \sum_{r=q}^{k+q-l} y_r \binom{r}{q} \binom{k-r}{l-q} a^{r-q} b^q = \sum_{i,j:i+j=2l} x_{ij} \qquad \forall l = 0,\dots,k.$$

Proof of Theorem 2. Recall that the dual of $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[f_n(\tilde{\rho})]$ for $\tilde{\rho}$ supported on the interval $[a, b]$ is given by (see Problem (9))

$$\begin{aligned} \sup_{\alpha,\beta,\gamma\in\mathbb{R}} \quad & \alpha d + \beta m + \gamma \\ \text{s.t.} \quad & \alpha|\rho - m| + \beta\rho + \gamma \le f_n(\rho) \qquad \forall \rho \in [a, b]. \end{aligned}$$

We can deal with the semi-infinite constraint separately for the cases $\rho \le m$ and $\rho \ge m$:

$$\begin{aligned} \sup_{\alpha,\beta,\gamma\in\mathbb{R}} \quad & \alpha d + \beta m + \gamma \\ \text{s.t.} \quad & \alpha(m - \rho) + \beta\rho + \gamma \le f_n(\rho) \qquad \forall \rho \in [a, m] \\ & \alpha(\rho - m) + \beta\rho + \gamma \le f_n(\rho) \qquad \forall \rho \in [m, b]. \end{aligned}$$

Substituting the definition of $f_n(\rho)$ in (3) and applying algebraic reductions yield the following polynomial inequalities:

$$\begin{aligned} &-(\alpha m + \gamma)\rho^0 + (R\mu - C - \beta + \gamma + \alpha m + \alpha)\rho + (-R\mu - \alpha + \beta)\rho^2 + (-R\mu + Cn + C + \alpha m + \gamma)\rho^{n+1} \\ &\quad + (R\mu - Cn - \alpha m - \alpha + \beta - \gamma)\rho^{n+2} + (\alpha - \beta)\rho^{n+3} \ge 0 \qquad \forall \rho \in [a, m], \quad \text{and} \\ &(\alpha m - \gamma)\rho^0 + (R\mu - C - \alpha m - \alpha - \beta + \gamma)\rho + (-R\mu + \alpha + \beta)\rho^2 + (-R\mu + Cn + C - \alpha m + \gamma)\rho^{n+1} \\ &\quad + (R\mu - Cn + \alpha m + \alpha + \beta - \gamma)\rho^{n+2} - (\alpha + \beta)\rho^{n+3} \ge 0 \qquad \forall \rho \in [m, b]. \end{aligned} \tag{11}$$

The inequalities are of the form $g_1(\rho) = \sum_{r=0}^{n+3} y_r \rho^r \ge 0$ for $\rho \in [a, m]$ and $g_2(\rho) = \sum_{r=0}^{n+3} z_r \rho^r \ge 0$ for $\rho \in [m, b]$, where $y = (y_1, \dots, y_{n+3})$ and $z = (z_1, \dots, z_{n+3})$ represent the coefficients of the respective polynomial inequalities. We now invoke the result of Lemma 3 with $k = n + 3$ to express the inequalities in (11) as semidefinite constraints. The resulting semidefinite problem is equivalent to the original problem, which completes the proof. □

Remark 1. In this subsection, we present two results. Theorem 1 provides a closed-form solution under certain prerequisites, whereas Theorem 2 derives an SDP reformulation for the general cases. It is worth noting that Theorem 1 requires the parameters to satisfy $n \le \frac{R\mu}{C} - 1$. By Proposition 1, there exists an optimal threshold $\hat{n}_s$ less than or equal to $\hat{n}_e$ (i.e., $\exists \hat{n}_s \le \hat{n}_e = \lfloor \frac{R\mu}{C} \rfloor$). Thus, for a strategic queue with mean arrival rate $m \le 1$, Theorem 1 can be applied to compute the worst-case expected social benefit rate for the first $\lfloor \frac{R\mu}{C} \rfloor - 1$ cases. This greatly speeds up the time to solve (4) because we only need to solve an SDP once for the remaining case $n = \lfloor \frac{R\mu}{C} \rfloor$. On the other hand, for a strategic queue with mean arrival rate $m > 1$, we cannot invoke Theorem 1 anymore, and we need to solve an SDP for each n satisfying $1 \le n \le \hat{n}_e$, $n \in \mathbb{Z}_+$.

3.2. Revenue Optimization

To determine an optimal joining threshold for a revenue maximizer, we compute the worst-case expected revenue rate $Z_r(n)$ for every $n \in \mathbb{Z}_+$, $1 \le n \le n_e$, and we choose an $\hat{n}_r$ such that $\hat{n}_r \in \arg\max_{n\in\mathbb{Z}_+} \{Z_r(n)\}$. To this end, we show how to compute the worst-case expected revenue for each n. Suppose the mean and MAD of the random parameter $\tilde{\rho}$ are known; then, the worst-case expected revenue rate is given by the following optimization problem:

$$
\begin{aligned}
Z_r(n) = \inf_{\nu \in \mathcal{M}_+} \quad & \int_{\Xi} r_n(\rho)\nu(\mathrm{d}\rho) \\
\text{s.t.} \quad & \int_{\Xi} |\rho - m|\, \nu(\mathrm{d}\rho) = d \\
& \int_{\Xi} \rho\, \nu(\mathrm{d}\rho) = m \\
& \int_{\Xi} \nu(\mathrm{d}\rho) = 1.
\end{aligned} \tag{12}
$$

To derive a tractable reformulation, we first investigate the property of the revenue rate function $r_n(\rho)$.

Lemma 4. *The revenue rate function $r_n(\rho)$ is concave for $\rho \in \mathbb{R}_+$.*

Equipped with Lemma 4, we now show that the worst-case expectation Problem (12) admits a closed-form solution.

Theorem 3. *For any $n \geq 1$, the worst-case expected revenue rate can be derived as*

$$Z_r(n) = \frac{d}{2(m-a)} r_n(a) + \left(1 - \frac{d}{2(m-a)} - \frac{d}{2(b-m)}\right) r_n(m) + \frac{d}{2(b-m)} r_n(b).$$

To prove this theorem, we invoke a classical result that characterizes the worst-case distribution from the MAD ambiguity set for a concave loss function.

Lemma 5 (Ben-Tal and Hochman 1972, Theorem 3). *Suppose $f(\rho)$ is a concave function and the ambiguity set is defined as $\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0([a,b]) : \mathbb{E}_{\mathbb{P}}[\tilde{\rho}] = m, \mathbb{E}_{\mathbb{P}}[|\tilde{\rho} - m|] = d\}$. The extremal distribution that solves $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[f(\tilde{\rho})]$ is a three-point distribution supported on $\rho_1 = a$, $\rho_2 = m$, $\rho_3 = b$ with probabilities*

$$p_1 = \frac{d}{2(m-a)},\ p_2 = 1 - \frac{d}{2(m-a)} - \frac{d}{2(b-m)},\ p_3 = \frac{d}{2(b-m)}. \tag{13}$$

Proof of Theorem 3. From Lemma 4, the revenue rate function $r_n(\rho)$ is concave. Therefore, applying Lemma 5 yields the result. □

Theorem 3 provides practical managerial insight for the decision maker. Observe that the extremal distribution is usually a discrete distribution supported on the mean and the lower and upper bounds of the support. Hence, instead of optimizing over the empirical distribution, the DRO scheme simplifies the problem into three cases: when the traffic intensity is extremely small ($\rho = a$), extremely large ($\rho = b$), or as expected ($\rho = m$). The weight for each scenario is determined by the MAD, which reflects the variation level of samples. This result aligns well with human intuition. To design a robust queue-regulating strategy, the decision maker may intuitively think about "how the queue behaves when the traffic intensity is extremely large, small, or as usual" and "what is the probability of these scenarios happening." The closed-form solution provides an answer to these questions. For example, the quantities p_1, p_2, p_3 answer the question "What are the probabilities of these scenarios happening?," whereas the quantities $r_n(a), r_n(b), r_n(m)$ answer the question "How does the queue behave when the traffic intensity is extremely large, small, or as expected?"

4. Extension to Data-Driven Problems

In this section, we apply the MAD ambiguity set to data-driven optimization problems. As we observed in the previous section, distributionally robust models with a moment ambiguity set necessitate decision makers to have access to exact values of the mean m or MAD d of the true unknown distribution, which may not be realistic in practice. A common approach is to construct such moment ambiguity sets by plugging in the point estimators generated from the historical samples. However, it is rarely the case that one can be entirely confident in these empirical estimators. For example, when the sample size is small, these empirical estimators might be far away from the true values; furthermore, some estimators, such as the empirical MAD, are even biased. In order to mitigate the adverse effects of the estimation errors, we design a data-driven MAD ambiguity set that contains the true underlying distribution with high confidence.

Unlike the setting in the previous section, here we assume that the queue system manager only has access to N independent and identically distributed samples of the traffic intensity given by $\{\hat{\rho}_i\}_{i\in[N]}$, where $\hat{\rho}_i = \hat{\lambda}_i/\mu$. In addition, we assume that decision makers have some prior knowledge or an educated estimate of the distribution support. Suppose the true mean and MAD of the underlying distribution are unknown and with high probabilities, belong to two confidence intervals $\mathcal{T} = [m_l, m_u]$ and $\mathcal{D} = [d_l, d_u]$ constructed using the samples. Then, the proposed data-driven distributionally robust model is formulated as

$$\overline{Z}(n) := \inf_{\mathbb{P}\in\mathcal{P}'_N} \mathbb{E}_{\mathbb{P}}[c_n(\tilde{\rho})], \tag{14}$$

where the modified data-driven ambiguity set is defined as

$$\mathcal{P}'_N = \{\mathbb{P} \in \mathcal{P}_0([a,b]) : m_l \le \mathbb{E}_{\mathbb{P}}[\tilde{\rho}] \le m_u,\ d_l \le \mathbb{E}_{\mathbb{P}}[|\tilde{\rho} - m|] \le d_u\}. \tag{15}$$

One can verify that the results of Propositions 1 and 2 still hold, and we can obtain the optimal value of (14) by solving $\overline{Z}(n)$ for each $n \in \mathbb{Z}_+$ satisfying $1 \le n \le n_e$ and select the one with the largest objective value.

We now derive the reformulations for the worst-case expected social benefit and revenue rates. To this end, we define the worst-case expected social benefit rate with the data-driven MAD ambiguity set by

$$\overline{Z}_s(n) := \inf_{\mathbb{P}\in\mathcal{P}'_N} \mathbb{E}_{\mathbb{P}}[f_n(\tilde{\rho})],$$

and we define the worst-case expected revenue rate with the data-driven MAD ambiguity set by

$$\overline{Z}_r(n) := \inf_{\mathbb{P}\in\mathcal{P}'_N} \mathbb{E}_{\mathbb{P}}[r_n(\tilde{\rho})].$$

The next theorem presents the reformulation of the worst-case expected social benefit rate. Some of the proofs of the results in this section are relegated to Appendix C.

