Vessel Deployment with Limited Information: Distributionally Robust Chance Constrained Models

Yue Zhao[1], Zhi Chen[2], Andrew Lim[1], Zhenzhen Zhang[3]
[1]Institute of Operations Research and Analytics, National University of Singapore, Singapore
[2]Department of Management Sciences, City University of Hong Kong, Hong Kong SAR, China
[3]School of Economics and Management, Tongji University, Shanghai, China

This paper studies the fundamental vessel deployment problem in the liner shipping industry, which decides the numbers of mixed-type ships and their sailing frequencies on fixed routes to provide sufficient vessel capacity for fulfilling stochastic shipping demands with high probability. In reality, it is usually difficult (if not impossible) to acquire a precise joint distribution of shipping demands, as they may fluctuate heavily due to the fast-changing economic environment or unpredictable events. To address this challenge, we leverage recent advances in distributionally robust optimization and propose distribution-free robust joint chance constrained models. In the first model, we only assume support, mean as well as lower-order dispersion information of the shipping demands and provide high-quality solutions via a sequential convex optimization algorithm. Comparing with existing literature that chiefly studies individual chance constraints based on concentration inequalities and the union bound, our approach yields solutions that are less conservative and less vulnerable to the magnitude of demand dispersion. We also extend to a data-driven model based on the Wasserstein distance, which suits well in situations where limited historical demand samples are available. Our distributionally robust chance constrained models could serve as a baseline model for vessel deployment, into which we believe additional practical constraints could be incorporated seamlessly.

Key words: maritime transportation, fleet deployment, liner shipping, distributionally robust optimization, joint chance constraints, data-driven optimization.
History: May 16, 2022

1. Introduction

International maritime trade, maintaining an average growth rate of 3% for the past few decades, accounts for more than 80% volume of the global merchandise trade in 2018 by far (Sirimanne et al. 2019). The liner container shipping sector plays a critical role in the shipping industry. A company in this sector transports cargo regularly according to pre-scheduled routes within a fixed planning horizon. In order to minimize the cost and gain the maximal profit, the numbers of mixed-type ships deployed in each route and their sailing frequencies have to be decided carefully. In the decision process, several practical constraints need to be met, among which to satisfy the shipping demand is central due to its inherent and significant uncertainty in real-world scenarios.

Pioneering works of this problem date back to the early 1990s by Jaramillo and Perakis (1991), Perakis and Jaramillo (1991), who propose mixed-integer linear program formulations for the problem. However, most existing studies assume that shipping demands are deterministic until a recent study of Meng and Wang (2010), which proposes chance constrained models with uncertain demands that follow known distributions. Considering independent normal distributions with given means and standard deviations, the *individual* chance constraint for each route can be reformulated as a deterministic constraint, and the resulting mixed-integer linear programming formulation can be directly solved by commercial solvers. Later, Meng et al. (2012) investigate the joint optimizationss of vessel deployment and container transshipment with normally distributed shipping demands, and develop a two-stage stochastic mixed-integer model that can be solved by the sample average approximation (SAA) approach. From a methodological point of view, when dealing with chance constraints where the uncertain demands are governed by a known probability distribution, one could also adopt nonlinear programming approaches (see, *e.g.*, Prékopa 2003, van Ackooij 2020 and references therein), which usually provide high-quality approximations of the probability function and efficient exact algorithms under Gaussian-like assumptions.

However, the true distribution is hard to obtain while a parametric distribution with prior knowledge based on estimation could be largely biased due to the fast-changing shipping industry and global economic landscape. To overcome this issue, a distribution-free model is proposed by Ng (2014) with information of only mean, variance and upper bound on the demand. Although this model has several interesting merits, such as leading to a mixed-integer linear program reformulation and providing interesting interpretations, it does not capture potential stochastic dependencies of demands across various routes. Therefore, Ng (2015) further relaxes the assumption on upper bounds of the demand and considers the *joint* chance constraint, which, by applying the union bound, can be conservatively approximated and reformulated as a collection of linear constraints. With the concern of being too conservative sometimes, Ng and Lin (2018) design a new model that bypasses the chance constraint and assumes known conditional demands based on partition of the sample space. For more details and outlook of these related problems, we refer to the reviews of Meng et al. (2014) as well as Wang and Meng (2017).

In this paper we focus on distributionally robust chance constrained models. Different from the works of Ng (2014, 2015), our approach does not rely on any probability inequalities that might be conservative when the given probability threshold or variance is large, or when considering stochastic dependencies (Xie et al. 2019). Our model is based on the general notion of robust optimization (Ben-Tal et al. 2009), and more precisely distributionally robust optimization with ambiguous chance constraints, see, *e.g.*, Postek et al. (2018). Distributionally robust optimization is a powerful tool to address problems with uncertainty, which does not require full knowledge of the distributional information and is usually computational tractable. The key idea of distributionally robust optimization is to specify an ambiguity set that contains all possible distributions sharing the given characteristics (such as support and moment conditions) and to find optimal decisions that perform the best in view of the worst-case distribution that may arise from the ambiguity set. Equipped with recent tools in distributionally robust optimization by Hanasusanto et al. (2015, 2017), Chen et al. (2018) and Xie (2019), we develop distributionally robust chance constrained models for the liner container shipping problem in the perspective of joint chance constraints. We refer interested readers to Chen et al. (2018) and references therein for a comprehensive review on distributionally robust chance constrained programs.

In practice, it is typically difficult, if not impossible, to obtain the precise description of the demand distribution, while either partial statistical information or empirical data is often accessible. Therefore, we propose to adopt two different kinds of distributionally robust joint chance constrained models, and we numerically compare their performance in the experiments. For the first type, we consider an ambiguity set with only mean and dispersion information. Although given the considered ambiguity set, it is generally intractable to solve the distributionally robust chance constrained model (Hanasusanto et al. 2015), we leverage the conic program reformulation of the chance constraint and follow a sequential convex optimization scheme to iteratively solve the whole problem. For the second type, we extend (in Appendix B) the vessel deployment problem to the *data-driven* setting where we study a distributionally robust chance constrained model based on the Wasserstein ambiguity set. The data-driven approach covers the SAA approach that considers the empirical distribution as a special case.

In the numerical studies, we show that (*i*) under relatively small risk threshold and large variance, our solution is more costly (within 5% of the existing best solution) while performs consistently and significantly better in the out-of-sample stress test; (*ii*) for small risk threshold and large variance, our solution is of the desirable amount while the existing models may be infeasible; (*iii*) our models better meet the joint chance constraint with much less extra cost than that of existing models; (*iv*) in the data-driven setting, the distributionally robust models could provide stable solutions and outperform the existing models in the out-of-sample stress test, and especially, much better than the SAA approach. Apart from these numerical observations, we would also like to share several managerial insights with decision-makers in the maritime industry. (*i*) Compared with meeting the demands in different routes individually, satisfying the demands jointly might be more practically relevant though is typically more computationally difficult to tackle. A loose approximation of joint satisfactory, such as the existing approach in the literature that is based on union bound, may result in a conservative decision with extra operational cost (see the illustrative Example 1 as well as Tables 4 and 5 for the comparison of solutions). (*ii*) The decision obtained from joint chance constrained models tends to be more robust against demand deviations, and this is also observed even when looking at the average individual probability of satisfactory (see, *e.g.*, Figure 2 for the situation under different demand deviations). (*iii*) If the decision-makers only use the SAA approach with historical data to model the problem, although the optimal cost will be lower than that of the robust approach, the out-of-sample performance of the SAA approach could be vulnerable to demand deviations and for many routes, it may not even meet the demands (see more details in Table 7 and Figure 7 for our data-driven experiments).

The remainder of this paper is organized as follows. In Section 2 we state the stochastic and distributionally robust vessel deployment models and point out that the stochastic model may be a loose approximation in view of distributionally robust joint chance constrained optimization. We then propose our approach that directly tackles the problem without any approximation in modeling. In Section 3, we formally introduce the mean and dispersion ambiguity set that captures a various distributional information and derive the corresponding equivalent reformulation. To showcase the practical usage of our model, we illustrate several concrete examples with the explanation of the parameters in the ambiguity set. A sequential convex optimization algorithm is then introduced to efficiently obtain practical and high-quality solutions of our reformulation. In the numerical study of Section 4, we show the merits of our models when dealing with high variance and low-risk thresholds. We also conduct perturbation analysis and out-of-sample stress tests under several commonly-used distributions. We extend our framework to the data-driven setting in Appendix B, and in Appendix C, we run extensive experiments to show the advantages of distributionally robust chance constrained models in the data-driven setting.

Notation. We use boldface uppercase and lowercase letters to denote matrices and vectors, respectively. Special vectors of appropriate dimensions include $\mathbf{0}$, $\boldsymbol{e}$, $\boldsymbol{e}_i$, which respectively correspond to the zero vector, the vector of all ones, and the i^{th} standard basis. Given $N \in \mathbb{Z}_+$, we use the shorthand $[N] = \{1, 2, \ldots, N\}$ to represent the set of all integers up to N. The dual norm of a general norm $\|\cdot\|$ is denoted by $\|\cdot\|_*$. Random variables are denoted by tilde signs, while their realizations are denoted by the same symbols without tildes. We use $\mathcal{P}(\mathbb{R}^K)$ to denote the set of probability distributions supported on $\mathbb{R}^K$. The indicator function $\mathbb{I}(\mathcal{E})$ takes a value of one if $\mathcal{E}$ is true and takes a value of zero otherwise.

2. Vessel Deployment under Uncertainty

We focus on the fundamental vessel deployment problem in Ng (2015) and Meng and Wang (2010), which can be potentially extended to incorporate some other constraints arising from practical considerations, operating patterns, as well as emerging phenomena and technologies. The problem asks for regularly transports cargo on the pre-scheduled routes, and the demand on each route

Sets	
$[S]$	set of ship types
$[K]$	set of routes
Parameters	
c_{sk}	cost of voyage for a ship of type $s \in [S]$ on route $k \in [K]$
h_s	cost of chartering in a ship of type $s \in [S]$
r_s	revenue of chartering out a ship of type $s \in [S]$
a_s	number of ships of type $s \in [S]$ available in the liner company's own fleet
m_s	maximum number of ships of type $s \in [S]$ that can be chartered from other ship owners
q_s	capacity of a ship of type $s \in [S]$ (in TEUs, *i.e.*, twenty foot equivalent unit)
n_k	number of voyages needed on route $k \in [K]$ to maintain the minimum sailing frequency
t_{sk}	transit time of a ship of type $s \in [S]$ to traverse route $k \in [K]$ (in days)
p	planning horizon under consideration (in days)
ε	risk threshold of not meeting customer demand
Random Variables	
$\tilde{d}_k$	maximal shipping demand among all legs of the voyage on route $k \in [K]$
Decision Variables	
u_{sk}	total number of ships of type $s \in [S]$ to be deployed on route $k \in [K]$
v_s	number of ships of type $s \in [S]$ to be chartered from other ship owners
w_s	number of ships of type $s \in [S]$ to be chartered out
x_{sk}	number of voyages that ships of type $s \in [S]$ have completed on route $k \in [K]$

Table 1 Variables description.

is uncertain. Throughout the planning horizon, each ship can perform multiple voyages on the route, which is different for heterogeneous ships. To respect the sailing frequency and to satisfy the demand with high probability on each route, sufficient number of mixed-type ships should be deployed. The objective is to minimize the total cost with a feasible deployment plan by either chartering in or out the ships regarding to the number of available ships. Detailed descriptions of notations and variables are summarized in Table 1.