Theorem 4. *For any $n \ge 1$, the worst-case expected social benefit rate $\overline{Z}_s(n)$ coincides with the optimal value of the following semidefinite program*:

$$\begin{array}{lll}
\textit{sup} & \gamma + \theta_1 d_l - \theta_2 d_u + \theta_3 m_l - \theta_4 m_u & \\
\textit{s.t.} & \gamma \in \mathbb{R}, \theta_1, \theta_2, \theta_3, \theta_4 \in \mathbb{R}_+, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^{n+3}, \boldsymbol{X}, \boldsymbol{X}' \in \mathbb{S}_+^{n+3} & \\
& y_1 = R\mu - C - y_0 + y_{n+3},\ y_2 = -R\mu - y_{n+3}, & \\
& y_3, \ldots, y_n = 0,\ y_{n+1} = -R\mu + C(n+1) - y_0, & \\
& y_{n+2} = R\mu - Cn + y_0 - y_{n+3}, & \\
& y_0 = (\theta_1 - \theta_2)\hat{m} + \gamma, y_{n+3} = -\theta_1 + \theta_2 + \theta_3 - \theta_4 & \\
& \displaystyle\sum_{i+j=2l-1} x_{ij} = 0 & \forall l \in [n+4] \\
& \displaystyle\sum_{i+j=2l} x_{ij} = \sum_{q=0}^{l} \sum_{r=q}^{n+3+q-l} y_r \binom{r}{q} \binom{n+3-r}{l-q} a^{r-q} \hat{m}^q & \forall l \in [n+4] \cup \{0\} \\
& z_1 = R\mu - C - z_0 + z_{n+3},\ z_2 = -R\mu - z_{n+3}, & \\
& z_3, \ldots, z_n = 0,\ z_{n+1} = -R\mu + C(n+1) - z_0, & \\
& z_{n+2} = R\mu - Cn + z_0 - z_{n+3} & \\
& z_0 = -(\theta_1 - \theta_2)\hat{m} + \gamma, z_{n+3} = \theta_1 - \theta_2 + \theta_3 - \theta_4 & \\
& \displaystyle\sum_{i+j=2l-1} x'_{ij} = 0 & \forall l \in [n+4] \\
& \displaystyle\sum_{i+j=2l} x'_{ij} = \sum_{q=0}^{l} \sum_{r=q}^{n+3+q-l} y'_r \binom{r}{q} \binom{n+3-r}{l-q} \hat{m}^{r-q} b^q & \forall l \in [n+4] \cup \{0\}.
\end{array}$$

Note that when $d_l = d_u$ and $m_l = m_u$, setting $\alpha = \theta_1 - \theta_2$ and $\beta = \theta_3 - \theta_4$ recovers the dual Problem (9) in the view of the primitive MAD ambiguity set, which corresponds to the case when we have absolute trust on the mean and MAD estimators.

The next theorem presents the reformulation of the worst-case expected revenue rate.

Theorem 5. *For any $n \geq 1$, the worst-case expected revenue rate $\overline{Z}_r(n)$ is equal to the optimal value of the following linear problem*:

$$\begin{aligned}
\sup_{\theta \in \mathbb{R}_+^4, \gamma \in \mathbb{R}} \quad & \gamma + \theta_1 d_l - \theta_2 d_u + \theta_3 m_l - \theta_4 m_u \\
\text{s.t.} \quad & (\theta_1 - \theta_2)|a - \hat{m}| + (\theta_3 - \theta_4)a + \gamma \leq r_n(a) \\
& (\theta_3 - \theta_4)\hat{m} + \gamma \leq r_n(\hat{m}) \\
& (\theta_1 - \theta_2)|b - \hat{m}| + (\theta_3 - \theta_4)b + \gamma \leq r_n(b).
\end{aligned}$$

Theorems 4 and 5 provide tractable reformulations for the social and revenue optimization problems. An advantage of the proposed data-driven model is that it can offer attractive finite-sample guarantees. Compared with the original MAD ambiguity set that imposes unique mean and MAD, the data-driven MAD ambiguity set allows these parameters to vary within the confidence intervals. In this way, we can assure that the set contains the true underlying distribution with a high probability, which immediately generates out-of-sample performance guarantees for the solution.

Theorem 6. *Let $\{\hat{\rho}_i\}_{i \in [N]}$ be a set of N samples generated independently at random from $\mathbb{P}^\star$ and $v^\star$ denote the optimal value of* (14). *Define $\hat{m}$ and $\hat{d}$ as the empirical mean and MAD obtained from samples $\{\hat{\rho}_i\}_{i \in [N]}$. By setting*

$$\begin{aligned}
\mathcal{T} &= \left[\hat{m} - (b-a)\sqrt{\frac{\log 4/\delta}{2N}}, \hat{m} + (b-a)\sqrt{\frac{\log 4/\delta}{2N}}\right] \\
\mathcal{D} &= \left[\hat{d} - (b-a)\sqrt{\frac{9 \log 4/\delta}{2N}}, \hat{d} + (b-a)\sqrt{\frac{9 \log 4/\delta}{2N}}\right],
\end{aligned} \tag{16}$$

we have

$$\text{Prob}(v^\star \leq \mathbb{E}_{\mathbb{P}^\star}[c_{\hat{n}}(\tilde{\rho})]) \geq 1 - \delta,$$

where $\hat{n}$ is the optimal threshold obtained from (14).

Proof. The error of the empirical MAD estimate is given by

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^N |\hat{\rho}_i - \hat{m}| - \mathbb{E}[|\tilde{\rho} - \hat{m}|]\right| \\
&= \max\left\{\frac{1}{N}\sum_{i=1}^N |\hat{\rho}_i - \hat{m}| - \mathbb{E}[|\tilde{\rho} - \hat{m}|], -\frac{1}{N}\sum_{i=1}^N |\hat{\rho}_i - \hat{m}| + \mathbb{E}[|\tilde{\rho} - \hat{m}|]\right\}.
\end{aligned}$$

We upper bound both terms inside the max operator. The first term is bounded by

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^N |\hat{\rho}_i - \hat{m}| - \mathbb{E}[|\tilde{\rho} - \hat{m}|] &\leq \frac{1}{N}\sum_{i=1}^N |\hat{\rho}_i - \hat{m}| - \mathbb{E}[||\tilde{\rho} - m| - |\hat{m} - m||] \\
&\leq \frac{1}{N}\sum_{i=1}^N |\hat{\rho}_i - \hat{m}| - \mathbb{E}[|\tilde{\rho} - m| - |\hat{m} - m|] \\
&\leq \left|\frac{1}{N}\sum_{i=1}^N |\hat{\rho}_i - \hat{m}| - \mathbb{E}[|\tilde{\rho} - m|]\right| + |\hat{m} - m|,
\end{aligned}$$

where the second inequality follows from reverse triangle inequality. Meanwhile, the second term is bounded by

$$\begin{aligned}
-\frac{1}{N}\sum_{i=1}^N |\hat{\rho}_i - \hat{m}| + \mathbb{E}[|\tilde{\rho} - \hat{m}|] &\leq -\frac{1}{N}\sum_{i=1}^N |\hat{\rho}_i - \hat{m}| + \mathbb{E}[|\tilde{\rho} - m| + |\hat{m} - m|] \\
&\leq \left|\frac{1}{N}\sum_{i=1}^N |\hat{\rho}_i - \hat{m}| - \mathbb{E}[|\tilde{\rho} - m|]\right| + |\hat{m} - m|.
\end{aligned}$$

Because both of these two terms have the same upper bound, we have

$$\left|\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-\hat{m}|-\mathbb{E}[|\tilde{\rho}-\hat{m}|]\right| \le \left|\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-\hat{m}|-\mathbb{E}[|\tilde{\rho}-m|]\right|+|\hat{m}-m|.$$

As $\mathbb{E}[\hat{m}]=\mathbb{E}[m]$ is an unbiased estimator, we can invoke the Hoeffding inequality to directly derive a confidence interval for the second term. However, the empirical MAD is biased (i.e., $\mathbb{E}[\frac{1}{N}\sum_{i=1}^{N}|\rho_i-\hat{m}|]\neq\mathbb{E}[|\rho-m|]$), making the Hoeffding inequality not applicable. To derive a confidence interval for this term, we rewrite it as

$$\begin{aligned}&\left|\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-\hat{m}|-\mathbb{E}[|\tilde{\rho}-m|]\right|\\&=\max\left\{\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-\hat{m}|-\mathbb{E}[|\tilde{\rho}-m|],\ -\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-\hat{m}|+\mathbb{E}[|\tilde{\rho}-m|]\right\}.\end{aligned}$$

We further upper bound the two terms inside the max operator. For the first term, we have

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-\hat{m}|-\mathbb{E}[|\tilde{\rho}-m|] &\le \frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-m|+|m-\hat{m}|-\mathbb{E}[|\tilde{\rho}-m|]\\&\le\left|\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-m|-\mathbb{E}[|\tilde{\rho}-m|]\right|+|m-\hat{m}|.\end{aligned}$$

For the second term, applying the reverse triangle inequality yields

$$\begin{aligned}\mathbb{E}[|\tilde{\rho}-m|]-\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-\hat{m}| &\le \mathbb{E}[|\tilde{\rho}-m|]-\frac{1}{N}\sum_{i=1}^{N}\big||\hat{\rho}_i-m|-|\hat{m}-m|\big|\\&\le \mathbb{E}[|\tilde{\rho}-m|]-\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-m|+|\hat{m}-m|\\&\le\left|\mathbb{E}[|\tilde{\rho}-m|]-\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-m|\right|+|\hat{m}-m|.\end{aligned}$$

Thus, we have

$$\left|\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-\hat{m}|-\mathbb{E}[|\tilde{\rho}-m|]\right| \le \left|\mathbb{E}[|\tilde{\rho}-m|]-\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-m|\right|+2|\hat{m}-m|.$$

Because both of these two terms are unbiased, we can apply the Hoeffding inequality and obtain

$$\begin{aligned}&\mathrm{Prob}\left(\left|\mathbb{E}[|\tilde{\rho}-m|]-\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-m|\right|\ge r_1\right)\le 2\exp\left(-\frac{2Nr_1^2}{(b-a)^2}\right)\quad\text{and}\\&\mathrm{Prob}(|\hat{m}-m|\ge r_2)\le 2\exp\left(-\frac{2Nr_2^2}{(b-a)^2}\right).\end{aligned}$$

By applying the union bound and setting $r_1=r_2=r/3$, we arrive at the desired confidence intervals that the true mean m and MAD d satisfy

$$\begin{aligned}\hat{m}-(b-a)\sqrt{\frac{\log 4/\delta}{2N}}&\le m\le\hat{m}+(b-a)\sqrt{\frac{\log 4/\delta}{2N}}\\\hat{d}-(b-a)\sqrt{\frac{9\log 4/\delta}{2N}}&\le d\le\hat{d}+(b-a)\sqrt{\frac{9\log 4/\delta}{2N}}\end{aligned}$$

with probability at least $1-\delta$. Therefore, by setting the confidence interval $\mathcal{T}$ and $\mathcal{D}$ as in (16), we have

$$\mathrm{Prob}(\mathcal{P}'_N\ni\mathbb{P}^\star)\ge 1-\delta,$$

where $\mathcal{P}'_N$ is the data-driven ambiguity set (15) constructed by N random samples drawn from the underlying distribution $\mathbb{P}^\star$. Because $v^\star:=\inf_{\mathbb{P}\in\mathcal{P}'_N}\mathbb{E}_{\mathbb{P}}[c_n(\tilde{\rho})]$ and the probability that $\mathcal{P}'_N$ contains the true distribution $\mathbb{P}^\star$ is at least $1-\delta$, we have

$$\text{Prob}(v^\star \leq \mathbb{E}_{\mathbb{P}^\star}[c_{\hat{n}}(\tilde{\rho})]) \geq 1-\delta,$$

which completes the proof. □

The theorem establishes that with judicious choices of the confidence interval lengths, the optimal value of the data-driven DRO model $v^\star$ provides a high-confidence lower bound on the expected benefit rate of the robust solution $\hat{n}$ under the true underlying distribution $\mathbb{P}^\star$.

Remark 2. An avid reader may be interested in employing the popular Wasserstein DRO model in the data-driven setting. Indeed, the model has been widely adopted because it can generate asymptotically consistent solutions and offer similarly attractive finite-sample guarantees. Unfortunately, the reformulation of this data-driven DRO model involves $\mathcal{O}(N)$ semidefinite constraints, which make the problem computationally intensive. For readers who are interested in the use of the Wasserstein ambiguity set, we provide a detailed discussion in Appendix B.

5. Numerical Experiments

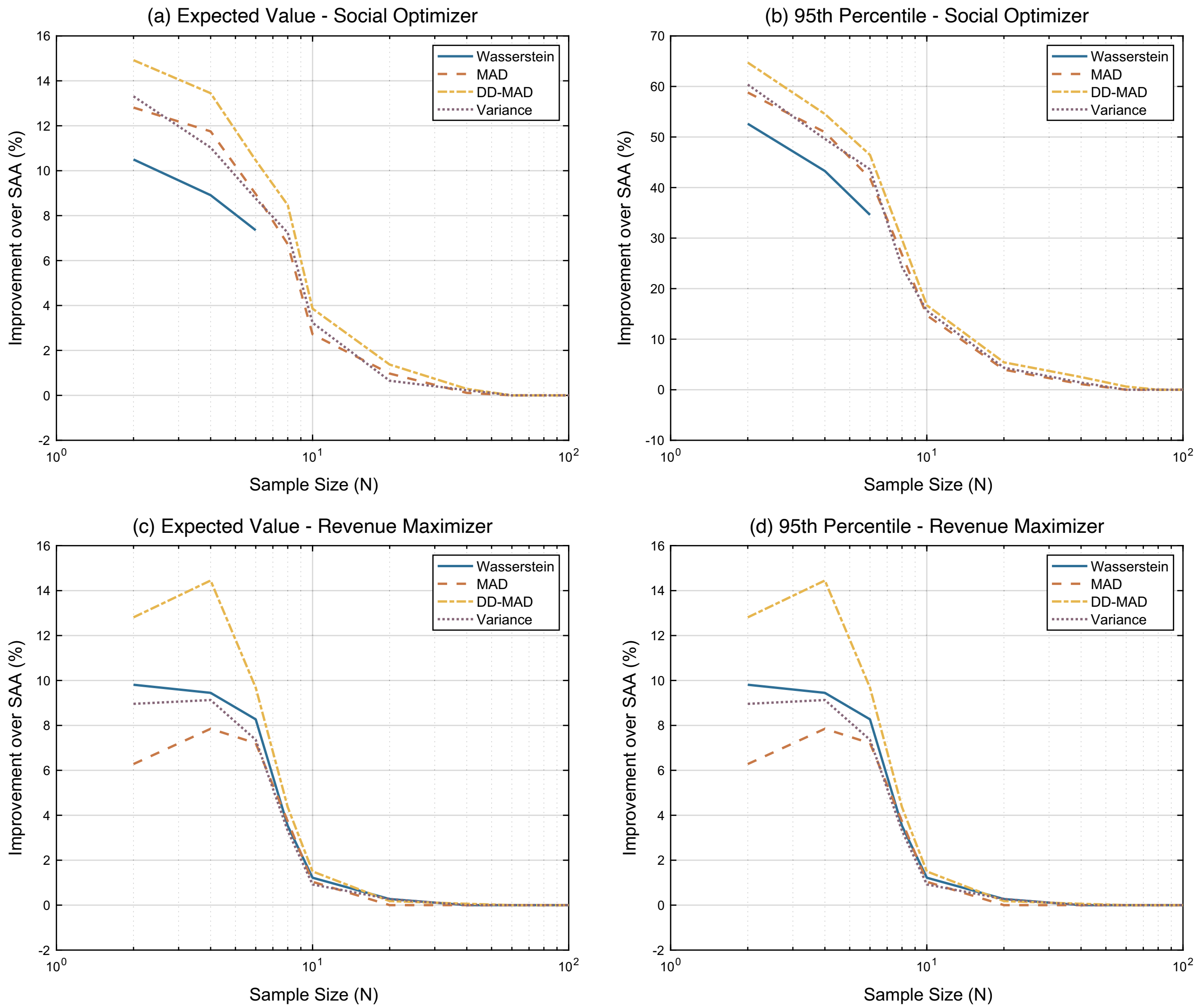
In this section, we present the numerical experiments and examine the performance of different DRO policies. All optimization problems are implemented in MATLAB and solved by SDPT3 (Toh et al. 1999) via the YALMIP interface (Lofberg 2004). The experiments are run on a 2.2-GHz Intel Core i7 CPU laptop with 8 GB RAM.

We assess the out-of-sample performance of the data-driven policies for a social optimizer and a revenue maximizer through a fair out-of-sample experiment. We assume we have access to N independent samples $\{\hat{\rho}_i\}_{i\in[N]}$ of the traffic intensity drawn from the true underlying distribution $\mathbb{P}^\star$, and we construct four ambiguity sets: an empirical MAD ambiguity set, an empirical variance ambiguity set, a DD-MAD ambiguity set, and a Wasserstein ambiguity set. The empirical MAD ambiguity set is defined in (7), where we directly substitute the empirical mean and MAD for m and d, respectively. The empirical variance model is another popular moment model that constructs its ambiguity set based on the empirical mean and variance (i.e., $\mathbb{E}[(\tilde{\rho}-m)^2]=\sigma^2$). Because its formulation and derivation parallel those of the empirical MAD model, we omit its discussion for brevity. The DD-MAD ambiguity set is defined in (15), where rather than carelessly plugging in the empirical estimators, we construct a confidence interval around the empirical mean and MAD. The Wasserstein ambiguity set (Esfahani and Kuhn 2018, Gao and Kleywegt 2023) is a popular data-driven ambiguity set. However, its complexity scales with the number of samples, making the problem computationally intensive with large sample sizes. We derive the reformulation of the Wasserstein model in Appendix D. Once we constructed the different ambiguity sets, we then proceed to compute the distributionally robust thresholds that maximize the respective worst-case expected benefit rates. Finally, we compare the three solutions in a fair out-of-sample experiment relative to the sample average approximation (SAA) method, which naively assumes that the empirical distribution generated from the N samples is the true underlying distribution. The SAA method also represents the stochastic model (Liu and Hasenbein 2019) under the empirical distribution.

We conduct the out-of-sample trials for data sets containing $N=2,4,\ldots,10,20,40\ldots,100$ independent samples. We assume the arrival rate is generated by $\lambda=4\tilde{b}$, where $\tilde{b}\sim\text{Beta}(0.1,0.5)$. In addition, we assume the experienced decision maker has an educated guess for the distribution support as $[0,5]$. In each trial, we draw N independent training samples and obtain $\{\hat{\rho}_i\}_{i\in[N]}$ from $\mathbb{P}^\star$. We then compute the optimal thresholds $\hat{n}_d$, n_v,$\hat{n}_{dd}$, and $\hat{n}_w$ for the MAD, variance, DD-MAD, and Wasserstein DRO models, respectively. We also compute the SAA threshold $\hat{n}_{SAA}$ by solving the sample average approximation model. Based on the scaling rates derived in Theorem 6 and Esfahani and Kuhn (2018, theorem 3.4), the size of the confidence intervals in (14) is set to be $C_1/\sqrt{N}$, and the Wasserstein radius is set to be $C_2/\sqrt{N}$, where C_1 and C_2 are chosen from the set $\{5,1,0.5,0.1,0.05,0.01\}$ using a *k-fold cross validation* procedure. Specifically, we partition the in-sample data $\{\hat{\rho}_i\}_{i\in[N]}$ into $k=\min\{N,5\}$ folds and repeat the following procedure for each fold; the ith fold is taken as a validation data set, and the remaining $k-1$ folds are merged to be a subtraining set. We repeat this process for each fold and choose the interval length that performs best in average. The reason why we do not directly plug in the theoretical values from Theorem 6 is that the bound holds for any underlying distributions, which can be overly conservative in practice. Finally, the out-of-sample expected benefit/revenue rate $\mathbb{E}_{\mathbb{P}^\star}[c_{\hat{n}}(\tilde{\rho})]$ for each of the strategies is then estimated at high accuracy using 10,000 test samples from $\mathbb{P}^\star$.

Figure 2 depicts the out-of-sample performances of a social optimizer and a revenue optimizer under different DRO policies with $R=10$, $C=1$, and $\mu=1$. The expected values and 95 percentiles are computed from 50 independent trials. The y axis represents the improvements of the DRO policies relative to the SAA policy, whereas

Figure 2. Improvements of the DRO Policies Relative to the SAA Policy in Terms of the Social Optimizer and the Revenue Maximizer, Respectively



the x axis denotes the sample size. In the social optimization problem, the curve of the Wasserstein model terminates at $n = 6$ because the solver fails to converge when the sample size reaches eight. Meanwhile, the Wasserstein model dominates the SAA model uniformly across all sample sizes in the revenue maximization problem, whereas the MAD, variance, and DD-MAD models outperform the SAA model for small to medium sample sizes. This is because the Wasserstein ambiguity set converges to the true distribution as the number of samples grows, whereas the moment ambiguity sets fail to converge to the true distribution. We also find that the MAD model performs poorly when the sample size is small because the empirical MAD constitutes a biased estimator with significant estimation errors. On the other hand, the DD-MAD model—by optimizing in view of the most adverse mean and MAD—mitigates the detrimental effects of poor empirical estimations and generates high-quality policies. Moreover, we notice that the empirical MAD and variance models yield similar scores, suggesting that measuring dispersion by MAD or variance does not influence the performance of the model. Finally, we observe that the advantages of the DRO policies relative to the SAA method are generally more substantial in terms of the 95th percentiles of improvements. This underlines a major advantage of incorporating the DRO scheme as it reduces the likelihood of realizing inferior performance in the out-of-sample tests.

Table 1 reports the computation time of different models with the sample sizes varying from 2 to 100. We set the length of the confidence intervals and the radius of the Wasserstein ball to 0.1. In this experiment, the running time limit of SDPT3 is set to 600 seconds, and the number of iterations is set to 5,000. All computational times are averaged over 10 trials.

Table 1. Running Time (in Seconds) of Different Methods

	Sample size N					
Model name	2	5	10	25	50	100
Social						
MAD	18.63	17.44	19.28	18.15	19.62	20.31
DD-MAD	31.46	32.58	30.19	33.64	32.84	31.52
Wasserstein	39.26	78.53	—	—	—	—
Variance	29.11	30.62	29.94	31.47	30.53	30.72
Revenue						
MAD	0.05	0.04	0.04	0.05	0.06	0.06
DD-MAD	1.48	1.52	1.66	1.92	1.73	1.70
Wasserstein	1.69	1.92	2.41	2.63	2.95	4.68
Variance	31.42	30.73	31.61	31.55	31.79	32.80

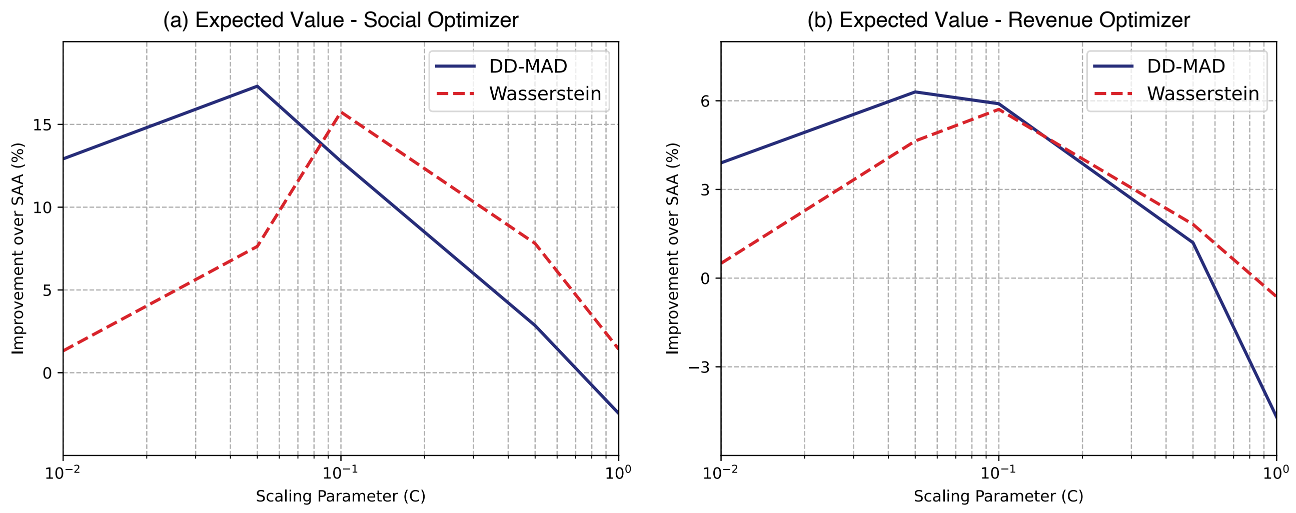
Note. The — symbol indicates that the model fails to converge in the maximal iteration/time.