2.1. Stochastic Chance Constrained Model

The basic stochastic formulation of the vessel deployment problem, originating from Meng and Wang (2010) and Ng (2015), is given as follows:

$$\min \sum_{s\in[S]} \sum_{k\in[K]} c_{sk} x_{sk} + \sum_{s\in[S]} h_s v_s - \sum_{s\in[S]} r_s w_s \tag{1a}$$

$$\text{s.t.}\;\; \mathbb{P}\left[\tilde{d}_k \leq \sum_{s\in[S]} q_s x_{sk} \quad \forall k\in[K]\right] \geq 1-\varepsilon \tag{1b}$$

$$\sum_{s\in[S]} x_{sk} \geq n_k \qquad \forall k\in[K] \tag{1c}$$

$$\sum_{k\in[K]} u_{sk} \leq a_s + v_s \qquad \forall s\in[S] \tag{1d}$$

$$v_s \leq m_s \qquad \forall s\in[S] \tag{1e}$$

$$w_s = a_s + v_s - \sum_{k\in[K]} u_{sk} \qquad \forall s\in[S] \tag{1f}$$

$$x_{sk} \leq u_{sk}\cdot \lfloor p/t_{sk}\rfloor \qquad \forall s\in[S],\ \forall k\in[K] \tag{1g}$$

$$v_s, w_s \geq 0 \qquad \forall s\in[S] \tag{1h}$$

$$u_{sk}, x_{sk} \in \mathbb{Z}_+ \qquad \forall s\in[S],\ \forall k\in[K]. \tag{1i}$$

Constraint (1b) is a stochastic chance constraint, stating that the joint probability to meet the demand in every route is at least $1-\varepsilon$. Constraints (1c) impose a minimum number of voyages to be completed on each route for maintaining the desired minimum sailing frequency. Constraints (1d) ensure the total number of deployed ships cannot exceed the available ones for each ship type, constraints (1e) impose a maximum number of ships to be chartered from other ship owners, and constraints (1f) are the ship flow equations among the numbers of ships to be chartered in/out and deployed and it implies that any available ship in the company is either deployed or chartered out. The number of completed voyages does not exceed the total number of ships deployed multiplied by the maximum times of voyages of each ship (estimated by the planning horizon divided by the transit time, *i.e.*, $\lfloor p/t_{sk}\rfloor$), captured by constraints (1g). The last two constraints are the non-negativity and integrality conditions of the decision variables. From the objective function and constraints (1d)-(1f), it is not hard to see that v_s and w_s must be integers when optimal solution is obtained. In other words, the integrality condition is only needed for variables u_{sk} and x_{sk}.

Let $\boldsymbol{x} = (x_{sk})_{s\in[S],k\in[K]}$, $\boldsymbol{u} = (u_{sk})_{s\in[S],k\in[K]}$, $\boldsymbol{v} = (v_s)_{s\in[S]}$, $\boldsymbol{w} = (w_s)_{s\in[S]}$, and then collectively denote $\boldsymbol{y} = (\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})$. With these notations, we can rewrite the objective function that minimizes a total cost consisting of the operating cost, the payment of chartering ships in, and the negative cost (*i.e.*, revenue) of chartering ships out, by defining

$$F(\boldsymbol{x},\boldsymbol{y}) = \sum_{s\in[S]}\sum_{k\in[K]} c_{sk}x_{sk} + \sum_{s\in[S]} h_s v_s - \sum_{s\in[S]} r_s w_s.$$

For ease of exposition, we use $(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}$ to denote that the decision variables $(\boldsymbol{x},\boldsymbol{y})$ satisfying all constraints (1c)-(1i) in problem (1), except the chance constraint (1b). This leads to the following equivalent representation of problem (1):

$$\begin{aligned} \min \quad & F(\boldsymbol{x},\boldsymbol{y}) \\ \text{s.t.} \quad & \mathbb{P}\left[\tilde{d}_k \le \sum_{s\in[S]} q_s x_{sk} \quad \forall k\in[K]\right] \ge 1-\varepsilon \\ & (\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}. \end{aligned} \tag{2}$$

Given a known demand distribution $\mathbb{P}$ that governs the random vector of uncertain shipping demands $\tilde{\boldsymbol{d}}=(\tilde{d}_k)_{k\in[K]}$, one could use the stochastic programming approach (see, *e.g.*, Shapiro et al. 2014, Birge and Louveaux 2011) to solve the chance constrained model (2). This approach, however, suffers notorious challenges from both computational and practical considerations. Computationally, checking the feasibility of a given decision $\boldsymbol{x}$ involves a multi-dimensional integration that already becomes increasingly hard when the dimension of the random vector grows or when the number of samples increases. Practically, due to the well-acknowledged *overfitting* phenomenon associated with the SAA approach, the out-of-sample performance might not be good. Indeed, our numerical evidence in Appendix C also raise similar concerns.

2.2. Distributionally Robust Chance Constrained Model

To address those issues mentioned in the previous section, we adopt the distributionally robust optimization approach (*e.g.*, Ben-Tal et al. 2009, Bertsimas et al. 2011) to consider the following joint chance constrained model:

$$\begin{aligned} \min \quad & F(\boldsymbol{x},\boldsymbol{y}) \\ \text{s.t.} \quad & \inf_{\mathbb{P}\in\mathcal{F}} \mathbb{P}\left[\tilde{d}_k \le \sum_{s\in[S]} q_s x_{sk} \quad \forall k\in[K]\right] \ge 1-\varepsilon \\ & (\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}. \end{aligned} \tag{3}$$

Here, instead of assuming a known demand distribution $\mathbb{P}$, we consider an ambiguity set $\mathcal{F}$ of distributions sharing some identical distributional information that are deemed reasonable, such as support, mean, and dispersion. By considering a distributionally robust joint chance constraint

$$\inf_{\mathbb{P}\in\mathcal{F}} \mathbb{P}\left[\tilde{d}_k \le \sum_{s\in[S]} q_s x_{sk} \quad \forall k\in[K]\right] \ge 1-\varepsilon \quad \Longleftrightarrow \quad \forall \mathbb{P}\in\mathcal{F}: \mathbb{P}\left[\tilde{d}_k \le \sum_{s\in[S]} q_s x_{sk} \quad \forall k\in[K]\right] \ge 1-\varepsilon,$$

we hedge against the worst-case scenario that may arise from the ambiguity set $\mathcal{F}$. The above ambiguous chance constraint has been receiving considerable interest in the distributionally robust optimization literature (see, *e.g.*, Chen et al. 2018, Postek et al. 2018, Xie and Ahmed 2018, and references therein). Indeed, the models of Ng (2015) that are based on classical statistical inequalities could be re-interpreted under the above distributionally robust optimization framework (3).

Proposition 1 **(Re-Interpretation of Proposition 1 in Ng 2015)**. *Under the mean and marginalized standard deviation ambiguity set*

$$\mathcal{F}=\left\{\mathbb{P}\in\mathcal{P}(\mathbb{R}^K)\;\middle|\;\begin{array}{l}\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{d}}]=\boldsymbol{\mu}\\ \mathbb{E}_{\mathbb{P}}[(\tilde{d}_k-\mu_k)^2]=\sigma_k^2\quad\forall k\in[K]\end{array}\right\},\tag{4}$$

for each $k\in[K]$, given the risk threshold ε_k, the worst-case individual chance constraint

$$\inf_{\mathbb{P}\in\mathcal{F}}\mathbb{P}\left[\tilde{d}_k\leq\sum_{s\in[S]}q_sx_{sk}\right]\geq 1-\varepsilon_k$$

is equivalent to

$$\mu_k+\sigma_k\sqrt{\frac{1-\varepsilon_k}{\varepsilon_k}}\leq\sum_{s\in[S]}q_sx_{sk}.$$

This re-interpretation follows from the Cantelli's inequality.

Proposition 2 **(Re-Interpretation of Proposition 4 in Ng 2015)**. *Under the mean and marginalized standard deviation ambiguity set* (4)*, if $\sum_{k\in[K]}\varepsilon_k\leq\varepsilon$, then the following collection of individual constraints*

$$\mu_k+\sigma_k\sqrt{\frac{1-\varepsilon_k}{\varepsilon_k}}\leq\sum_{s\in[S]}q_sx_{sk}\quad\forall k\in[K]$$

is a relaxation of the worst-case joint chance constraint in problem (3)*. In other words, if replacing the worst-case joint chance constraint in problem* (3) *by the above collection of individual constraints, one then obtains a feasible (but not necessarily optimal) solution to problem* (3)*.*

This proposition is a direct application of the union bound (or Boole's inequality), which implies that the joint chance constraint could be approximated by a collection of individual ones:

$$\inf_{\mathbb{P}\in\mathcal{F}}\mathbb{P}\left[\tilde{d}_k\leq\sum_{s\in[S]}q_sx_{sk}\right]\geq 1-\varepsilon_k\quad\forall k\in[K]\quad\Longrightarrow\quad\inf_{\mathbb{P}\in\mathcal{F}}\mathbb{P}\left[\tilde{d}_k\leq\sum_{s\in[S]}q_sx_{sk}\quad\forall k\in[K]\right]\geq 1-\varepsilon.\tag{5}$$

Based on the above propositions, Ng (2015) relaxes the ambiguous joint chance constraint to a collection of linear inequalities and with a common choice of identical $\varepsilon_k=\varepsilon/K$ for all $k\in[K]$, obtains a mixed-integer linear program reformulation of the distributionally robust joint chance constrained model (3). The author claims that stochastic dependencies among demands on various routes could be considered by this approach, which, however, does not take the potential correlation between different routes into full consideration. The approximation of joint probability constraint by individual ones (*i.e.*, Proposition 2) is at the core of this approach. Unfortunately, such an approximation could be coarse to capture the correlation among different routes: in the following example, we show that such an approximation could lead to a large financial loss.

Example 1 (Ignoring Potential Correlation). Consider three ports A, B, and C as well as two routes $A \to B \to A$ (route 1) and $C \to B \to C$ (route 2). Suppose ports A and C provide the supply and only port B has demand. The setting of different routes share one or multiple intermediate ports is commonly seen in maritime industry, for example, see the 8-routes example in Meng and Wang (2010). Here, we make it simple for the convenience of elucidating the insights. We assume that the demand of port B maintains at a constant level such that

$$\mathbb{P}(\tilde{d}_1 = i, \tilde{d}_2 = 101 - i) = \frac{1}{100} \quad \forall i \in [100],$$

where $\tilde{d}_1$ and $\tilde{d}_2$ are the demand in route 1 and route 2, respectively. It is easy to calculate the marginal distributions of $\tilde{d}_1$ and $\tilde{d}_2$ as

$$\mathbb{P}(\tilde{d}_1 = i) = \frac{1}{100}, \ \mathbb{P}(\tilde{d}_2 = i) = \frac{1}{100} \quad \forall i \in [100].$$

Let D_1 and D_2 (in TEU) be the respective demand satisfactory levels in route 1 and route 2 determined by decision-makers, and suppose that the risk threshold for joint probability satisfactory is 0.98, *i.e.*, $\mathbb{P}(\tilde{d}_1 \le D_1, \tilde{d}_2 \le D_2) \ge 0.98$. Then by the union bound approach in Ng (2015), it is required that $\mathbb{P}(\tilde{d}_1 \le D_1) \ge 0.99$ and $\mathbb{P}(\tilde{d}_2 \le D_2) \ge 0.99$, implying $D_1 \ge 99$ and $D_2 \ge 99$.