The results in Table 1 indicate that the computational times of the MAD, variance, and DD-MAD models are size invariant because the number of constraints is independent of the number of samples. For the social optimization problem, the Wasserstein model is applicable to small-size instances. However, it encounters computational difficulties for moderate-size problem instances; when the sample size reaches 10, the SDP solver fails to converge within the time/iteration limit. Benefiting from the closed-form solution, the MAD model is more efficient than the variance and DD-MAD models in both the social and revenue optimization problems. For the revenue optimization problem, the DD-MAD and Wasserstein models admit a linear programming (LP) reformulation, whereas the variance model still leads to an SDP reformulation. This underlines a major advantage of using MAD at the moment ambiguity set as it can significantly improve the model's efficiency.

The DD-MAD model is still size invariant, and its linear programming reformulation yields a much shorter computational time than the SDP reformulation for the social optimization problem. In addition, the Wasserstein model can be solved efficiently even for large sample sizes, benefiting from the linear programming reformulation.

Finally, we report the performance of the DRO models under different scaling parameters C in Figure 3. The expected improvements are computed with $n = 5$ samples from 50 independent trials. We observe that both models have large variations in performance with different scaling parameters. The Wasserstein and DD-MAD models yield unimodal curves in both the social and revenue optimization problems, implying a trade-off between performance and conservatism. Intuitively, including robustness can improve the out-of-sample performance, whereas being too conservative may also adversely affect the results. To achieve the best performance, one could set the size of the ambiguity set or confidence interval to the best radius. Unfortunately, we do not have access to this information. Although one can plug in the theoretical values obtained from concentration inequalities, these values are usually too conservative. In practice, decision makers can rely on a crossvalidation

Figure 3. Improvements of the DRO Policies Relative to the SAA Policy with Different Scaling Parameters



or bootstrap procedure to obtain a suitable size for the ambiguity set (Gotoh et al. 2021, Bates et al. 2023). From the figure, we further observe that the two models perform quite differently when the scaling parameter C is small; the DD-MAD model achieves significant improvement, whereas the Wasserstein model only yields a slight improvement. The reason is that the Wasserstein ambiguity set is centered on the empirical distribution. When the scaling parameter is small, all the distributions within this ambiguity set are close to the empirical distribution. Thus, it generates similar results to SAA and cannot achieve a large improvement. Conversely, the DD-MAD model converges to the empirical MAD model when $C = 0$. As illustrated in the first set of experiments, the MAD model outperforms the SAA method for small sample sizes. Hence, the DD-MAD method yields a substantial improvement when the scaling parameter is small.

6. Conclusion

This paper developed an extension of the Naor (1969) strategic queue model with uncertain arrival rates using the DRO framework. We showed that under the DRO setting, the optimal threshold of an individual optimizer coincides with the original result of Naor (1969), and there exist optimal thresholds of the social and revenue optimizers not larger than the optimal individual threshold. We then proved that the revenue rate function is concave, whereas the social benefit rate function is concave or unimodal under some mild conditions. These nice properties lead to a closed-form solution for the revenue maximization problem and an analytical solution for the social optimization problem.

Next, we considered the data-driven optimization setting, where decision makers only have access to limited historical samples. We proposed a data-driven MAD model by introducing an extra layer of robustness to the primitive MAD ambiguity set. As the model mitigates the detrimental estimation errors from the empirical mean and MAD, it achieves attractive performance in out-of-sample tests. We derived an SDP reformulation for the social optimization problem and a linear programming reformulation for the revenue maximization problem. We further established finite-sample guarantees for the data-driven model, which provide valuable guidance for choosing the robustness parameters in practice. Our experimental results demonstrate that a system manager who disregards ambiguities in the arrival rate distribution as well as errors from the empirical parameter estimations may incur large out-of-sample costs. Future work includes extending the DRO scheme to the unobservable strategic queues, where newly arrived customers cannot observe the current length of the queue system.

Appendix A. Proofs of Section 2

Proof of Proposition 1. It is established in Naor (1969, equation 30) that for any deterministic arrival rate λ and service rate μ, the optimal threshold from the perspective of a public goods regulator will be less than or equal to the optimal threshold of an individual customer. Suppose that every optimal threshold that maximizes the worst-case expected social benefit rate is strictly greater than the optimal threshold of an individual customer (i.e., $\hat{n}_s > \hat{n}_e$ for all $\hat{n}_s \in \arg\max_{n\in\mathbb{Z}_+} \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P}[f_n(\tilde{\rho})]$). Then, based on our previous statement, for any fixed ρ and any optimal $\hat{n}_s$, we have $n_s(\rho) \le n_e = \hat{n}_e < \hat{n}_s$, where $n_s(\rho)$ is the corresponding optimal social threshold under the deterministic setting. Because $f_n(\rho)$ is discretely unimodal for any fixed ρ (Naor 1969, p. 20), the relationship of the benefit rate can consequently be derived as

$$f_{n_s(\rho)}(\rho) \ge f_{\hat{n}_e}(\rho) \ge f_{\hat{n}_s}(\rho) \quad \forall \rho \in \mathbb{R}_+.$$

Using this relationship, one can further establish that for any ambiguity set $\mathcal{P}$,

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P}[f_{\hat{n}_e}(\tilde{\rho})] \ge \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P}[f_{\hat{n}_s}(\tilde{\rho})].$$

Conversely, by the definition of $\hat{n}_s$, we also have $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P}[f_{\hat{n}_e}(\tilde{\rho})] \le \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P}[f_{\hat{n}_s}(\tilde{\rho})]$. This implies that $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P}[f_{\hat{n}_e}(\tilde{\rho})] = \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P}[f_{\hat{n}_s}(\tilde{\rho})]$. Therefore, $\hat{n}_e$ is also an optimal threshold of the social optimization problem, which contradicts our previous assumption. This completes the proof. □

Proof of Proposition 2. The proof parallels that of Proposition 1—we omit it for brevity. □

Appendix B. Proofs of Section 3

Lemma B.1. *The first and second derivatives of the social benefit rate function $f_n(\rho)$ are continuous.*

Proof. To show the continuity of the first and second derivatives of $f_n(\rho)$, we will show that

$$f_n(\rho) = R\mu\left(1 - \frac{1}{\sum_{k=0}^{n} \rho^k}\right) - C\left(\frac{\rho(\sum_{k=0}^{n-1} \rho^k) + \rho^2(\sum_{k=0}^{n-2} \rho^k) + \cdots + \rho^n}{(\sum_{k=0}^{n} \rho^k)}\right), \tag{B.1}$$

which has continuous first and second derivatives.

First, we perform the transformation for the term $\frac{\rho(1-\rho^n)}{1-\rho^{n+1}}$ when $\rho \neq 1$. Note that $\frac{\rho(1-\rho^n)}{1-\rho^{n+1}} = 1 - \frac{1-\rho}{1-\rho^{n+1}}$, and the denominator is equal to $(1-\rho)(1+\rho+\rho^2+\cdots+\rho^n)$. We can consequently rewrite the first term as

$$\frac{\rho(1-\rho^n)}{1-\rho^{n+1}} = 1 - \frac{1}{\sum_{k=0}^{n}\rho^k}.$$

Next, we prove the equivalence of the remaining part $\frac{(n+1)\rho^{n+1}}{1-\rho^{n+1}} - \frac{\rho}{1-\rho}$ when $\rho \neq 1$. Similarly, by the fact that $(1-\rho^{n+1}) = (1-\rho)(\sum_{k=0}^{n}\rho^k)$, we can rewrite this part as

$$\begin{aligned}
\frac{(n+1)\rho^{n+1}}{1-\rho^{n+1}} - \frac{\rho}{1-\rho} &= \frac{(n+1)\rho^{n+1}}{(1-\rho)(\sum_{k=0}^{n}\rho^k)} - \frac{\rho(\sum_{k=0}^{n}\rho^k)}{(1-\rho)(\sum_{k=0}^{n}\rho^k)} \\
&= \frac{-\rho-\rho^2-\cdots-\rho^n+n\rho^{n+1}}{(1-\rho)(\sum_{k=0}^{n}\rho^k)} \\
&= \frac{\rho^{n+1}-\rho+\rho^{n+1}-\rho^2+\cdots+\rho^{n+1}-\rho^n}{(1-\rho)(\sum_{k=0}^{n}\rho^k)} \\
&= \frac{\rho(\rho-1)(1+\rho+\cdots+\rho^{n-1})+\rho^2(\rho-1)(1+\rho+\cdots+\rho^{n-2})+\cdots+\rho^n(\rho-1)}{(1-\rho)(\sum_{k=0}^{n}\rho^k)} \\
&= -\frac{\rho(\sum_{k=0}^{n-1}\rho^k)+\rho^2(\sum_{k=0}^{n-2}\rho^k)+\cdots+\rho^n}{(\sum_{k=0}^{n}\rho^k)}.
\end{aligned}$$

When $\rho = 1$, $R\mu\left(1-\frac{1}{\sum_{k=0}^{n}\rho^k}\right) - C\left(\frac{\rho(\sum_{k=0}^{n-1}\rho^k)+\rho^2(\sum_{k=0}^{n-2}\rho^k)+\cdots+\rho^n}{(\sum_{k=0}^{n}\rho^k)}\right) = R\mu\left(1-\frac{1}{1+n}\right) - C\frac{n}{2}$, which coincides with $f_n(1)$. Therefore, $f_n(\rho)$ is equal to (B.1). One can verify that the first and second derivatives of (B.1) are continuous; hence, $f_n(\rho)$ also has these properties. □

Lemma B.2. *The function $h_n(\rho) = \frac{\rho(1-\rho^n)}{1-\rho^{n+1}}$ is strictly concave and monotone increasing on $[0,1)\cup(1,\infty)$.*

Proof. When $\rho \in [0,1)\cup(1,\infty)$, the first derivative of $h_n(\rho)$ is

$$h_n'(\rho) = \frac{n\rho^{n+1}-(n+1)\rho^n+1}{(1-\rho^{n+1})^2}.$$

Define the numerator as $\varphi_n(\rho) = n\rho^{n+1}-(n+1)\rho^n+1$. The first derivative of $\varphi_n(\rho)$ is given by $\varphi_n'(\rho) = n(n+1)\rho^{n-1}(\rho-1)$. Note that when $0<\rho<1$, $\varphi_n'(\rho)$ is negative and that when $\rho>1$, $\varphi_n'(\rho)$ is positive. Therefore, the function $\varphi_n(\rho)$ is decreasing on (0, 1) and increasing on $(1,\infty)$. Meanwhile, by the fact that $\varphi_n(1) = 1+n-(n+1) = 0$, we know that the numerator $\varphi_n(\rho)$ is positive on $[0,1)\cup(1,\infty)$. Because the denominator $(1-\rho^{n+1})^2$ is positive, the first derivative $h_n'(\rho)$ is positive on $[0,1)\cup(1,\infty)$. Thus, we conclude that $h_n(\rho)$ is increasing on $[0,1)\cup(1,\infty)$.