On the contrary, if one considers the correlation between the two routes, then by the joint distribution of $\tilde{d}_1$ and $\tilde{d}_2$ we have

$$\mathbb{P}(\tilde{d}_1 + \tilde{d}_2 = 101, \ \tilde{d}_1, \tilde{d}_2 \in \mathbb{Z}_+) = 1, \tag{6}$$

That is to say, any value of $D_1, D_2 \in \mathbb{Z}_+$ that satisfy $D_1 + D_2 \ge 101$ will meet the demand. In this case, $99 \times 2 - 101 = 97$ TEU capacity is wasted in the decision stage, invoking around 96% ($\approx 97/101$) additional budget compared to the optimal solution. ♣

We further analyze the drawback of Proposition 2 in three aspects. First, from the theoretical perspective, the implication in (5), called the Bonferroni approximation, is inherently hard to obtain a satisfying approximation. It is known that the choice of $\varepsilon_k = \varepsilon / K$ for all $k \in [K]$ could be coarse and thus could lead to a conservative approximation (or even hindering the problem being feasible). A more appropriate choice of ε_k's would generally lead to a less conservative approximation. In fact, the question of an optimized Bonferroni approximation remains open (Nemirovski and Shapiro 2007) and is partially addressed by Xie et al. (2019) under certain conditions. Unfortunately, it is typically hard to optimize and find the optimal ε_k's. Second, from the numerical perspective, in Tables 4 and 5 we show that as the risk threshold shrinks or the variance increases, the solution obtained from the union bound approach quickly goes to the boundary of the feasible region and then the model may become infeasible. In other words, such an approach could be very sensitive to the perturbation of parameters, which is due to the inherent drawbacks in the modeling stage. Finally, from the practical perspective, different routes in maritime industry may share many intermediate ports which suggests the correlation among them. Such approach still ignores the potential correlation of different routes, as illustrated in the above Example 1, which may result in a large financial loss to the liner shipping companies.

To address these issues, we next propose the distributionally robust joint chance constrained model that could precisely consider the joint probability satisfactory in the modeling stage, with an ambiguity set that could largely capture the correlation between routes. Our encouraging numerical results show that under parameter perturbations, the proposed distributionally robust optimization approach could be much more cost-saving than the union bound approach (Proposition 2).

3. Ambiguity Set, Reformulation, and Solution Approach

In this section, we adopt recent advances in the distributionally robust optimization literature to directly reformulate the distributionally robust joint chance constrained model (3) into a mixed-integer program that can be solved via a sequential convex optimization algorithm.

3.1. Mean and Dispersion Ambiguity Set

The global liner shipping industry is extremely vulnerable to disruptions such as natural disasters and political conflicts (Taylor 2012), making the demand largely fluctuate. Therefore, it is desirable to have some measure about the demand fluctuation that is robust against outliers or large deviations. This motivates us to leverage a dispersion measure (or to be more specific, mean absolute deviation) that is popular in robust statistics (Casella and Berger 2002). Its successful applications have appeared in many operations management problems recently, *e.g.*, inventory management problem (Levi et al. 2011) for its advantages over other statistical estimators when only limited historical samples are provided, and portfolio optimization problem (DeMiguel and Nogales 2009) as it could yield portfolio weights that are more robust against the outliers. However, few works have considered the dispersion measure in the liner shipping problem before. In particular, we focus on the following mean and dispersion ambiguity set

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}(\mathbb{R}^K) \;\middle|\; \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{d}}] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}[\|\boldsymbol{Q}_\ell(\tilde{\boldsymbol{d}} - \boldsymbol{\mu})\|] \le \sigma_\ell \quad \forall \ell \in [L] \\ \mathbb{P}[\boldsymbol{A}\tilde{\boldsymbol{d}} \le \boldsymbol{b}] = 1 \end{array} \right\}, \tag{7}$$

where $\boldsymbol{Q}_\ell \in \mathbb{R}^{N_\ell \times K}$, $\boldsymbol{A} \in \mathbb{R}^{B \times K}$, $\boldsymbol{b} \in \mathbb{R}^B$, and $\|\cdot\|$ is a general norm of proper dimension. To ensure the strong duality, we assume $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_L) > \boldsymbol{0}$.

This ambiguity set, as its name suggests, specifies information about mean and dispersion around the mean of demand distribution $\mathbb{P}$. It has also been popularly investigated in the distributionally robust optimization literature (see, *e.g.*, Postek et al. 2018, Ghosal and Wiesemann 2020). There are several parameters which could be chosen and tailored to the needs of decision-makers. The parameter $\boldsymbol{Q}_\ell$ is a weight matrix that captures the relation among demand deviations in different routes while the parameters $\boldsymbol{A}$ and $\boldsymbol{b}$ form a general polyhedron[1] to restrict the possible range of demands, *e.g.*, commonly used lower and upper bounds of demands or a limit on the total demand of some related routes. Meanings of these parameters will become more explicit in the subsequent illustrative examples.

Before proceeding, we present two examples wherein the dispersion information could be well captured by the mean and dispersion ambiguity set (7) and we also comment how to use the proposed ambiguity set to model a large class of distributions.

Example 2 (Marginalized Dispersion). *Let $\boldsymbol{Q}_\ell = \boldsymbol{e}_\ell^\top$ for some $\ell \in [K]$ and $\|\cdot\|$ be the ℓ_1-norm. Then the expectation constraint*

$$\mathbb{E}_\mathbb{P}[\|\boldsymbol{Q}_\ell(\tilde{\boldsymbol{d}} - \boldsymbol{\mu})\|] = \mathbb{E}_\mathbb{P}[|\tilde{d}_\ell - \mu_\ell|] \leq \sigma_\ell$$

specifies an upper bound σ_ℓ on the first-order dispersion of the random demand $\tilde{d}_\ell$ in a single route ℓ. It is common to consider the first-order dispersion measure in robust statistics (Casella and Berger 2002) for its better performance against outliers and deviations. It also has computational advantages in optimization to consider first-order information in the ambiguity set, compared to the covariance matrix (second-order information) where semidefinite programming is typically needed (Delage and Ye 2010, Zymler et al. 2013). ♣

Example 3 (Coupled Dispersion). *Sometimes, estimating the deviation of a single route might be risky and a large deviation might be observed. For example, consider a group of routes with many intermediate ports or linked in an end-to-end manner, with the total demand in the ports being relatively stable. As such, we would better estimate the deviation of the sum of demands in this group of routes instead of estimating them individually.* Suppose that $\boldsymbol{Q}_\ell = \sum_{k\in\mathcal{K}} \boldsymbol{e}_k^\top$ where $\mathcal{K}$ is a subset of $[K]$ and let $\|\cdot\|$ be ℓ_1-norm. Then the expectation constraint

$$\mathbb{E}_\mathbb{P}[\|\boldsymbol{Q}_\ell(\tilde{\boldsymbol{d}} - \boldsymbol{\mu})\|] = \mathbb{E}_\mathbb{P}\left[\left|\sum_{k\in\mathcal{K}}(\tilde{d}_k - \mu_k)\right|\right] = \mathbb{E}_\mathbb{P}\left[\left|\sum_{k\in\mathcal{K}}\tilde{d}_k - \sum_{k\in\mathcal{K}}\mu_k\right|\right] \leq \sigma_\ell$$

would specify the upper bound σ_ℓ on the dispersion of the sum of random demands in a collection $\mathcal{K} \subseteq [K]$ of routes. Indeed, the correlation in equation (6) of Example 1 could be captured by this constraint if one sets $\sigma_l = 0$ and $\mathcal{K} = \{1, 2\}$. ♣

[1] Indeed, the polyhedral support is rich to include many common choices, including the box support set and the budgeted support set (Bertsimas and Sim 2004), among others.

REMARK 1. The proposed ambiguity set (7) could capture a large class of distributions including the empirical distribution, and commonly used ones such as the uniform or normal distribution. For example, let the number of historical demand data be N. Most commonly, we could obtain an empirical distribution of the demand data with probability $1/N$ in each data point. Then the empirical distribution will have (*i*) a mean value, (*ii*) a finite first-order dispersion, and (*iii*) a polyhedral support (a possible choice is a box estimated by the minimum and maximum of the demand data in each dimension). Therefore, we could construct an ambiguity set with the above information obtained from the historical data. If one assumes the demand data is uniformly distributed, we could still construct a suitable ambiguity set following the same procedure as above; if one assumes it is a normal distribution, it also has finite mean and first-order dispersion, the support could also be set as the whole real space by letting $\boldsymbol{Q}, \boldsymbol{A}$ and $\boldsymbol{b}$ be zeros. ♣

3.2. Deterministic Reformulation

Given any decision $\boldsymbol{x}$ and an ambiguity set in (7), the distributionally robust joint chance constraint in (3) can be represented into a system of finitely many constraints.

THEOREM 1. *Under the assumption that $\boldsymbol{\sigma} > \boldsymbol{0}$, for any completed voyages $\boldsymbol{x}$, the distributionally robust joint chance constraint*

$$\inf_{\mathbb{P} \in \mathcal{F}} \mathbb{P}\left[\tilde{d}_k \leq \sum_{s \in [S]} q_s x_{sk} \quad \forall k \in [K]\right] \geq 1 - \varepsilon$$

with the ambiguity set $\mathcal{F}$ in (7) is satisfied if and only if there exist variables

$$\begin{cases} \boldsymbol{\rho}_k \in \mathbb{R}^B_+ & \forall k \in [K] \\ \boldsymbol{\phi}_\ell \in \mathbb{R}^{N_\ell}, \ \|\boldsymbol{\phi}_\ell\|_* \leq \beta_\ell & \forall \ell \in [L] \\ \boldsymbol{\psi}_{k\ell} \in \mathbb{R}^{N_\ell}, \ \|\boldsymbol{\psi}_{k\ell}\|_* \leq \beta_\ell & \forall k \in [K], \ \forall \ell \in [L] \\ \boldsymbol{\alpha} \in \mathbb{R}^K, \ \boldsymbol{\beta} \in \mathbb{R}^L_+, \ \boldsymbol{\gamma} \in \mathbb{R}^K_+, \ \boldsymbol{\rho}_0 \in \mathbb{R}^B_+, \ \lambda \in \mathbb{R}, & \end{cases}$$

that satisfy the following constraint system

$$\begin{cases} \lambda - \boldsymbol{\sigma}^\top \boldsymbol{\beta} \geq 1 - \varepsilon & \\ \lambda + (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{\mu})^\top \boldsymbol{\rho}_0 \leq 1 & \\ \boldsymbol{\alpha} + \sum_{\ell \in [L]} \boldsymbol{Q}_\ell^\top \boldsymbol{\phi}_\ell = \boldsymbol{A}^\top \boldsymbol{\rho}_0 & \\ \lambda + (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{\mu})^\top \boldsymbol{\rho}_k \leq \gamma_k \left(\sum_{s \in [S]} q_s x_{sk} - \mu_k\right) & \forall k \in [K] \\ \boldsymbol{\alpha} + \sum_{\ell \in [L]} \boldsymbol{Q}_\ell^\top \boldsymbol{\psi}_{k\ell} + \gamma_k \boldsymbol{e}_k = \boldsymbol{A}^\top \boldsymbol{\rho}_k & \forall k \in [K]. \end{cases}$$

Proof of Theorem 1. First, observe that we could always apply the variable substitution $\tilde{\boldsymbol{d}} \leftarrow \tilde{\boldsymbol{d}} - \boldsymbol{\mu}$ and consider the corresponding distributionally robust chance constraint

$$\inf_{\mathbb{P} \in \mathcal{F}'} \mathbb{P}\left[\tilde{d}_k \leq \sum_{s\in[S]} q_s x_{sk} - \mu_k \quad \forall k \in [K]\right] \geq 1 - \varepsilon \tag{8}$$

with a shifted ambiguity set

$$\mathcal{F}' = \left\{ \mathbb{P} \in \mathcal{P}(\mathbb{R}^K) \;\middle|\; \begin{array}{ll} \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{d}}] = \boldsymbol{0} & \\ \mathbb{E}_{\mathbb{P}}[\|\boldsymbol{Q}_\ell \tilde{\boldsymbol{d}}\|] \leq \sigma_\ell & \forall \ell \in [L] \\ \mathbb{P}[\boldsymbol{A}\tilde{\boldsymbol{d}} \leq \boldsymbol{b} - \boldsymbol{A}\boldsymbol{\mu}] = 1 & \end{array} \right\}.$$

The worst-case uncertainty quantification problem on the left-hand side of (8) can be considered as a moment problem (see, *e.g.*, section 3 of Shapiro 2001):

$$\begin{array}{lll} \min & \mathbb{E}_{\mathbb{P}}[\mathbb{I}_{\{\tilde{d}_k \leq \sum_{s\in[S]} q_s x_{sk} - \mu_k \ \ \forall k\in[K]\}}] & \\ \text{s.t.} & \mathbb{P} \in \mathcal{P}(\mathbb{R}^K) & \\ & \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{d}}] = \boldsymbol{0} & \\ & \mathbb{E}_{\mathbb{P}}[\|\boldsymbol{Q}_\ell \tilde{\boldsymbol{d}}\|] \leq \sigma_\ell & \forall \ell \in [L] \\ & \mathbb{E}_{\mathbb{P}}[\mathbb{I}_{\{\boldsymbol{A}\tilde{\boldsymbol{d}} \leq \boldsymbol{b} - \boldsymbol{A}\boldsymbol{\mu}\}}] = 1, & \end{array} \tag{9}$$

which admits the following dual reformulation

$$\begin{array}{ll} \max & \lambda - \boldsymbol{\sigma}^\top \boldsymbol{\beta} \\ \text{s.t.} & \lambda + \boldsymbol{d}^\top \boldsymbol{\alpha} - \sum_{\ell\in[L]} \|\boldsymbol{Q}_\ell \boldsymbol{d}\| \beta_\ell \leq \mathbb{I}_{\{\boldsymbol{d} \leq \boldsymbol{x}^\top \boldsymbol{q} - \boldsymbol{\mu}\}} \quad \forall \boldsymbol{d} : \boldsymbol{A}\boldsymbol{d} \leq \boldsymbol{b} - \boldsymbol{A}\boldsymbol{\mu} \\ & \boldsymbol{\alpha} \in \mathbb{R}^K, \ \boldsymbol{\beta} \in \mathbb{R}^L_+, \ \lambda \in \mathbb{R}. \end{array} \tag{10}$$

Indeed, strong duality holds under the imposed assumption $\boldsymbol{\sigma} > \boldsymbol{0}$; see Proposition 3.4 of Shapiro (2001). Introducing an auxiliary epigraphical variable $\boldsymbol{g} \in \mathbb{R}^L$ and defining a conic representable set $\mathcal{D} = \{(\boldsymbol{d}, \boldsymbol{g}) \in \mathbb{R}^K \times \mathbb{R}^L : \boldsymbol{b} \geq \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{d}, \ (\boldsymbol{Q}_\ell \boldsymbol{d}, g_\ell) \in \mathcal{C}_{N_\ell} \ \forall \ell \in [L]\}$ with $\mathcal{C}_{N_\ell} = \{(\boldsymbol{\phi}, \eta) \in \mathbb{R}^{N_\ell} \times \mathbb{R} : \|\boldsymbol{\phi}\| \leq \eta\}$, we can then rewrite the first constraint into two different collections of constraints

$$\begin{cases} \lambda + \boldsymbol{d}^\top \boldsymbol{\alpha} - \boldsymbol{g}^\top \boldsymbol{\beta} \leq 1 \ \ \forall (\boldsymbol{d}, \boldsymbol{g}) \in \mathcal{D} \\ \lambda + \boldsymbol{d}^\top \boldsymbol{\alpha} - \boldsymbol{g}^\top \boldsymbol{\beta} \leq 0 \ \ \forall k \in [K], \ (\boldsymbol{d}, \boldsymbol{g}) \in \mathcal{D} : d_k > \sum_{s\in[S]} q_s x_{sk} - \mu_k. \end{cases}$$

Rearranging terms in its expression, the first constraint is equivalent to

$$\max_{(\boldsymbol{d}, \boldsymbol{g}) \in \mathcal{D}} \left\{ \boldsymbol{d}^\top \boldsymbol{\alpha} - \boldsymbol{g}^\top \boldsymbol{\beta} \right\} \leq 1 - \lambda,$$

which, by strong conic duality of the maximization problem on the left-hand side[2], is satisfiable if and only if the following constraint system is feasible:

$$\left[\begin{array}{lll} \min & (\boldsymbol{b}-\boldsymbol{A}\boldsymbol{\mu})^\top \boldsymbol{\rho}_0 & \\ \text{s.t.} & \boldsymbol{\alpha} + \sum_{\ell\in[L]} \boldsymbol{Q}_\ell^\top \boldsymbol{\phi}_\ell - \boldsymbol{A}^\top \boldsymbol{\rho}_0 = \boldsymbol{0} & \\ & -\boldsymbol{\beta} + \boldsymbol{\eta} = \boldsymbol{0} & \\ & \boldsymbol{\rho}_0 \in \mathbb{R}_+^B, \boldsymbol{\eta} \in \mathbb{R}^L & \\ & \boldsymbol{\phi}_\ell \in \mathbb{R}^{N_\ell},\ \|\boldsymbol{\phi}_\ell\|_* \le \eta_\ell & \forall \ell \in [L] \end{array}\right] \le 1-\lambda.$$

In fact, the above constraint system is equivalent to

$$\begin{cases} (\boldsymbol{b}-\boldsymbol{A}\boldsymbol{\mu})^\top \boldsymbol{\rho}_0 \le 1-\lambda \\ \boldsymbol{\alpha} + \sum_{\ell\in[L]} \boldsymbol{Q}_\ell^\top \boldsymbol{\phi}_\ell = \boldsymbol{A}^\top \boldsymbol{\rho}_0 \\ \boldsymbol{\rho}_0 \in \mathbb{R}_+^B \\ \boldsymbol{\phi}_\ell \in \mathbb{R}^{N_\ell},\ \|\boldsymbol{\phi}_\ell\|_* \le \beta_\ell \qquad \forall \ell \in [L]. \end{cases} \tag{11}$$

With this observation and applying similar arguments, we can see that each $k \in [K]$ constraint in the second collection is satisfiable if and only if the corresponding constraint system

$$\begin{cases} \lambda + (\boldsymbol{b}-\boldsymbol{A}\boldsymbol{\mu})^\top \boldsymbol{\rho}_k \le \gamma_k \Big(\sum_{s\in[S]} q_s x_{sk} - \mu_k \Big) \\ \boldsymbol{\alpha} + \sum_{\ell\in[L]} \boldsymbol{Q}_\ell^\top \boldsymbol{\psi}_{k\ell} + \gamma_k \boldsymbol{e}_k = \boldsymbol{A}^\top \boldsymbol{\rho}_k \\ \boldsymbol{\gamma} \in \mathbb{R}_+^K,\ \boldsymbol{\rho}_k \in \mathbb{R}_+^B \\ \boldsymbol{\psi}_{k\ell} \in \mathbb{R}^{N_\ell},\ \|\boldsymbol{\psi}_{k\ell}\|_* \le \beta_\ell \qquad \forall \ell \in [L] \end{cases} \tag{12}$$

is feasible. The remaining proof straightforwardly follows from first substituting constraints (11) and (12) back to the dual problem (10), then substituting the resulting dual problem into the distributionally robust chance constraint (8). □

[2] Here, we note that the underlying cone is $\mathbb{R}_+^B \times \mathcal{C}_{N_1} \times \cdots \times \mathcal{C}_{N_L}$ with dual variables $\boldsymbol{\rho}_0 \in \mathbb{R}_+^B$ and $(\boldsymbol{\phi}_1, \eta_1) \in \mathcal{C}_{N_1}^*, \ldots, (\boldsymbol{\phi}_L, \eta_L) \in \mathcal{C}_{N_L}^*$, where $\mathcal{C}_{N_\ell}^* = \{(\boldsymbol{\phi}, \eta) \in \mathbb{R}^{N_\ell} \times \mathbb{R} : \|\boldsymbol{\phi}\|_* \le \eta\}$.

Equipped with Theorem 1, we can now reformulate the distributionally robust chance constrained model (3) as the following optimization problem:

$$
\begin{array}{lll}
\min & F(\boldsymbol{x},\boldsymbol{y}) & \\
\text{s.t.} & \lambda-\boldsymbol{\sigma}^\top\boldsymbol{\beta}\geq 1-\varepsilon & \\
& \lambda+(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{\mu})^\top\boldsymbol{\rho}_0\leq 1 & \\
& \boldsymbol{\alpha}+\sum_{\ell\in[L]}\boldsymbol{Q}_\ell^\top\boldsymbol{\phi}_\ell=\boldsymbol{A}^\top\boldsymbol{\rho}_0 & \\
& \lambda+(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{\mu})^\top\boldsymbol{\rho}_k\leq\gamma_k\left(\sum_{s\in[S]}q_sx_{sk}-\mu_k\right) & \forall k\in[K] \\
& \boldsymbol{\alpha}+\sum_{\ell\in[L]}\boldsymbol{Q}_\ell^\top\boldsymbol{\psi}_{k\ell}+\gamma_k\boldsymbol{e}_k=\boldsymbol{A}^\top\boldsymbol{\rho}_k & \forall k\in[K] \\
& \boldsymbol{\rho}_k\in\mathbb{R}^B_+ & \forall k\in[K] \\
& \boldsymbol{\phi}_\ell\in\mathbb{R}^{N_\ell},\ \|\boldsymbol{\phi}_\ell\|_*\leq\beta_\ell & \forall \ell\in[L] \\
& \boldsymbol{\psi}_{k\ell}\in\mathbb{R}^{N_\ell},\ \|\boldsymbol{\psi}_{k\ell}\|_*\leq\beta_\ell & \forall k\in[K],\ \forall\ell\in[L] \\
& \boldsymbol{\alpha}\in\mathbb{R}^K,\ \boldsymbol{\beta}\in\mathbb{R}^L_+,\ \boldsymbol{\gamma}\in\mathbb{R}^K_+,\ \boldsymbol{\rho}_0\in\mathbb{R}^B_+,\ \lambda\in\mathbb{R} & \\
& (\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}. &
\end{array}
\tag{13}
$$

Due to product terms of decision variables (*e.g.*, γ_k and x_{sk}) in the fourth collection of constraints, reformulation (13) is not a conic program that can be directly solved by off-the-shelf solvers. It is worth noting that under a certain condition $\boldsymbol{A}\boldsymbol{\mu}=\boldsymbol{b}$ on parameters $\boldsymbol{A},\boldsymbol{b}$ and $\boldsymbol{\mu}$ in the ambiguity set (7), this reformulation becomes a mixed-integer second-order cone program that can be efficiently solved by the state-of-the-art solvers. We first discuss two examples for which the condition $\boldsymbol{A}\boldsymbol{\mu}=\boldsymbol{b}$ naturally holds, then we establish this result formally in a subsequent proposition.