Next, we show that the second derivative of $h_n(\rho)$ is negative. We have

$$h_n''(\rho) = \frac{(n+1)\rho^{n-1}[n\rho^{n+2}-(n+2)\rho^{n+1}+(n+2)\rho-n]}{(1-\rho^{n+1})^3}.$$

Because the term $\frac{(n+1)\rho^{n-1}}{(1-\rho^{n+1})^3}$ is positive on $[0,1)$ and is negative on $(1,\infty)$, we simply need to determine the sign of $[n\rho^{n+2}-(n+2)\rho^{n+1}+(n+2)\rho-n]$. For convenience, define

$$\psi_n(\rho) := n\rho^{n+2}-(n+2)\rho^{n+1}+(n+2)\rho-n.$$

Note that $\psi_n(0) = -n < 0$ and $\psi_n(1) = 0$, whereas $\lim_{\rho\to\infty}\psi_n(\rho) = +\infty$. Therefore, if $\psi_n(\rho)$ is increasing on $[0,1)\cup(1,\infty)$, the second derivative $h_n''(\rho)$ will be negative on $[0,1)\cup(1,\infty)$. To show this, we take the first derivative of $\psi_n(\rho)$ and obtain

$$\psi_n'(\rho) = n(n+2)\rho^{n+1}-(n+2)(n+1)\rho^n+(n+2).$$

Taking specific values into this function, we can obtain $\psi_n'(0) = n+2 > 0$, $\psi_n'(1) = 0$, and $\lim_{\rho\to\infty}\psi_n(\rho) = +\infty$. Similarly, if $\psi_n'(\rho)$ is decreasing on $[0,1)$ and increasing on $(1,\infty)$, then $\psi_n'(\rho)$ will be positive on $[0,1)\cup(1,\infty)$. To verify this, we can take the second derivative of $\psi_n(\rho)$, which gives

$$\psi_n''(\rho) = (n+2)(n+1)n\rho^{n-1}(\rho-1).$$

One can verify that $\psi_n''(\rho)$ is negative on $[0,1)$ and positive on $(1,\infty)$. Thus, we have established that $h_n''(\rho)$ is negative on $[0,1)\cup(1,\infty)$ and $h_n(\rho)$ is concave on $[0,1)\cup(1,\infty)$. □

Lemma B.3. *For any $v\in\mathbb{R}$, $v\geq 1$, the function $g_v(\rho) = \frac{v\rho}{1-\rho^v} - \frac{\rho}{1-\rho}$ is concave on $[0,1)$.*

Proof. For any $v \in \mathbb{R}$, $v \geq 1$, one can verify that $g_v(\rho)$ is continuous and second-order differentiable on $[0,1)$. Thus, $g_v(\rho)$ is concave if and only if its second derivative

$$g_v''(\rho) = \frac{v^2\rho^{v-1}(1+v+\rho^v(v-1))}{(1-\rho^v)^3} - \frac{2}{(1-\rho)^3}$$

is nonpositive for every $\rho \in [0,1)$. Notice that when $\rho = 0$, $g_v''(\rho) = -2$ is less than zero. We now prove that the second derivative is also nonpositive on (0, 1). We first observe that $g_v''(\rho) = 0$ at $v = 1$ for all $\rho \in [0,1)$. Consider the partial derivative with respect to v:

$$\partial_v g_v''(\rho) = \frac{v(((v^2-v)\ln(\rho) - 3v+2)\rho^{3v-1} + (4v^2\ln(\rho)-4)\rho^{2v-1} + \rho^{v-1}((v^2+v)\ln(\rho)+3v+2))}{(1-\rho^v)^4}.$$

If this function is nonpositive for all $v \in \mathbb{R}$, $v \geq 1$, then we can establish that the second derivative $g_v''(\rho)$ is nonpositive for all $\rho \in (0,1)$.

Consider a fixed $v \in \mathbb{R}$, $v \geq 1$. Defining $\psi(\rho)$ as the product of $\partial_v g_v''(\rho)$ and $\frac{(1-\rho^v)^4}{v\rho^{v-1}} > 0$ yields

$$\psi_v(\rho) := (v(v-1)\rho^{2v} + 4v^2\rho^v + v^2 + v)\ln(\rho) + (-3v+2)\rho^{2v} - 4\rho^v + 3v + 2.$$

We show that $\psi_v(\rho)$ is nonpositive for $\rho \in (0,1)$. Observe that $\psi_n(\rho)$ goes to negative infinity as $\rho \to 0_+$ and equals to zero at $\rho = 1$. Thus, it is sufficient to show that $\psi_v(\rho)$ is increasing on $\rho \in (0,1)$ for every fixed v. Taking the derivative with respect to ρ and dividing it by $v\rho^{v-1} > 0$ yields

$$\tau_n(\rho) := \frac{\psi_n'(\rho)}{v\rho^{v-1}} = 2v((v-1)\rho^v + 2v)\ln(\rho) + (v+1)\rho^{-v} + (-5v+3)\rho^v + 4v - 4.$$

Similarly, one can verify that this expression goes to positive infinity as $\rho \to 0_+$ and is equal to zero at $\rho = 1$. Therefore, to show that $\psi_n'(\rho)$ is positive on (0, 1), it is sufficient to show that $\tau_n(\rho)$ is decreasing on (0, 1). Again, taking the derivative with respect to ρ and dividing it by $v\rho^{v-1} > 0$, we get

$$\varphi(\rho) := \frac{\tau_n'(\rho)}{v\rho^{v-1}} = 2v(v-1)\ln(\rho) - (v+1)\rho^{-2v} + 4v\rho^{-v} - 3v + 1.$$

This expression again vanishes at $\rho = 1$ and goes to negative infinity as $\rho \to 0_+$. Thus, it is sufficient to show that it is increasing on $(0,1)$. Taking the derivative with respect to ρ and multiplying with $\frac{\rho^{2v+1}}{2v} > 0$ yield

$$\theta_n(\rho) := \frac{\varphi'(\rho)\rho^{2v+1}}{2v} = (v-1)\rho^{2v} - 2v\rho^v + v + 1.$$

At $\rho = 0$, $\theta_n(\rho)$ is equal to $v+1$, which is greater than zero, and vanishes at $\rho = 1$. Taking the derivative with respect to ρ and dividing by $2n\rho^{n-1} > 0$, we have

$$\phi_n(\rho) := \frac{\theta_n'(\rho)}{2n\rho^{n-1}} = (v-1)\rho^v - v.$$

One can verify that when $v \geq 1$, $\phi(\rho)$ is always nonpositive, which completes our proof. □

Proof of Lemma 1, Statement (1). Using the lemmas, we are ready to show that when $\frac{R\mu}{C} \geq n+1$, the social benefit rate function $f_n(\rho)$ is strictly concave on $[0,1]$. For $\rho \in [0,1)$, we can rewrite $f_n(\rho)$ as

$$f_n(\rho) = (R\mu - C(n+1))\frac{\rho(1-\rho^n)}{1-\rho^{n+1}} + C\frac{(n+1)\rho}{1-\rho^{n+1}} - \frac{C\rho}{1-\rho}.$$

From Lemma B.2 and Lemma B.3, we know that $\frac{\rho(1-\rho^n)}{1-\rho^{n+1}}$ is strictly concave and that $\frac{(n+1)\rho}{1-\rho^{n+1}} - \frac{\rho}{1-\rho}$ is concave. Therefore, $f_n(\rho)$ is the sum of a strictly concave function and a concave function, which is strictly concave for $\rho \in [0,1)$. □

Proof of Lemma 1, Statement (2). When $n = 1$, one can verify that $f_1(\rho)$ is a concave increasing function for $\rho \in \mathbb{R}_+$. We now proceed to show that the function is unimodal for $n \geq 2$. A sufficient condition for $f_n(\rho)$ to be unimodal is $f_n'(0) > 0$, $\lim_{\rho\to\infty} f_n'(\rho) < 0$, and $f_n'(\rho) = 0$ has a unique solution. Taking the derivative of $f_n(\rho)$ yields

$$f_n'(\rho) = \begin{cases} R\mu\left(\frac{n\rho^{n+1} - (n+1)\rho^n + 1}{(\rho^{n+1}-1)^2}\right) + C\left(\frac{(n+1)(n\rho^{n+1}+1)}{(\rho^{n+1}-1)^2} - \frac{1}{(\rho-1)^2}\right) & \text{if } \rho \neq 1 \\ \lim_{\rho\to1} R\mu\left(\frac{n\rho^{n+1} - (n+1)\rho^n + 1}{(\rho^{n+1}-1)^2}\right) + C\left(\frac{(n+1)(n\rho^{n+1}+1)}{(\rho^{n+1}-1)^2} - \frac{1}{(\rho-1)^2}\right) & \text{if } \rho = 1. \end{cases}$$

Showing that $f_n'(\rho) = 0$ has exactly one positive root directly is nontrival. However, it is equivalent to showing that $(1-\rho)^2 f_n'(\rho) = 0$ has exactly three positive roots. One can verify that this new term can be written explicitly as $(1-\rho)^2 f_n'(\rho) = R\mu\left(\frac{n\rho^{n+1}-(n+1)\rho^n+1}{(1+\rho+\cdots+\rho^n)^2}\right) + C\left(\frac{(n+1)(n\rho^{n+1}+1)}{(1+\rho+\cdots+\rho^n)^2}-1\right),\ \forall \rho \in \mathbb{R}_+$. We then reformulate the root equation to a polynomial form:

$$
\begin{aligned}
&R\mu\left(\frac{n\rho^{n+1}-(n+1)\rho^n+1}{(1+\rho+\cdots+\rho^n)^2}\right) + C\left(\frac{(n+1)(n\rho^{n+1}+1)}{(1+\rho+\cdots+\rho^n)^2}-1\right) = 0\\
&\Longleftrightarrow\\
&R\mu(n\rho^{n+1}-(n+1)\rho^n+1) + C(n+1)(n\rho^{n+1}+1) = C(1+\rho+\cdots+\rho^n)^2\\
&\Longleftrightarrow\\
&C(1+\rho+\cdots+\rho^n)^2 - n(R\mu+C(n+1))\rho^{n+1} + R\mu(n+1)\rho^n - R\mu - C(n+1) = 0.
\end{aligned}
$$

The left-hand side of the equation is a single-variable polynomial, and one can verify that it has three sign changes. Based on Descartes' rule of signs, the number of positive roots is at most three. By the fact that $f_n'(0) > 0$ and $\lim_{\rho\to\infty} f_n'(\rho) < 0$, $f_n'(\rho)$ must has at least one root. Because the term $(1-\rho)^2$ has two roots, we know that this polynomial has at least three roots. Therefore, this polynomial has exactly three roots, and $f_n'(\rho)$ has exactly one root. This shows that $f_n(\rho)$ is a unimodal function. □

Proof of Lemma 1, Statement (3). The second derivative of $f_n(\rho)$ is

$$
f_n''(\rho) = \begin{cases}
R\mu\dfrac{(n+1)\rho^{n-1}((n+1)(\rho-1)(\rho^{n+1}+1)-2\rho(\rho^{n+1}-1))}{(1-\rho^{n+1})^3} \\
\qquad + C\left(\dfrac{(n+1)^2\rho^n(2+n+n\rho^{n+1})}{(1-\rho^{n+1})^3} - \dfrac{2}{(1-\rho)^3}\right) & \text{if } \rho \neq 1\\
\lim\limits_{\rho\to 1} R\mu\dfrac{(n+1)\rho^{n-1}((n+1)(\rho-1)(\rho^{n+1}+1)-2\rho(\rho^{n+1}-1))}{(1-\rho^{n+1})^3} \\
\qquad + C\left(\dfrac{(n+1)^2\rho^n(2+n+n\rho^{n+1})}{(1-\rho^{n+1})^3} - \dfrac{2}{(1-\rho)^3}\right) & \text{if } \rho = 1.
\end{cases}
$$

Showing that $f_n''(\rho) = 0$ only has one root is equivalent to showing that $(1-\rho)^3 f_n(\rho) = 0$ has exactly four roots. One can check that $(1-\rho)^3 f_n''(\rho)$ coincides with $R\mu\frac{(n+1)\rho^{n-1}((n+1)(\rho-1)(\rho^{n+1}+1)-(\rho+1)(\rho^{n+1}-1))}{(1+\rho+\cdots+\rho^n)^3} + C\left(\frac{(n+1)^2\rho^n(2+n+n\rho^{n+1})}{(1+\rho+\cdots+\rho^n)^3}-2\right)$. Similar to the previous proof, we transform the root equation to a polynomial form:

$$
\begin{aligned}
&(1-\rho)^3 f_n''(\rho) = 0\\
&\Longleftrightarrow\\
&R\mu\frac{(n+1)\rho^{n-1}((n+1)(\rho-1)(\rho^{n+1}+1)-(\rho+1)(\rho^{n+1}-1))}{(1+\rho+\cdots+\rho^n)^3} + C\frac{(n+1)^2\rho^n(2+n+n\rho^{n+1})}{(1+\rho+\cdots+\rho^n)^3}\\
&\qquad -2C = 0\\
&\Longleftrightarrow\\
&2C(1+\rho+\cdots+\rho^n)^3 + (n+1)\rho^{n-1}(R\mu(n+1) - (4C(n+1)+R\mu(n+3))\rho - R\mu(n+1)\rho^2\\
&\ + R\mu(n+1)\rho^{n+1} + (R\mu(n-1)+C(n^2+n)\rho^{n+2}) = 0.
\end{aligned}
$$

One can verify that this polynomial has four sign changes. Based on Descartes' rule of signs, the number of positive roots is four or two. Because the term $(1-\rho)^3$ already has three roots, $f_n''(\rho)$ has exactly one root, which also implies that the sign of $f_n''(\rho)$ changes at most once. □

Proof of Lemma 2. We first show that strong duality holds, and both the primal and dual optimal solutions are attained, which is a sufficient condition for complementary slackness. To show this, we need to prove that both the primal and dual problems have interior points.

Showing the existence of interior points of the primal problem is equivalent to finding a point $(1, m, d)$ that resides in the interior of the convex cone

$$
\mathcal{V} = \left\{ (l,t,u) \in \mathbb{R}^3 : \exists \nu \in \mathcal{M}_+ \text{ such that } \begin{array}{l} \int_\Xi \nu(\mathrm{d}\rho) = l \\ \int_\Xi \rho\,\nu(\mathrm{d}\rho) = t \\ \int_\Xi |\rho - t|\,\nu(\mathrm{d}\rho) = u \end{array} \right\},
$$

where $\Xi = [a, b]$. Let $\mathbb{B}_\kappa(c)$ be the closed Euclidean ball of radius $\kappa \geq 0$ centered at c. We will show that $\mathbb{B}_\kappa(1) \times \mathbb{B}_\kappa(m) \times \mathbb{B}_\kappa(d)$ for a sufficiently small $\kappa > 0$. To this end, choose any point $(l_s, t_s, u_s) \in \mathbb{B}_\kappa(1) \times \mathbb{B}_\kappa(m) \times \mathbb{B}_\kappa(d)$, and consider the measure

$$\nu_s = \frac{n_s}{2(m_s - a)} \cdot \delta_a + \left(l_s - \frac{u_s}{2(t_s - a)} - \frac{u_s}{2(b - t_s)}\right) \cdot \delta_t + \frac{u_s}{2(b - t_s)} \cdot \delta_b,$$

where $s \cdot \delta_m$ denotes a measure that places mass s at m. By construction, this measure satisfies $\int_\rho \nu_s(\mathrm{d}\rho) = l_s$, $\int_\rho \rho\, \nu_s(\mathrm{d}\rho) = t_s$, and $\int_\rho |\rho - t_s| \nu_s(\mathrm{d}\rho) = u_s$ for a sufficiently small κ (because $m \in (a, b)$ and $d \in (0, \frac{2(m-a)(b-m)}{b-a})$).

Therefore, $(1, m, d)$ is an interior point of $\mathcal{V}$, and strong duality holds (i.e., the optimal values of the primal and dual problems coincide). Moreover, because there exists an interior point of the primal problem and because the common optimal value is finite, we have that the dual optimal solution is also attained (Shapiro 2001, proposition 3.4). Noticing that the support $[a, b]$ is compact, whereas the social benefit rate function $f_n(\rho)$ and the moment functions ρ and $|\rho - m|$ are continuous, we can invoke Shapiro (2001, corollary 3.1) to establish that the primal optimal solution is attained.

In summary, we have strong duality and the attainment of both the primal and dual optimal solutions, which imply that complementary slackness holds (Shapiro 2001, proposition 2.1). □

Proof of Lemma 4. We know that the revenue rate function is continuous for $\rho \in \mathbb{R}_+$. Therefore, employing Lemma B.2 yields the desired result. □

Proof of Proposition 3. The dual Problem (9) can be equivalently written as

$$\begin{aligned} \sup_{\alpha, \beta, \gamma \in \mathbb{R}} \quad & \mathbb{E}_{\mathbb{P}}[\alpha|\rho - m| + \beta\rho + \gamma] \\ \text{s.t.} \quad & \alpha|\rho - m| + \beta\rho + \gamma \leq f_n(\rho) \qquad \forall \rho \in [a, b], \end{aligned}$$

where $\mathbb{P} \in \mathcal{P}$ is an arbitrary probability measure in the ambiguity set. Observe that the left-hand side of the constraint is a two-piece piecewise affine function with a breakpoint at the mean m. Therefore, we can interpret the dual problem as finding a feasible two-piece piecewise affine function with the largest expected value. We now use this interpretation to derive the desired results.

First, we illustrate the case when $f_n(b) + f_n'(b)(m - b) \geq f_n(m)$. The constraint of the dual problem indicates that $f_n(\rho)$ majorizes $\alpha|\rho - m| + \beta\rho + \gamma$. One can verify that the two-piece piecewise affine function with the largest expected value is the one that touches $f_n(\rho)$ at three points: $\rho = a, m$, and b; see Figure 1(a) for an illustrative example. By complementary slackness in Lemma 2, the optimal distribution can only assign positive mass to these three points, which yields the following system of linear equations:

$$\begin{aligned} p_1(a - m) + p_2(m - m) + p_3(b - m) &= m \\ p_1|a - m| + p_2|m - m| + p_3|b - m| &= d \\ p_1 + p_2 + p_3 &= 1. \end{aligned}$$

Solving this system of linear equations leads to the first result in Proposition 3.

Next, we prove the two cases when $f_n(b) + f_n'(b)(m - b) \leq f_n(m)$. If $0 < d < d_0 := \frac{2(m-a)(\rho_t - m)}{\rho_t - a}$, we claim that the extremal distribution that solves (8) is a three-point distribution. To see this, we know that complementary slackness holds from Lemma 2, which means that the extremal distribution is supported on points where the dual constraint is binding. Because the two-piece piecewise affine function can touch $f_n(\rho)$ on at most three points under constraint

$$\alpha|\rho - m| + \beta\rho + \gamma \leq f_n(\rho) \qquad \forall \rho \in [a, b],$$

the extremal distribution is a one-point, two-point, or three-point distribution. We readily exclude the possibility that the extremal distribution is a one-point distribution because the mean-absolute deviation of a one-point distribution is zero. Next, we illustrate why the extremal distribution cannot be a two-point distribution. Suppose there exists a two-point distribution supported on $\{\rho_1, \rho_2\}$ that solves the worst-case expectation problem. Then, by complementary slackness, the dual constraint $f_n(\rho) = \alpha|\rho - m| + \beta\rho + \gamma$ will be binding at these two points. Without loss of generality, we assume $\rho_1 \in [a, m)$ and $\rho_2 \in (m, b]$. Because $f_n(\rho)$ is strictly concave for $\rho \in [a, m)$ and the dual constraint requires $\alpha|\rho - t| + \beta\rho + \gamma \leq f_n(\rho)$, we thus have $\rho_1 = a$. Because ρ_t is defined as the ρ coordinate of the point such that the line segment between $(m, f_n(m))$ and $(\rho_t, f_n(\rho_t))$ is tangent with $f_n(\rho)$, we must have $\rho_2 \geq \rho_t$; otherwise, the dual constraint will be violated. Because $\rho_2 - \rho_1 \geq \rho_t - a$, the corresponding mean-absolute deviation will be greater than d_0. Therefore, the extremal distribution cannot be a two-point distribution (i.e., it is a three-point distribution). Next, it can be shown that if $f_n(\rho)$ intersects $\alpha|\rho - m| + \beta\rho + \gamma$ at three points, then these three points must be $\rho = a, m$, and ρ_t. Therefore, we have the following system of linear equations:

$$\begin{aligned} p_1(a - m) + p_2(m - m) + p_3(\rho_t - m) &= m \\ p_1|a - m| + p_2|m - m| + p_3|\rho_t - m| &= d \\ p_1 + p_2 + p_3 &= 1. \end{aligned}$$

Solving this system of linear equations leads to the second result in Proposition 3.

We now establish that if $d_0 \leq d$, the extremal distribution is a two-point distribution. Similarly, by the fact that the extremal distribution is a discrete distribution supported on at most three points, we just need to show that there does not exist a one-point or three-point extremal distribution that solves (8). We can exclude the possibility of one-point distribution easily because its mean-absolute deviation is zero. As we described previously, the extremal three-point distribution is supported on $\rho = a, \rho_t$, and b, and the largest mean-absolute deviation that can be achieved within this support is given by $d_0 := \frac{2(m-a)(\rho_t - m)}{\rho_t - a}$. Because $d \geq d_0$, the extremal distribution can only be a two-point distribution. One of the support points is given by $\rho = a$, whereas the other one is determined by the value of d, which yields the following linear equations:

$$
\begin{aligned}
p_1(a-m) + p_2(\rho_2 - m) &= m \\
p_1|a-m| + p_2|\rho_2 - m| &= d \\
p_1 + p_2 &= 1.
\end{aligned}
\tag{B.2}
$$

Solving this system of equations, we obtain the optimal solution explicitly as

$$
p_1 = \frac{d}{2(m-a)},\ \rho_1 = a;\ p_2 = 1 - \frac{d}{2(m-a)},\ \rho_2 = \frac{da + 2m(a-m)}{d + 2(a-m)}.
$$

This completes the proof. □

Appendix C. Proofs of Section 4

Proof of Theorem 4. Problem (14) can be equivalently written as

$$
\begin{aligned}
\inf_{\nu \in \mathcal{M}_+} \quad & \int_\Xi f_n(\rho)\nu(\mathrm{d}\rho) \\
\text{s.t.} \quad & \int_\Xi |\rho - \hat{m}|\, \nu(\mathrm{d}\rho) = d \\
& \int_\Xi \rho\, \nu(\mathrm{d}\rho) = m \\
& \int_\Xi \nu(\mathrm{d}\rho) = 1 \\
& m_l \leq m \leq m_u \\
& d_l \leq d \leq d_u.
\end{aligned}
$$