Example 4 (Partially Deterministic Demands). *In practice, it could be possible that demands in some routes of a subset $\mathcal{K}\subseteq[K]$ are deterministic, e.g., by committing to certain pre-orders.* Then the support constraint would be

$$\mathbb{P}[\boldsymbol{A}\tilde{\boldsymbol{d}}\leq\boldsymbol{b}]=\mathbb{P}\big[\tilde{d}_k=\mu_k\ \forall k\in\mathcal{K}\big]=1,$$

where the matrices are, respectively, given by

$$\boldsymbol{A}=\begin{pmatrix}\operatorname{diag}\left(\sum_{k\in\mathcal{K}}\boldsymbol{e}_k\right)\\ -\operatorname{diag}\left(\sum_{k\in\mathcal{K}}\boldsymbol{e}_k\right)\end{pmatrix}\quad\text{and}\quad\boldsymbol{b}=\begin{pmatrix}\sum_{k\in\mathcal{K}}\mu_k\boldsymbol{e}_k\\ -\sum_{k\in\mathcal{K}}\mu_k\boldsymbol{e}_k\end{pmatrix}.$$

It is not hard to verify that $\boldsymbol{A}\boldsymbol{\mu}=\boldsymbol{b}$ holds in this scenario. ♣

Example 5 (Competitive Routes). *Competitions between shipping lines have received considerable interest recently (Lee and Song 2017). In a simple competition model, the market size of the liner shipping industry between two competitors remains relatively stable in a short period of time. That is to say, the total demand across the routes provided by two competitors stays at a* *constant level and could be well bounded by the sum of their mean values, for an increase in one route may result in a decline in another and vice versa.* The support constraint can be written as

$$\mathbb{P}[\boldsymbol{A}\tilde{\boldsymbol{d}} \le \boldsymbol{b}] = \mathbb{P}\left[\sum_{k\in\mathcal{K}} \tilde{d}_k \le \sum_{k\in\mathcal{K}} \mu_k\right] = 1,$$

where $\boldsymbol{A} = \sum_{k\in\mathcal{K}} \boldsymbol{e}_k^\top$ and $\boldsymbol{b} = \sum_{k\in\mathcal{K}} \mu_k$, and $\mathcal{K}$ could be the set of routes between two countries. It is clear that $\boldsymbol{A\mu} = \boldsymbol{b}$ herein. In fact, equation (6) in Example 1 is a special case of this setting where $\tilde{d}_1 + \tilde{d}_2 = 101$ with probability 1. ♣

PROPOSITION 3. *In the same setting of Theorem 1, if* $\boldsymbol{A\mu} = \boldsymbol{b}$ *then the distributionally robust chance constrained model* (3) *reduces to the following mixed-integer second-order cone program:*

$$\begin{array}{lll}
\min & F(\boldsymbol{x},\boldsymbol{y}) & \\
\text{s.t.} & 1 - \boldsymbol{\sigma}^\top\boldsymbol{\beta} \ge 1 - \varepsilon & \\
& \boldsymbol{\alpha} + \sum_{\ell\in[L]} \boldsymbol{Q}_\ell^\top \boldsymbol{\phi}_\ell = \boldsymbol{A}^\top \boldsymbol{\rho}_0 & \\
& \left\|\left(2, \gamma_k - \sum_{s\in[S]} q_s x_{sk} + \mu_k\right)\right\|_2 \le \gamma_k + \sum_{s\in[S]} q_s x_{sk} - \mu_k & \forall k \in [K] \\
& \boldsymbol{\alpha} + \sum_{\ell\in[L]} \boldsymbol{Q}_\ell^\top \boldsymbol{\psi}_{k\ell} + \gamma_k \boldsymbol{e}_k = \boldsymbol{A}^\top \boldsymbol{\rho}_k & \forall k \in [K] \\
& \boldsymbol{\rho}_k \in \mathbb{R}_+^B & \forall k \in [K] \\
& \boldsymbol{\phi}_\ell \in \mathbb{R}^{N_\ell},\ \|\boldsymbol{\phi}_\ell\|_* \le \beta_\ell & \forall \ell \in [L] \\
& \boldsymbol{\psi}_{k\ell} \in \mathbb{R}^{N_\ell},\ \|\boldsymbol{\psi}_{k\ell}\|_* \le \beta_{k\ell} & \forall k \in [K],\ \forall \ell \in [L] \\
& \boldsymbol{\alpha} \in \mathbb{R}^K,\ \boldsymbol{\beta} \in \mathbb{R}_+^L,\ \boldsymbol{\gamma} \in \mathbb{R}_+^K,\ \boldsymbol{\rho}_0 \in \mathbb{R}_+^B & \\
& (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}. &
\end{array}$$

Proof of Proposition 3. Since $\boldsymbol{A\mu} = \boldsymbol{b}$, the first constraints in systems (11) and (12) reduce, respectively, to $\lambda \le 1$ and

$$\lambda \le \gamma_k \left(\sum_{s\in[S]} q_s x_{sk} - \mu_k\right) \quad \forall k \in [K].$$

It remains to argue that one can add $\lambda = 1$ for free and hence the collection of constraints above admits the following second-order cone representation:

$$\left\|\left(2, \gamma_k - \sum_{s\in[S]} q_s x_{sk} + \mu_k\right)\right\|_2 \le \gamma_k + \sum_{s\in[S]} q_s x_{sk} - \mu_k \quad \forall k \in [K].$$

Indeed, suppose the distributionally robust chance constraint (8) is feasible and $(\boldsymbol{\beta}_\star, \lambda_\star, \boldsymbol{\Gamma}_\star)$ is an optimal solution to the dual problem (10), where $\boldsymbol{\Gamma}_\star$ collectively denotes $\boldsymbol{\alpha}^\star$ and other auxiliary decision variables that may arise. Since $\boldsymbol{\beta}_\star \ge 0$ and the chance constraint is satisfied, we have

$$\lambda_\star \ge 1 - \varepsilon + \boldsymbol{\sigma}^\top \boldsymbol{\beta}_\star > 0.$$

Note that the constraints in (11) and (12) (except for $(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{\mu})^\top\boldsymbol{\rho}\le 1-\lambda$, *i.e.*, $\lambda\le 1$ when $\boldsymbol{A}\boldsymbol{\mu}=\boldsymbol{b}$) are positive homogeneous, that is, for any $c_0>0$, $(c_0\boldsymbol{\beta}_\star, c_0\lambda_\star, c_0\boldsymbol{\Gamma}_\star)$ remains feasible to these constraints. If $\lambda_\star<1$, then $(c_0\boldsymbol{\beta}_\star, c_0\lambda_\star, c_0\boldsymbol{\Gamma}_\star)$ with $c_0=1/\lambda_\star$ is also a feasible solution. However, $c_0\lambda_\star - c_0\boldsymbol{\sigma}^\top\boldsymbol{\beta}_\star > \lambda_\star - \boldsymbol{\sigma}^\top\boldsymbol{\beta}_\star$ would contradict to the optimality assumption. Hence, $\lambda_\star=1$. □

REMARK 2. *The key to the above proof is the positive homogeneity that enables us to positively scale* $(\boldsymbol{\beta}_\star,\lambda_\star,\boldsymbol{\Gamma}_\star)$ *as long as* $\lambda_\star\le 1$. *This will not affect the optimal solution of* $\boldsymbol{x}$ *and* $\boldsymbol{y}$, *thus we may assume that* $\lambda_\star=1$ *in the first place. The proof also suggests that whenever* $\lambda_\star<1$, *the objective function in the dual problem* (10) *will not be maximized.* ♣

Apart from Examples 4 and 5, a simpler example where the condition $\boldsymbol{A}\boldsymbol{\mu}=\boldsymbol{b}$ holds is the no-support case such that $\boldsymbol{A}=\boldsymbol{0}$ and $\boldsymbol{b}=\boldsymbol{0}$ in (7). On the one hand, estimating the support information may not be an easy task in many practical situations, showing the need for no-support ambiguity sets. On the other hand, the optimal solution under a no-support ambiguity set can be obtained from solving a mixed-integer second-order cone program and it can be used as the initial input for a sequential convex optimization algorithm that solves reformulation (13). Before we proceed to introduce this algorithm formally, we give two corollaries that also fall within the general format of ambiguity set (7) in Theorem 1.

COROLLARY 1. *Consider an instance of the ambiguity set* (7) *with box support and upper bounds on mean absolute deviations of each individual demand as well as that of their sum*

$$\mathcal{F}=\left\{\mathbb{P}\in\mathcal{P}(\mathbb{R}^K)\ \middle|\ \begin{array}{ll}\mathbb{E}_\mathbb{P}[\tilde{\boldsymbol{d}}]=\boldsymbol{\mu} & \\ \mathbb{E}_\mathbb{P}[|\boldsymbol{e}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{\mu})|]\le\sigma_0 & \\ \mathbb{E}_\mathbb{P}[|\tilde{d}_k-\mu_k|]\le\sigma_k & \forall k\in[K] \\ \mathbb{P}\big[\tilde{\boldsymbol{d}}\in[\underline{\boldsymbol{d}},\bar{\boldsymbol{d}}]\big]=1 & \end{array}\right\}. \tag{14}$$

The corresponding distributionally robust chance constrained model (3) *is equivalent to the following optimization problem*

$$
\begin{aligned}
\min\quad & F(\boldsymbol{x},\boldsymbol{y}) \\
\text{s.t.}\quad & \lambda - \sigma_0\beta_0 - \boldsymbol{\sigma}^\top\boldsymbol{\beta} \geq 1-\varepsilon \\
& \lambda + (\boldsymbol{\mu}-\underline{\boldsymbol{d}})^\top \underline{\boldsymbol{\rho}}_0 + (\bar{\boldsymbol{d}}-\boldsymbol{\mu})^\top \bar{\boldsymbol{\rho}}_0 \leq 1 \\
& \boldsymbol{\alpha} + \phi_0\boldsymbol{e} + \boldsymbol{\phi} = \bar{\boldsymbol{\rho}}_0 - \underline{\boldsymbol{\rho}}_0 \\
& \lambda + (\boldsymbol{\mu}-\underline{\boldsymbol{d}})^\top \underline{\boldsymbol{\rho}}_k + (\bar{\boldsymbol{d}}-\boldsymbol{\mu})^\top \bar{\boldsymbol{\rho}}_k \leq \gamma_k\left(\sum_{s\in[S]} q_s x_{sk} - \mu_k\right) && \forall k\in[K] \\
& \boldsymbol{\alpha} + \psi_{k0}\boldsymbol{e} + \boldsymbol{\psi}_k + \gamma_k\boldsymbol{e}_k = \bar{\boldsymbol{\rho}}_k - \underline{\boldsymbol{\rho}}_k && \forall k\in[K] \\
& \begin{pmatrix}\bar{\boldsymbol{\rho}}_k \\ \underline{\boldsymbol{\rho}}_k\end{pmatrix} \in \mathbb{R}^{2K}_+ && \forall k\in[K] \\
& \phi_0\in\mathbb{R},\ \boldsymbol{\phi}\in\mathbb{R}^K,\ -\beta_0\leq\phi_0\leq\beta_0,\ -\boldsymbol{\beta}\leq\boldsymbol{\phi}\leq\boldsymbol{\beta} \\
& \psi_{k0}\in\mathbb{R},\ \boldsymbol{\psi}_k\in\mathbb{R}^K,\ -\beta_0\leq\psi_{k0}\leq\beta_0,\ -\boldsymbol{\beta}\leq\boldsymbol{\psi}_k\leq\boldsymbol{\beta} && \forall k\in[K] \\
& \boldsymbol{\alpha}\in\mathbb{R}^K,\ \beta_0\in\mathbb{R}_+,\ \boldsymbol{\beta}\in\mathbb{R}^K_+,\ \boldsymbol{\gamma}\in\mathbb{R}^K_+,\ \begin{pmatrix}\bar{\boldsymbol{\rho}}_0 \\ \underline{\boldsymbol{\rho}}_0\end{pmatrix}\in\mathbb{R}^{2K}_+,\ \lambda\in\mathbb{R} \\
& (\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}.
\end{aligned}
$$

Proof of Corollary 1. We first observe that the expectation constraints

$$\mathbb{E}_{\mathbb{P}}[|\boldsymbol{e}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{\mu})|]\leq\sigma_0 \quad \text{and} \quad \mathbb{E}_{\mathbb{P}}[|\tilde{d}_k-\mu_k|]\leq\sigma_k \quad \forall k\in[K]$$

can be collectively represented in the form of (7) with $\boldsymbol{Q}_0=\boldsymbol{e}^\top$ and $\boldsymbol{Q}_\ell=\boldsymbol{e}_\ell^\top$ for each $\ell\in[K]$. Since the support set can be represented as

$$\{\boldsymbol{d}\mid\underline{\boldsymbol{b}}\leq\boldsymbol{d}\leq\bar{\boldsymbol{b}}\}=\{\boldsymbol{d}\mid\boldsymbol{A}\boldsymbol{d}\leq\boldsymbol{b}\} \quad \text{with} \quad \boldsymbol{A}=\begin{pmatrix}\boldsymbol{I}\\-\boldsymbol{I}\end{pmatrix} \quad \text{and} \quad \boldsymbol{b}=\begin{pmatrix}\bar{\boldsymbol{d}}\\-\underline{\boldsymbol{d}}\end{pmatrix},$$

then in the proof of Theorem 1 we can define $\boldsymbol{\rho}_0=(\bar{\boldsymbol{\rho}}_0,\underline{\boldsymbol{\rho}}_0)$ and $\boldsymbol{\rho}_k=(\bar{\boldsymbol{\rho}}_k,\underline{\boldsymbol{\rho}}_k)$, $k\in[K]$. Finally, the remaining proof then follows from plugging the above notations in Theorem 1. □

Since the mean absolute deviation is not larger than the standard deviation, a proper design of the ambiguity set in Corollary 1 can contain the mean and marginalized standard deviation ambiguity set (4) considered in Ng (2015). In particular, by assuming the no-support condition $\boldsymbol{A}=\boldsymbol{0}$ and $\boldsymbol{b}=\boldsymbol{0}$ in Proposition 3 we have the following result immediately.