Dualizing this optimization problem yields

$$
\begin{aligned}
\sup_{\theta \in \mathbb{R}_+^4, \gamma \in \mathbb{R}} \quad & \gamma + \theta_1 d_l - \theta_2 d_u + \theta_3 m_l - \theta_4 m_u \\
\text{s.t.} \quad & (\theta_1 - \theta_2)|\rho - \hat{m}| + (\theta_3 - \theta_4)\rho + \gamma \leq f_n(\rho) \qquad \forall \rho \in [a, b].
\end{aligned}
$$

Applying algebraic reductions and invoking Lemma 3 lead to the desired reformulation. The derivation straightforwardly follows that of Theorem 1, and we omit for brevity. □

Proof of Theorem 5. The dual problem is given by

$$
\begin{aligned}
\sup_{\theta \in \mathbb{R}_+^4, \gamma \in \mathbb{R}} \quad & \gamma + \theta_1 d_l - \theta_2 d_u + \theta_3 m_l - \theta_4 m_u \\
\text{s.t.} \quad & (\theta_1 - \theta_2)|\rho - \hat{m}| + (\theta_3 - \theta_4)\rho + \gamma \leq r_n(\rho) \qquad \forall \rho \in [a, b].
\end{aligned}
$$

Because the revenue rate function $r_n(\rho)$ is concave for $\rho \geq 0$, the semi-infinite constraints are satisfied if and only if each constraint is satisfied at points $\rho = a, \hat{\rho}_i, b$, which completes the proof. □

Appendix D. Distributionally Robust Model with a Wasserstein Ambiguity Set

In this section, we study the DRO model with a Wasserstein ambiguity set (Esfahani and Kuhn 2018, Gao and Kleywegt 2023). We develop solution schemes to find the optimal threshold strategies for a social optimizer and a revenue maximizer given by $\hat{n}_s$ and $\hat{n}_r$, respectively, such that the worst-case expected benefit rates are maximized. Here, the worst case is taken over the Wasserstein ambiguity set containing all probability distributions (discrete or continuous) sufficiently close to the discrete empirical distribution, where the closeness between two distributions is measured in terms of the Wasserstein metric (Esfahani et al. 2018).

Definition D.1 (Wasserstein Metric). For any $r \geq 1$, let $\mathcal{M}^r(\Xi)$ be the set of all probability distributions $\mathbb{P}$ supported on Ξ satisfying $\mathbb{E}_{\mathbb{P}}[\|\boldsymbol{\xi}\|^r] = \int_\Xi \|\boldsymbol{\xi}\|^r \mathbb{P}(\mathrm{d}\boldsymbol{\xi}) < \infty$. The r-Wasserstein distance between two distributions $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}_0^r(\Xi)$ is defined as

$$
\mathcal{W}^r(\mathbb{P}_1, \mathbb{P}_2) = \inf\left\{ \left(\int_{\Xi^2} \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|^r \mathbb{Q}(\mathrm{d}\boldsymbol{\xi}_1, \mathrm{d}\boldsymbol{\xi}_2) \right)^{\frac{1}{r}} \right\},
$$

where $\mathbb{Q}$ is a joint distribution of $\tilde{\boldsymbol{\xi}}_1$ and $\tilde{\boldsymbol{\xi}}_2$ with marginals $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively.

The Wasserstein distance $\mathcal{W}^r(\mathbb{P}_1, \mathbb{P}_2)$ can be viewed as the (rth root of the) minimum cost for moving the distribution $\mathbb{P}_1$ to $\mathbb{P}_2$, where the cost of moving a unit mass from $\boldsymbol{\xi_1}$ to $\boldsymbol{\xi_2}$ amounts to $\|\boldsymbol{\xi_1} - \boldsymbol{\xi_2}\|^r$. The joint distribution $\mathbb{Q}$ of $\tilde{\boldsymbol{\xi}}_1$ and $\tilde{\boldsymbol{\xi}}_2$ is, therefore, naturally interpreted as a mass transportation plan (Esfahani et al. 2018). Similarly to the data-driven setting in Section 4, we assume that we have observed a finite set of N independent realizations given by $\{\hat{\rho}_i\}_{i\in[N]}$, where $\hat{\rho}_i = \hat{\lambda}_i/\mu$. Using the observations, we define the empirical distribution $\hat{\mathbb{P}}_N := \frac{1}{N}\sum_{i\in[N]}\delta_{\hat{\rho}_i}$ as the discrete uniform distribution on the samples.

In this paper, we consider the Wasserstein ambiguity set defined as

$$\mathcal{B}_\epsilon(\hat{\mathbb{P}}_N) := \{\mathbb{P} \in \mathcal{P}_0(\Xi) : \mathcal{W}^1(\mathbb{P}, \hat{\mathbb{P}}_N) \le \epsilon\}, \tag{D.1}$$

which is a neighborhood around the empirical distribution. The ambiguity set contains all distributions supported on Ξ that are of type 1 Wasserstein distance less than or equal to ϵ from $\hat{\mathbb{P}}_N$. By adjusting the radius ϵ of the ball, one can control the degree of conservatism of the DRO model. If $\epsilon = 0$, the Wasserstein ball shrinks to a singleton set containing only the empirical distribution $\hat{\mathbb{P}}_N$. One can further show that this data-driven DRO model converges to the corresponding true stochastic program as the sample size N tends to infinity (Esfahani and Kuhn 2018).

We derive the optimal threshold strategies $\hat{n}_s$ and $\hat{n}_r$ for a social optimizer and a revenue maximizer, respectively. As stated in Section 2, the optimal joining threshold $\tilde{n}_e$ for an individual customer is independent of the arrival rate, and we have $\tilde{n}_e = n_e$ from (2).

D.1. Social Optimizer

The objective of a social optimizer is to obtain an optimal joining threshold $\hat{n}_s$ that maximizes the worst-case expected benefit: that is, $\hat{n}_s \in \arg\max_{n\in\mathbb{Z}_+}\{Z_s(n)\}$, where

$$Z_s(n) := \inf_{\mathbb{P}\in\mathcal{B}_\epsilon(\hat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}}[f_n(\tilde{\rho})]. \tag{D.2}$$

The worst-case expectation is computed over all distributions in the Wasserstein ambiguity set $\mathcal{B}_\epsilon(\hat{\mathbb{P}}_N)$ with the support set $\Xi = [a, b]$.

Theorem D.1. *For any $n \ge 1$ and $\mathcal{P} = \mathcal{B}_\epsilon(\hat{\mathbb{P}}_N)$, the worst-case expectation $Z_s(n)$ coincides with the optimal objective value of the following semidefinite program*:

$$\begin{array}{lll}
sup & -\alpha\epsilon + \dfrac{1}{N}\displaystyle\sum_{i\in[N]} s_i & \\
s.t. & \alpha \in \mathbb{R}_+, s \in \mathbb{R}^N \boldsymbol{y^i}, \boldsymbol{z^i} \in \mathbb{R}^{n+3}, \boldsymbol{X^i}, \boldsymbol{W^i} \in \mathbb{S}_+^{n+3} & \forall i \in [N] \\
& y_0^i = -s_i + \alpha\hat{\rho}_i,\ y_1^i = s_i - \alpha - \alpha\hat{\rho}_i + R\mu - C, y_2^i = \alpha - R\mu & \\
& y_3^i, \ldots, y_n^i = 0,\ y_{n+1}^i = -s_i - \alpha\hat{\rho}_i - R\mu + C(n+1), & \\
& y_{n+2}^i = -s_i + \alpha + \alpha\hat{\rho}_i + R\mu + C - C(n+1),\ y_{n+3}^i = -\alpha & \forall i \in [N] \\
& z_0^i = -s_i - \alpha\hat{\rho}_i,\ z_1^i = s_i + \alpha + \alpha\hat{\rho}_i + R\mu - C, z_2^i = -\alpha - R\mu & \\
& z_3^i, \ldots, z_n^i = 0,\ z_{n+1}^i = -s_i + \alpha\hat{\rho}_i - R\mu + C(n+1), & \\
& z_{n+2}^i = -s_i - \alpha - \alpha\hat{\rho}_i + R\mu + C - C(n+1),\ z_{n+3}^i = \alpha & \forall i \in [N] \\
& \displaystyle\sum_{u+v=2l-1} x_{uv}^i = 0 & \forall l \in [n+3]\, i \in [N] \\
& \displaystyle\sum_{q=0}^{l}\sum_{r=q}^{n+3+q-l} y_r^i \binom{r}{q}\binom{n+3-r}{l-q} a^{r-q}\hat{\rho}_i^q = \sum_{u+v=2l} x_{uv}^i & \forall l \in [n+3] \cup \{0\}\, i \in [N] \\
& \displaystyle\sum_{u+v=2l-1} w_{uv}^i = 0 & \forall l \in [n+3]\, i \in [N] \\
& \displaystyle\sum_{q=0}^{l}\sum_{r=q}^{n+3+q-l} z_r^i \binom{r}{q}\binom{n+3-r}{l-q} \hat{\rho}_i^{r-q} b^q = \sum_{u+v=2l} w_{uv}^i & \forall l \in [n+3] \cup \{0\}\, i \in [N].
\end{array}$$

Proof. The distributionally robust model with the ambiguity set (D.1) can be equivalently written as

$$\begin{array}{ll}
\inf & \dfrac{1}{N}\displaystyle\sum_{i\in[N]}\int_\Xi f_n(\rho)\mathbb{P}_i(\mathrm{d}\rho) \\
\text{s.t.} & \mathbb{P}_i \in \mathcal{P}_0(\Xi) \quad \forall i \in [N] \\
& \dfrac{1}{N}\displaystyle\sum_{i\in[N]}\int_\Xi \|\rho - \hat{\rho}_i\|\mathbb{P}_i(\mathrm{d}\rho) \le \epsilon.
\end{array}$$

Its strong dual problem is given by Esfahani and Kuhn (2018, theorem 4.2):

$$\begin{aligned}
\sup_{\alpha\in\mathbb{R}_+,\, s\in\mathbb{R}^N} \quad & -\alpha\epsilon + \frac{1}{N}\sum_{i\in[N]} s_i \\
\text{s.t.} \quad & s_i - \alpha\|\rho - \hat{\rho}_i\| \le f_n(\rho) \qquad \forall i\in[N]\ \ \forall \rho\in[a,b].
\end{aligned}$$

We can deal with each constraint separately for the cases $\rho \le \hat{\rho}_i$ and $\rho \ge \hat{\rho}_i$, and consequently, we have

$$\begin{aligned}
\sup_{\alpha\in\mathbb{R}_+,\, s\in\mathbb{R}^N} \quad & -\alpha\epsilon + \frac{1}{N}\sum_{i\in[N]} s_i \\
\text{s.t.} \quad & s_i + \alpha(\rho - \hat{\rho}_i) \le f_n(\rho) \qquad \forall i\in[N]\ \ \forall \rho\in[a,\hat{\rho}_i] \\
& s_i - \alpha(\rho - \hat{\rho}_i) \le f_n(\rho) \qquad \forall i\in[N]\ \ \forall \rho\in[\hat{\rho}_i,b].
\end{aligned}$$