Corollary 2. *Consider the marginalized mean absolute deviation ambiguity set*

$$\mathcal{F}=\left\{\mathbb{P}\in\mathcal{P}(\mathbb{R}^K)\ \middle|\ \begin{array}{l}\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{d}}]=\boldsymbol{\mu}\\ \mathbb{E}_{\mathbb{P}}[|\tilde{d}_k-\mu_k|]\leq\sigma_k\quad\forall k\in[K]\end{array}\right\}. \tag{15}$$

The corresponding distributionally robust chance constrained model (3) *is equivalent to the following optimization problem*

$$
\begin{aligned}
\min \quad & F(\boldsymbol{x},\boldsymbol{y}) \\
\text{s.t.} \quad & 1-\boldsymbol{\sigma}^\top\boldsymbol{\beta} \geq 1-\varepsilon \\
& \left\| \left(2, \gamma_k - \sum_{s\in[S]} q_s x_{sk} + \mu_k\right) \right\|_2 \leq \gamma_k + \sum_{s\in[S]} q_s x_{sk} - \mu_k \quad \forall k \in [K] \\
& -\boldsymbol{\beta} \leq \boldsymbol{\alpha} \leq \boldsymbol{\beta} \\
& -\boldsymbol{\beta} \leq \boldsymbol{\alpha} + \gamma_k \boldsymbol{e}_k \leq \boldsymbol{\beta} \quad \forall k \in [K] \\
& \boldsymbol{\alpha} \in \mathbb{R}^K,\ \boldsymbol{\beta} \in \mathbb{R}^K_+,\ \boldsymbol{\gamma} \in \mathbb{R}^K_+ \\
& (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}.
\end{aligned}
$$

Note that the ambiguity set (15) is a superset of the ambiguity set (14) in Corollary 1. Hence, to solve the formulation provided in Corollary 1, the optimal solution obtained in Corollary 2 could be served as the initial feasible input fed to our sequential convex approximation algorithm, which we formally introduce next.

3.3. Sequential Convex Optimization Algorithm

In general, due to product terms of decision variables (*e.g.*, γ_k and x_{sk}) in the fourth collection of constraints, reformulation (13) is not a conic program that can be directly solved by off-the-shelf solvers. Nevertheless, as advocated in Hanasusanto et al. (2015), high-quality approximate solution to problem (13) can be found and sequentially improved via an algorithm that consists of solving convex optimization problems. Detailed procedures of such a sequential convex optimization algorithm are summarized as follows.

The idea is originated from Hanasusanto et al. (2015) and is based on a critical observation that the distributionally robust joint chance constrained model (3) can be decomposed into an uncertainty quantification problem that evaluates for a fixed decision $\boldsymbol{x}$ the worst-case probability of the chance constraint in problem (3) and a policy improvement problem to improve the current decision $\boldsymbol{x}$. To this end, let $\boldsymbol{\Omega}$ collectively combine all decision variables in reformulation (13), excluding $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{\gamma}$. By their definitions, $\boldsymbol{\gamma}$ is the only decision variable that is coupled with $\boldsymbol{x}$,[3] while any decision variable in $\boldsymbol{\Omega}$ is not. With these notations, we can rewrite reformulation (13) of the distributionally robust chance constrained model (3) as follows:

$$
\begin{aligned}
\min \quad & F(\boldsymbol{x},\boldsymbol{y}) \\
\text{s.t.} \quad & \lambda-\boldsymbol{\sigma}^\top\boldsymbol{\beta} \geq 1-\varepsilon \\
& (\boldsymbol{\gamma},\boldsymbol{\Omega}) \in \mathcal{G}(\boldsymbol{x}) \\
& (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X},
\end{aligned}
$$

[3] Two variable are “coupled” if there is a multiplication of them.

where $(\boldsymbol{\gamma}, \boldsymbol{\Omega}) \in \mathcal{G}(\boldsymbol{x})$ denotes collectively all other constraints in model (13) except the coming two constraints: (i) $\lambda - \boldsymbol{\sigma}^\top \boldsymbol{\beta} \geq 1 - \varepsilon$ and (ii) $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}$. Observe that given $\boldsymbol{x}$, the uncertainty quantification problem

$$\inf_{\mathbb{P} \in \mathcal{F}} \mathbb{P}\left[\tilde{d}_k \leq \sum_{s \in [S]} q_s x_{sk} \quad \forall k \in [K]\right]$$

is equivalent to the following problem

$$\begin{aligned} \max \quad & \lambda - \boldsymbol{\sigma}^\top \boldsymbol{\beta} \\ \text{s.t.} \quad & (\boldsymbol{\gamma}, \boldsymbol{\Omega}) \in \mathcal{G}(\boldsymbol{x}). \end{aligned}$$

We are now ready to present the algorithm.

Sequential Convex Optimization Algorithm (Hanasusanto et al. 2015)

Input: An initial feasible solution $(\boldsymbol{x}^0, \boldsymbol{y}^0)$ and the maximum number of iterations $i_{\max}$.

1. Let $i \leftarrow 0$ and $f^0 \leftarrow F(\boldsymbol{x}^0, \boldsymbol{y}^0)$.
2. For $i < i_{max}$, let $i \leftarrow i + 1$ and do:

    Uncertainty Quantification Problem.

    Let $(\boldsymbol{\gamma}^\star, \boldsymbol{\Omega}^\star)$ be the solution of

    $$\max_{(\boldsymbol{\gamma}, \boldsymbol{\Omega})} \left\{ \lambda - \boldsymbol{\sigma}^\top \boldsymbol{\beta} : (\boldsymbol{\gamma}, \boldsymbol{\Omega}) \in \mathcal{G}(\boldsymbol{x}^{i-1}) \right\}.$$

    Set $\boldsymbol{\gamma}^i \leftarrow \boldsymbol{\gamma}^\star$.

    Policy Improvement Problem.

    Let $(\boldsymbol{x}^\star, \boldsymbol{y}^\star, \boldsymbol{\Omega}^\star)$ be the solution to

    $$\min_{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\Omega})} \left\{ F(\boldsymbol{x}, \boldsymbol{y}) : \lambda - \boldsymbol{\sigma}^\top \boldsymbol{\beta} \geq 1 - \varepsilon,\ (\boldsymbol{\gamma}^i, \boldsymbol{\Omega}) \in \mathcal{G}(\boldsymbol{x}),\ (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \right\}.$$

    Set $(\boldsymbol{x}^i, \boldsymbol{y}^i, \boldsymbol{\Omega}^i) \leftarrow (\boldsymbol{x}^\star, \boldsymbol{y}^\star, \boldsymbol{\Omega}^\star)$ and $f^i \leftarrow F(\boldsymbol{x}^i, \boldsymbol{y}^i)$.
3. Go to the output step if $|f^i - f^{i-1}|$ is smaller than some convergence threshold[4], otherwise go to step 2.

Output: optimal solution $(\boldsymbol{x}^\star, \boldsymbol{y}^\star)$ and optimal objective value $F(\boldsymbol{x}^\star, \boldsymbol{y}^\star)$.

We refer interested readers to Gorski et al. (2007) for convergence analysis of the sequential convex optimization algorithm. In particular, the designed algorithm is a special case of Algorithm 4.1 in Gorski et al. (2007), whose convergence is guaranteed by Theorem 4.5 therein. Because accessing

[4] Indeed, in all of our numerical experiments the algorithm converges to a fixed point after tens of steps. Here, the threshold is to check whether f^i is equal to f^{i-1} up to some precision.

to the global optimum is not easy in general, it remains rather challenging to quantify the optimality gap analytically or even numerically. Nevertheless, experiment results in Section 4 suggest that our sub-optimal heuristic performs reasonably well. Besides, as observed in both Hanasusanto et al. (2015) and our experiments, the algorithm is practically efficient: (i) a good initial input could be obtained by considering no-support ambiguity sets (Proposition 3) and (ii) both uncertainty qualification and policy improvement problems are indeed mixed-integer linear programs if we consider ambiguity sets as in Corollary 1.

4. Numerical Experiments

We adopt from the existing literature (Ng 2014, 2015) an 8-route Orient Overseas Container Line shipping line example, where the planning horizon (in days) is $p = 180$, the number of voyages on each route to maintain the minimal sailing frequency is $n = 26$, and the expected shipping demands (in TEUs) are $\mathbb{E}[\tilde{\boldsymbol{d}}] = \boldsymbol{\mu} = (78000, 52000, 52000, 130000, 78000, 52000, 78000, 26000)$. Other parameters are summarized in Table 2. We denote the model of Ng (2015) by "CUB"[5] and our model as in Corollary 1 by "DRO". In particular, our model is solved by the sequential convex optimization (SCO) algorithm with a maximal iteration $i_{\max} = 200$. All results were produced on an Intel Xeon 3.00GHz processor with 32GB memory in single-thread mode using Gurobi 9.0. Further numerical experiments under the data-driven setting are relegated to Appendix C. The main message conveyed therein is similar as in this section.

4.1. Comparison of Solutions

We first repeat the experiment in Ng (2015), where $\varepsilon = 0.05$ and $\bar{\boldsymbol{\sigma}} = (\bar{\sigma}_k)_{k \in [K]}$ with $\bar{\sigma}_k = \sqrt{0.005}\mu_k \ \forall k \in [K]$. We shall call $\bar{\boldsymbol{\sigma}}$ the baseline standard deviation. We set $[\underline{\boldsymbol{d}}, \bar{\boldsymbol{d}}] = [\boldsymbol{0}, 2\boldsymbol{\mu}]$, where the lower bound guarantees that shipping demands are naturally non-negative and the upper bound ensures the mean to be the midpoint of the support set. Observe that for any distribution $\mathbb{P}$, the mean absolute deviation of a random scalar is bounded from above by the corresponding standard deviation via Jensen inequality, that is, $\mathbb{E}_{\mathbb{P}}[|\tilde{d} - \mu|] \leq \sqrt{\mathbb{E}_{\mathbb{P}}[(\tilde{d} - \mu)^2]}$. Thus we could use the baseline standard deviation $\bar{\boldsymbol{\sigma}}$ as an upper bound on the mean absolute deviation in the DRO model. For the upper bound σ_0 on the mean absolute deviation of the sum of shipping demands, we set it as $\sigma_0 = \boldsymbol{e}^\top \bar{\boldsymbol{\sigma}}/2$, half of the sum of baseline standard deviations.

The results are reported in Table 3.[6] Although chartering fewer ships from other owners, our model indeed yields a vessel deployment quite similar to that of Ng (2015)—the difference in optimal cost is less than 5%. This is because distributionally robust chance constraints to be tackled in both models are not very tight; see the evidence that chartered ships of all types have not reached their upper bound $\boldsymbol{m} = (10, 10, 10, 6, 6)$.

[5] The "C" stands for the Chebyshev inequality and "UB" stands for the union bound.

[6] 'Optimal' for the DRO model refers to the solution obtained from the SCO algorithm, not the global optimum.