Substituting the definition of $f_n(\rho)$ in (3) and applying algebraic reductions yield the following polynomial inequalities for each $i\in[N]$:

$$\begin{aligned}
&(-s_i + \alpha\hat{\rho}_i)\rho^0 + (s_i - \alpha - \alpha\hat{\rho}_i + R\mu - C)\rho + (\alpha - R\mu)\rho^2 + (s_i - \alpha\hat{\rho}_i - R\mu + C(n+1))\rho^{n+1} \\
&+ (-s_i + \alpha + \alpha\hat{\rho}_i + R\mu + C - C(n+1))\rho^{n+2} - \alpha\rho^{n+3} \ge 0 \quad \forall\rho\in[a,\hat{\rho}_i], \\
&(-s_i - \alpha\hat{\rho}_i)\rho^0 + (s_i + \alpha + \alpha\hat{\rho}_i + R\mu - C)\rho + (-\alpha - R\mu)\rho^2 + (s_i + \alpha\hat{\rho}_i - R\mu + C(n+1))\rho^{n+1} \\
&+ (-s_i - \alpha - \alpha\hat{\rho}_i + R\mu + C - C(n+1))\rho^{n+2} + \alpha\rho^{n+3} \ge 0 \quad \forall\rho\in[\hat{\rho}_i,b].
\end{aligned} \tag{D.3}$$

The inequalities are of the form $g_1^i(\rho) = \sum_{r=0}^{n+3} y_r^i\rho^r \ge 0$ for $\rho\in[a,\hat{\rho}_i]$ and $g_2^i(\rho) = \sum_{r=0}^{n+3} z_r^i\rho^r \ge 0$ for $\rho\in[\hat{\rho}_i,b]$, where y^i and z^i represent the coefficients of the respective polynomial inequalities. We next invoke the result of Lemma 3 for every $i\in[N]$ to express the inequalities in (D.3) as semidefinite constraints. This leads to the desired semidefinite program, which completes the proof. □

To determine an optimal joining threshold, we compute the worst-case expected benefit rate $Z_s(n)$ for every $n\in\mathbb{Z}_+$, $1 \le n \le n_e$, using the result of Theorem D.1, and then, we select the best threshold $\hat{n}_s \in \arg\max_{n\in\mathbb{Z}_+}\{Z_s(n)\}$.

D.2. Revenue Maximizer

The objective of a revenue maximizer is to find an optimal threshold $\hat{n}_r$ that maximizes the worst-case expected revenue rate of a firm (i.e., $\hat{n}_r \in \arg\max_{n\in\mathbb{Z}_+}\{Z_r(n)\}$, where the worst-case expectation is computed over all the distributions in the Wasserstein ambiguity set $\mathcal{B}_\epsilon(\hat{\mathbb{P}}_N)$ defined by (D.1) with support set $\Xi = [a.b]$). The worst-case expected profit rate $Z_r(n)$ is given by

$$Z_r(n) := \inf_{\mathcal{P}\in\mathcal{B}_\epsilon(\hat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}}[r_n(\tilde{\rho})]. \tag{D.4}$$

Theorem D.2. *For any $n \ge 1$, the worst-case expectation $Z_r(n)$ coincides with the optimal objective value of the following linear program:*

$$\begin{aligned}
\mathit{sup}_{\alpha\in\mathbb{R}_+,\, s\in\mathbb{R}^N} \quad & -\alpha\epsilon + \frac{1}{N}\sum_{i\in[N]} s_i \\
\mathit{s.t.} \quad & s_i + \alpha(a - \hat{\rho}_i) \le r_n(a) \quad \forall i\in[N] \\
& s_i \le r_n(\hat{\rho}_i) \quad \forall i\in[N] \\
& s_i - \alpha(b - \hat{\rho}_i) \le r_n(b) \quad \forall i\in[N].
\end{aligned}$$

Proof. The strong dual problem of $\inf_{\mathbb{P}\in\mathcal{B}_\epsilon(\hat{\mathbb{P}}_N)}\mathbb{E}_{\mathbb{P}}[r_n(\tilde{\rho})]$ is given by

$$\begin{aligned}
\sup_{\alpha\in\mathbb{R}_+,\, s\in\mathbb{R}^N} \quad & -\alpha\epsilon + \frac{1}{N}\sum_{i\in[N]} s_i \\
\text{s.t.} \quad & s_i - \alpha\|\rho - \hat{\rho}_i\| \le r_n(\rho) \qquad \forall i\in[N]\ \ \forall \rho\in[a,b].
\end{aligned}$$

Because the revenue rate function $r_n(\rho)$ is concave for $\rho \ge 0$, the semi-infinite constraints are satisfied if and only if each constraint is satisfied at three points $\rho = a, \hat{\rho}_i, b$. Consequently, we have

$$\begin{aligned}
Z_r(n) := \sup_{\alpha\in\mathbb{R}_+,\, s\in\mathbb{R}^N} \quad & -\alpha\epsilon + \frac{1}{N}\sum_{i\in[N]} s_i \\
\text{s.t.} \quad & s_i + \alpha(a - \hat{\rho}_i) \le r_n(a) \qquad \forall i\in[N] \\
& s_i + \alpha(\hat{\rho}_i - \hat{\rho}_i) \le r_n(\hat{\rho}_i) \qquad \forall i\in[N] \\
& s_i - \alpha(b - \hat{\rho}_i) \le r_n(b) \qquad \forall i\in[N],
\end{aligned}$$

and thus, the claim follows. □

To determine an optimal joining threshold $\hat{n}_r$, we compute the worst-case expected profit rate $Z_r(n)$ for every $n \in \mathbb{Z}_+$, $1 \le n \le \hat{n}_e$ using the result of Theorem D.2, and we select $\hat{n}_r \in \arg\max_{n \in \mathbb{Z}_+}\{Z_r(n)\}$.

References

Abu-Shawiesh MOA, Banik S, Kibria B (2018) Confidence intervals based on absolute deviation for population mean of a positively skewed distribution. *Internat. J. Comput. Theoret. Statist.* 5(1):1–13.

Afèche P, Ata B (2013) Bayesian dynamic pricing in queueing systems with unknown delay cost characteristics. *Manufacturing Service Oper. Management* 15(2):292–304.

ApS (2022) MOSEK optimizer API for Python Version 9.2.49. Accessed April 13, 2022, https://docs.mosek.com/9.2/pythonapi/index.html.

Arachchige CN, Prendergast LA (2019) Confidence intervals for median absolute deviations. Preprint, submitted November 1, https://arxiv.org/abs/1910.00229.

Ardestani-Jaafari A, Delage E (2021) Linearized robust counterparts of two-stage robust optimization problems with applications in operations management. *INFORM J. Comput.* 33(3):1138–1161.

Bandi C, Bertsimas D, Youssef N (2015) Robust queueing theory. *Oper. Res.* 63(3):676–700.

Bates S, Hastie T, Tibshirani R (2023) Cross-validation: What does it estimate and how well does it do it? *J. Amer. Statist. Assoc.*, ePub ahead of print May 15, https://doi.org/10.1080/01621459.2023.2197686.

Ben-Tal A, Hochman E (1972) More bounds on the expectation of a convex function of a random variable. *J. Appl. Probab.* 9(4):803–812.

Ben-Tal A, El Ghaoui L, Nemirovski A (2009) *Robust Optimization*, vol. 28 (Princeton University Press, Princeton, NJ).

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

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

Bhat UN (2008) The general queue G/G/1 and approximations. *An Introduction to Queueing Theory: Modeling and Analysis in Applications* (Birkhäuser, Boston), 169–183.

Bonett DG, Seier E (2003) Confidence intervals for mean absolute deviations. *Amer. Statist.* 57(4):233–236.

Burnetas A, Economou A (2007) Equilibrium customer strategies in a single server Markovian queue with setup times. *Queueing Systems* 56(3–4):213–228.

Chen Y, Hasenbein JJ (2020) Knowledge, congestion, and economics: Parameter uncertainty in Naor's model. *Queueing Systems* 96(1):83–99.

Debo L, Veeraraghavan S (2014) Equilibrium in queues under unknown service times and service value. *Oper. Res.* 62(1):38–57.

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

Economou A, Kanta S (2008) Equilibrium balking strategies in the observable single-server queue with breakdowns and repairs. *Oper. Res. Lett.* 36(6):696–699.

Esfahani PM, Kuhn D (2018) Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Programming* 171(2018):115–166.

Esfahani PM, Shafieezadeh-Abadeh S, Hanasusanto GA, Kuhn D (2018) Data-driven inverse optimization with imperfect information. *Math. Programming* 167(1):191–234.

Gao R, Kleywegt A (2023) Distributionally robust stochastic optimization with Wasserstein distance. *Math. Oper. Res.* 48(2):603–655.

Gotoh Jy, Kim MJ, Lim AE (2021) Calibration of distributionally robust empirical optimization models. *Oper. Res.* 69(5):1630–1650.

Guo P, Hassin R (2011) Strategic behavior and social optimization in Markovian vacation queues. *Oper. Res.* 59(4):986–997.

Guo P, Zipkin P (2007) Analysis and comparison of queues with different levels of delay information. *Management Sci.* 53(6):962–970.

Hanasusanto GA, Kuhn D, Wallace SW, Zymler S (2015) Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Math. Programming* 152(1–2):1–32.

Hassin R, Haviv M (2003) *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*, vol. 59 (Springer Science & Business Media, New York).

Hassin R, Haviv M, Oz B (2023) Strategic behavior in queues with arrival rate uncertainty. *Eur. J. Oper. Res.* 309(1):217–224.

Haviv M, Oz B (2016) Regulating an observable M/M/1 queue. *Oper. Res. Lett.* 44(2):196–198.

Herrey EM (1965) Confidence intervals based on the mean absolute deviation of a normal sample. *J. Amer. Statist. Assoc.* 60(309):257–269.

Li X, Natarajan K, Teo CP, Zheng Z (2014) Distributionally robust mixed integer linear programs: Persistency models with applications. *Eur. J. Oper. Res.* 233(3):459–473.

Liu C, Hasenbein JJ (2019) Naor's model with heterogeneous customers and arrival rate uncertainty. *Oper. Res. Lett.* 47(6):594–600.

Lofberg J (2004) Yalmip: A toolbox for modeling and optimization in MATLAB. *2004 IEEE Internat. Conf. Robotics Automation IEEE Catalog Number 04CH37508* (IEEE, Piscataway, NJ), 284–289.

Naor P (1969) The regulation of queue size by levying tolls. *Econometrica* 37(1):15–24.

Pflug G, Wozabal D (2007) Ambiguity in portfolio selection. *Quant. Finance* 7(4):435–442.

Postek K, Ben-Tal A, Den Hertog D, Melenberg B (2018) Robust optimization with ambiguous stochastic constraints under mean and dispersion information. *Oper. Res.* 66(3):814–833.

Scarf HE (1957) A min-max solution of an inventory problem. Technical report, RAND Corporation, Santa Monica, CA.

Shafieezadeh-Abadeh S, Esfahani PM, Kuhn D (2015) Distributionally robust logistic regression. *Adv. Neural Inform. Processing Systems* 1(2015):1576–1584.

Shapiro A (2001) On duality theory of conic linear problems. Goberna MÁ, López MA, eds. *Semi-Infinite Programming*, Nonconvex Optimization and Its Applications, vol. 57 (Springer, Boston), 135–165.

Shapiro A, Kleywegt A (2002) Minimax analysis of stochastic problems. *Optim. Methods Software* 17(3):523–542.

Toh KC, Todd MJ, Tütüncü RH (1999) SDPT3—A MATLAB software package for semidefinite programming, version 1.3. *Optim. Methods Software* 11(1–4):545–581.

van Eekelen W, den Hertog D, van Leeuwaarden JS (2022) MAD dispersion measure makes extremal queue analysis simple. *INFORMS J. Comput.* 34(3):1681–1692.

Wiesemann W, Kuhn D, Sim M (2014) Distributionally robust convex optimization. *Oper. Res.* 62(6):1358–1376.

Žáčková J (1966) On minimax solutions of stochastic linear programming problems. *Časopis Pro Pěstování Matematiky* 91(4):423–430.