Ship type s	1	2	3	4	5
q_s (in TEUs)	2808	3218	4500	5714	8063
c_{sk} (in thousand \$/day)	19.8	22.5	30.9	38.8	54.2
r_s (in million \$)	1.82	2.34	3.21	4.32	5.12
h_s (in million \$)	2	2.6	3.5	4.7	6
a_s	2	2	9	2	12
m_s	10	10	10	6	6
t_{s1} (in days)	25.2	24.1	21.9	21.6	21.0
t_{s2} (in days)	20.7	19.7	17.9	17.6	17.2
t_{s3} (in days)	15.1	14.4	13.1	12.9	12.6
t_{s4} (in days)	38.9	37.1	33.8	33.2	32.4
t_{s5} (in days)	63.8	60.9	55.4	54.5	53.2
t_{s6} (in days)	22.8	21.7	19.8	19.4	19.0
t_{s7} (in days)	58.0	55.4	50.4	49.5	48.4
t_{s8} (in days)	2.1	2.0	1.8	1.8	1.8

Table 2 Parameters.

	Optimal cost = \$ 1340 million $\boldsymbol{v}=(9,1,3,1,0),\boldsymbol{w}=(0,0,0,0,0)$								Optimal cost = \$ 1398 million $\boldsymbol{v}=(5,2,2,0,3),\boldsymbol{w}=(0,0,0,0,0)$							
Route	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
u_{1k}	1		1		2	2	4	1	1	1	1		1		2	1
u_{2k}		2	1						1		1			1	1	
u_{3k}	2	1	1	2	4	2				2	1	2	3	2	1	
u_{4k}					2		1					1	1			
u_{5k}	1			5	2		4		2			4	4		5	
x_{1k}	4		4		4	11	12	26	7	8	3		2		5	26
x_{2k}		18	9						3		12			8	3	
x_{3k}	16	9	13	10	12	15				19	13	10	9	18	3	
x_{4k}					6		3					5	3			
x_{5k}	8			25	6		12		16			20	12		15	

Table 3 Optimal solutions of the CUB model (left) and the DRO model (right), when $n=26$, $\varepsilon=0.05$, and $\boldsymbol{\sigma}=\bar{\boldsymbol{\sigma}}$. The mean runtime over 50 times of both models are 2.05 and 219.39 seconds respectively.

On the other hand, when the distributionally robust chance constraint to be tackled becomes tighter, our model remains quite stable and outperforms that of Ng (2015). Tables 4 and 5 show

	Optimal cost = \$ 2130 million $\boldsymbol{v}=(7,4,10,6,6), \boldsymbol{w}=(0,0,0,0,0)$								Optimal cost = \$ 1426 million $\boldsymbol{v}=(6,\ 0,\ 2,\ 0,\ 4), \boldsymbol{w}=(0,0,0,0,0)$							
Route	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
u_{1k}	1	2	1				4	1	1	1	2				3	1
u_{2k}			1				4		1					1		
u_{3k}		1		3	13	1	1			2		1	5	2	1	
u_{4k}	2				2	1	3				1		1			
u_{5k}	2	1	1	8	1	1	4		2			6	3		5	
x_{1k}	5	14	4				12	28	7	7	15				8	26
x_{2k}			10				12		3					8		
x_{3k}		8		15	39	9	3			19		5	15	18	3	
x_{4k}	16				6	9	9				11		3			
x_{5k}	16	10	14	40	3	8	12		16			30	9		15	

Table 4 Optimal solutions of the CUB model (left) and the DRO model (right), when $n=26$, $\varepsilon=0.01$, and $\boldsymbol{\sigma}=\bar{\boldsymbol{\sigma}}$. The mean runtime over 50 times of both models are 22.54 and 80.68 seconds respectively.

the results of a stricter risk threshold (for the fixed value of demand dispersion in Table 3) and a larger value of demand dispersion (for the fixed risk threshold in Table 3), respectively. In both cases of Tables 4 and 5, distributionally robust chance constraints are stricter than those in the corresponding case of Table 3. It can be seen that the model of Ng (2015) almost uses up allowable chartered ships, leading to a deployment that is much more costly than ours. Indeed, the model of Ng (2015) will quickly become infeasible if $\varepsilon < 0.01$ or $\boldsymbol{\sigma} > 2.64\bar{\boldsymbol{\sigma}}$. In terms of interpretation to real-world scenarios, the stricter risk threshold represents a higher level of service satisfactory commitment in the routes, while the larger demand dispersion represents higher uncertainty in the liner shipping industry, which could be the result of unpredictable events in international trading. Unsurprisingly, both conditions will invoke higher cost, but the robust approach is much more cost-saving than the existing model, which demonstrates its ability to better handle the uncertainty.

4.2. Perturbation Analysis

We also test the performance of both models with different n, the number of voyages required to maintain the minimum sailing frequency in each route, and $\boldsymbol{\sigma}$, the standard deviation of the demand in each route. In particular, we first run experiments on $n \in \{14, 20, 26, 32, 38\}$, which represents different service levels of minimum sailing frequency from low to high. Other parameters are remained the same as those in Table 3; similarly, we then test $\boldsymbol{\sigma}=\kappa\bar{\boldsymbol{\sigma}}$ with different choices of $\kappa \in \{0.5, 1.0, 1.5, 2.0, 2.5\}$, which represents the different level of variations of demand. The results are presented in Figure 1, where two implications are worth noting. First, when the number of

	Optimal cost = \$ 2404 million $\boldsymbol{v}=(10,10,10,6,6),\boldsymbol{w}=(0,0,0,0,0)$								Optimal cost = \$ 1420 million $\boldsymbol{v}=(8,0,2,0,4),\boldsymbol{w}=(0,0,0,1,0)$							
Route	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
u_{1k}							11	1	2	1	1	1	1		3	1
u_{2k}	1	1					10				1			1		
u_{3k}		2	3	2	12					2	1	1	4	2	1	
u_{4k}	1	1		1	3	2						1				
u_{5k}	3			9	2	1	3		2			5	4		5	
x_{1k}							33	32	10	7	6	3	2		8	26
x_{2k}	7	9					30				9			8		
x_{3k}		20	39	10	36					19	13	5	12	18	3	
x_{4k}	8	10		5	9	18						5				
x_{5k}	24			45	6	9	9		16			25	12		15	

Table 5 Optimal solutions of the CUB model (left) and the DRO model (right), when $n=26$, $\varepsilon=0.05$, $\boldsymbol{\sigma}=2.64\bar{\boldsymbol{\sigma}}$. The mean runtime over 50 times of both models are 0.81 and 70.04 seconds respectively.

voyages n needed in each route increases, the optimal costs of both models also increase while the gap between them diminishes, which is consistent with the result in Figure 2 of Ng (2015). The insight behind this observation is that when the minimum number of voyages required at each route increases, the chance constraint actually becomes redundant since satisfying the required minimum sailing frequency already implies the satisfaction of the chance constraint on demands. Second, the DRO model is more robust to the change of demand deviation. In other words, the cost does not increase severely as the demand deviation increases. The reason is that the robust approach is not directly determined by the standard deviation, other information such as the support set also matters. On the contrary, the optimal cost of CUB model almost increases linearly with respect to $\boldsymbol{\sigma}$, which makes the CUB model quickly become infeasible when the variance of demand increases. An explanation of this observation is that the required total number of voyages is linearly dominated by the demand deviation, as suggested by the linear constraint $\mu_k+\sigma_k\sqrt{(1-\varepsilon_k)/\varepsilon_k}\le\sum_{s\in[S]}q_s x_{sk}$ in Proposition 2.

4.3. Out-of-Sample Test of Solutions

We next present the out-of-sample results on the solution of Ng (2015) and that of our model. Throughout the out-of-sample test, we use the mean $\boldsymbol{\mu}$ and baseline standard deviation $\bar{\boldsymbol{\sigma}}$ as assumed previously. We consider the following commonly-used distributions in our experiment.

- **Uni.** Uniform distribution: $\tilde{d}_k\sim\mathbb{U}(\mu_k-\sqrt{3}\sigma_k,\mu_k+\sqrt{3}\sigma_k)\ \forall k\in[K]$.
- **MN.** Multivariate normal distribution: $\tilde{\boldsymbol{d}}\sim\mathbb{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}=(\mathrm{diag}(\sigma_1^2,\cdots,\sigma_k^2)+\boldsymbol{\sigma}\boldsymbol{\sigma}^\top)/2$. Here, dependency between customer demands are captured by the term $\boldsymbol{\sigma}\boldsymbol{\sigma}^\top$.

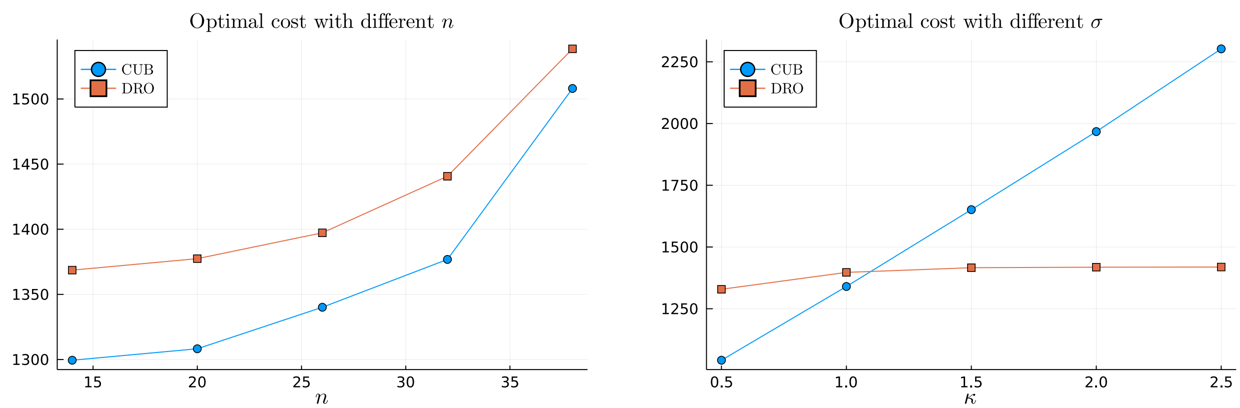

Figure 1 Perturbation analysis with different n and $\boldsymbol{\sigma} = \kappa\bar{\boldsymbol{\sigma}}$.

- **MLN.** Multivariate log-normal distribution: $\tilde{\boldsymbol{d}} \sim \mathbb{L}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$, where $\hat{\mu}_k = \ln(\mu_k^2/\sqrt{\mu_k^2+\sigma_k^2})$ and $\hat{\sigma}_k^2 = \ln(1+\sigma_k^2/\mu_k^2)$ in order to obtain mean and variance that are consistent with our assumption. Here, we recall that the mean and variance of a one-dimensional log-normal distribution $\mathbb{L}(\mu, \sigma^2)$ are $\exp(\mu+\sigma^2/2)$ and $(\exp(\sigma^2)-1)\exp(2\mu+\sigma^2)$, respectively. For the same reason as in the case of multivariate normal distribution case, we use $\hat{\boldsymbol{\Sigma}} = (\text{diag}(\hat{\sigma}_1^2, \cdots, \hat{\sigma}_k^2) + \hat{\boldsymbol{\sigma}}\hat{\boldsymbol{\sigma}}^\top)/2$.
- **T-Uni.** Truncated Uniform distribution: truncate the above uniform distribution so that the marginal support set of demand in each route is $[0, +\infty)$.
- **T-MN.** Truncated Normal distribution: truncate the normal distribution $\mathbb{N}(\mu_k, \sigma_k^2)$ $\forall k \in [K]$ to non-negative marginal support set $[0, +\infty)$. Note that here the demand in each route is assumed to be independent.

In particular, we look at two metrics for the out-of-sample test, namely average individual probability (AIP) and average joint probability (AJP) that are formally defined as follows:

$$\text{AIP} = \frac{1}{NK} \sum_{n\in[N]} \sum_{k\in[K]} \mathbb{I}(\hat{d}_{nk} < \sum_{s\in[S]} q_s x_{sk}) \times 100\%$$
$$\text{AJP} = \frac{1}{N} \sum_{n\in[N]} \mathbb{I}(\hat{d}_{nk} < \sum_{s\in[S]} q_s x_{sk} \ \forall k \in [K]) \times 100\%,$$

where $\hat{\boldsymbol{d}}_n = (\hat{d}_{nk})_{k\in[K]}, n \in [N]$ are randomly generated testing demand samples. By definition, AJP is not larger than AIP as the former is a stricter metric on probability satisfaction.

We test the optimal solutions as obtained in Tables 3, 4 and 5, using $50,000$ randomly generated demand samples from these distributions with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ as given in the corresponding table. It turns out that for all listed testing distributions, both models perform very well with respect to the risk threshold, attaining a probability of one in both metrics AIP and AJP. This is indeed expected: the distributionally robust chance constraint hedges against any distribution that satisfies partial distributional information specified in the ambiguity set.

	$\kappa=3$		$\kappa=6$		$\kappa=9$		$\kappa=3$		$\kappa=6$		$\kappa=9$	
	AIP	AJP	AIP	AJP	AIP	AJP	AIP	AJP	AIP	AJP	AIP	AJP
Uni	100.0	100.0	100.0	100.0	91.7	49.8	100.0	100.0	100.0	100.0	96.3	73.7
MN	100.0	100.0	98.5	91.6	93.1	71.0	100.0	100.0	99.1	94.5	94.8	76.4
MLN	99.9	99.3	94.1	74.5	85.7	52.9	99.9	99.6	95.4	78.9	87.3	56.2
T-Uni	100.0	100.0	100.0	100.0	91.3	47.9	100.0	100.0	100.0	100.0	96.2	72.8
T-MN	100.0	100.0	98.5	88.3	92.6	53.6	100.0	100.0	99.1	93.2	94.6	64.1

Table 6 AIP and AJP (in %) of optimal solutions of the CUB model (left) and the DRO model (right) in Tables 3, $\boldsymbol{\sigma}=\kappa\bar{\boldsymbol{\sigma}}$.

To further evaluate solution quality, we conduct an out-of-sample stress test where the standard deviation $\boldsymbol{\sigma}$ of a distribution that generates testing samples is larger than that in Table 3, that is, $\boldsymbol{\sigma}=\kappa\bar{\boldsymbol{\sigma}}$ with $\kappa\geq 1$ (the only exception is that we set $\hat{\sigma}_k=\kappa\sqrt{\ln(1+\bar{\sigma}_k^2/\mu_k^2)}$ $\forall k\in[K]$ in the case of MLN). Our consideration is two folds. First, the baseline standard deviation $\bar{\boldsymbol{\sigma}}$ in Table 3 is quite small, thus it would be interesting to see what is the situation when it becomes larger. Second, the distributionally robust chance constraint naturally guarantees probability satisfaction given the partial distributional information and we would like to see when the probability satisfaction will begin to diminish as the dispersion grows.

Results of $\kappa=3,6,9$ (representing demand deviation levels of low, median, and high, respectively) are reported in Table 6, while complete results for a larger range on the value of κ are presented in Figures 2-3 for uniform and multi-normal distributions. More results on different test distributions are relegated to Appendix A. Our model is consistently better than that of Ng (2015) and significantly outperforms in a wide range of large κ values. It is also worth mentioning that neither AIP nor AJP decreases from 100% until $\kappa\geq 3$. These encouraging results reveal the robustness of our proposed chance constrained model.

5. Conclusion

In this paper, we introduce to the vessel deployment problem recent advances in distributionally robust chance constraints, which provide a promising way to hedge against distributional ambiguity about shipping demand distribution, using support, mean, and dispersion information, as well as historical samples. High-quality solutions of our model can be effectively obtained via a sequential convex optimization algorithm or directly solving a mixed-integer linear program (in the data-driven approach). Computational studies show the encouraging and robust performance of our distributionally robust chance constrained models against the existing literature that is chiefly based on concentration inequalities and the union bound. In summary, although we chiefly focus

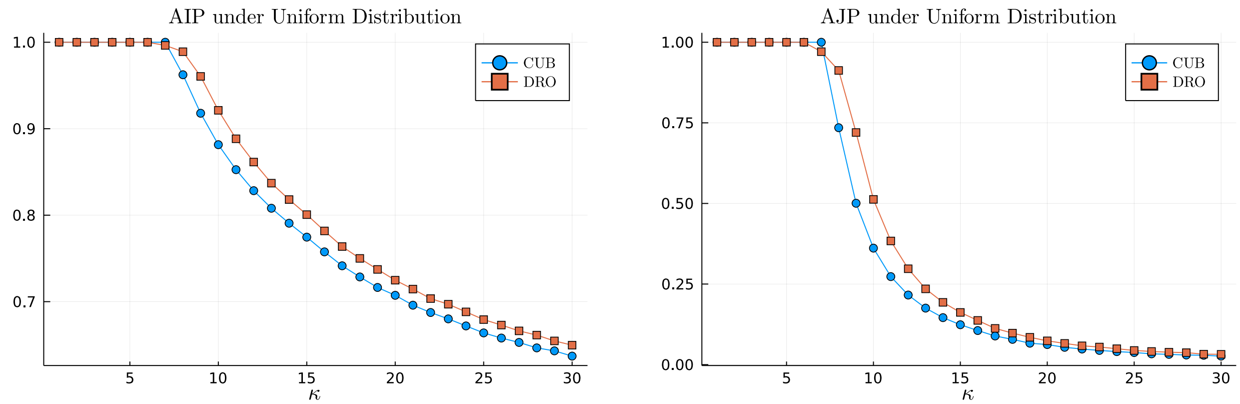

Figure 2 Out-of-sample stress test under uniform distribution.

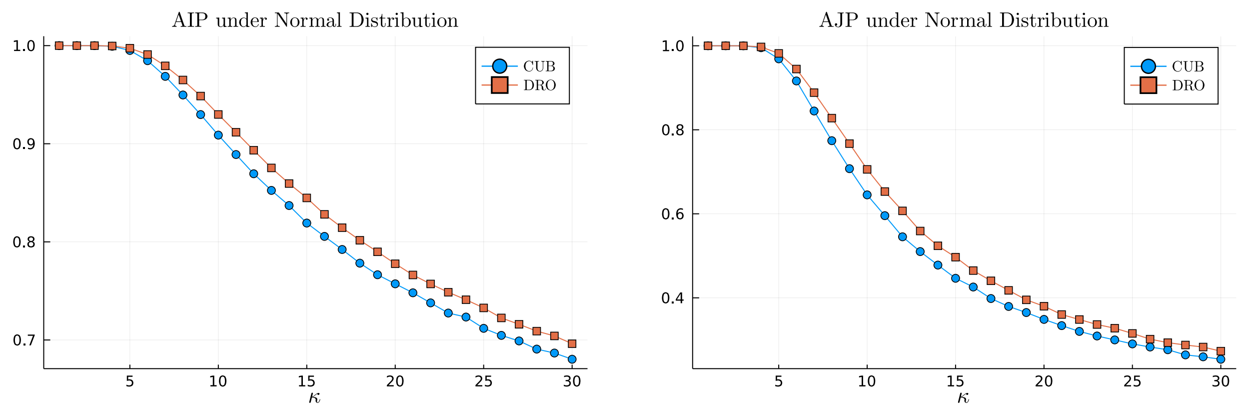

Figure 3 Out-of-sample stress test under normal distribution.

on a basic formulation of the fundamental vessel deployment problem, we firmly believe that the distributionally robust chance constraints remain open to more involved models that further include other features arising from practical operations in the maritime industry.

Acknowledgements

The authors are grateful to the review team for the thorough review and invaluable comments that lead to substantial improvement of the paper. This research was partially supported by the National Natural Science Foundation of China (Grants 72101187 and 72021002), the Fundamental Research Funds for the Central Universities as well as the Early Career Scheme of the Hong Kong Research Grants Council (CityU 21502820).

References

Ben-Tal, Aharon, Laurent El Ghaoui, Arkadi Nemirovski. 2009. *Robust optimization*, vol. 28. Princeton University Press.

Bertsimas, Dimitris, David B Brown, Constantine Caramanis. 2011. Theory and applications of robust optimization. *SIAM Review* **53**(3) 464–501.

Bertsimas, Dimitris, Melvyn Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.

Birge, John R, Francois Louveaux. 2011. *Introduction to stochastic programming*. Springer Science & Business Media.

Blanchet, Jose, Karthyek Murthy. 2019. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research* **44**(2) 565–600.

Casella, George, Roger L Berger. 2002. *Statistical inference*, vol. 2. Duxbury Pacific Grove, CA.

Chen, Zhi, Daniel Kuhn, Wolfram Wiesemann. 2018. Data-driven chance constrained programs over Wasserstein balls. *arXiv preprint arXiv:1809.00210*.

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

DeMiguel, Victor, Francisco J Nogales. 2009. Portfolio selection with robust estimation. *Operations Research* **57**(3) 560–577.

Gao, Rui, Anton J Kleywegt. 2016. Distributionally robust stochastic optimization with Wasserstein distance. *arXiv preprint arXiv:1604.02199*.

Ghosal, Shubhechyya, Wolfram Wiesemann. 2020. The distributionally robust chance-constrained vehicle routing problem. *Operations Research* **68**(3) 716–732.

Gorski, Jochen, Frank Pfeuffer, Kathrin Klamroth. 2007. Biconvex sets and optimization with biconvex functions: a survey and extensions. *Mathematical Methods of Operations Research* **66**(3) 373–407.

Hanasusanto, Grani A, Vladimir Roitch, Daniel Kuhn, Wolfram Wiesemann. 2015. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming* **151**(1) 35–62.

Hanasusanto, Grani A, Vladimir Roitch, Daniel Kuhn, Wolfram Wiesemann. 2017. Ambiguous joint chance constraints under mean and dispersion information. *Operations Research* **65**(3) 751–767.

Ho-Nguyen, Nam, Fatma Kılınç-Karzan, Simge Küçükyavuz, Dabeen Lee. 2021. Distributionally robust chance-constrained programs with right-hand side uncertainty under Wasserstein ambiguity. *Mathematical Programming* 1–32.

Jaramillo, DI, Anastassios N Perakis. 1991. Fleet deployment optimization for liner shipping Part 2. Implementation and results. *Maritime Policy and Management* **18**(4) 235–262.

Lee, Chung-Yee, Dong-Ping Song. 2017. Ocean container transport in global supply chains: Overview and research opportunities. *Transportation Research Part B: Methodological* **95** 442–474.

Levi, Retsef, Georgia Perakis, Joline Uichanco. 2011. Regret optimization for stochastic inventory models with spread information. *Working paper*.

Meng, Qiang, Shuaian Wang, Henrik Andersson, Kristian Thun. 2014. Containership routing and scheduling in liner shipping: overview and future research directions. *Transportation Science* **48**(2) 265–280.

Meng, Qiang, Tingsong Wang. 2010. A chance constrained programming model for short-term liner ship fleet planning problems. *Maritime Policy and Management* **37**(4) 329–346.

Meng, Qiang, Tingsong Wang, Shuaian Wang. 2012. Short-term liner ship fleet planning with container transshipment and uncertain container shipment demand. *European Journal of Operational Research* **223**(1) 96–105.

Mohajerin Esfahani, Peyman, Daniel Kuhn. 2018. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming* **171**(1) 115–166.

Nemirovski, Arkadi, Alexander Shapiro. 2007. Convex approximations of chance constrained programs. *SIAM Journal on Optimization* **17**(4) 969–996.

Ng, ManWo. 2014. Distribution-free vessel deployment for liner shipping. *European Journal of Operational Research* **238**(3) 858–862.

Ng, ManWo. 2015. Container vessel fleet deployment for liner shipping with stochastic dependencies in shipping demand. *Transportation Research Part B: Methodological* **74** 79–87.

Ng, ManWo, Dung-Ying Lin. 2018. Fleet deployment in liner shipping with incomplete demand information. *Transportation Research Part E: Logistics and Transportation Review* **116** 184–189.

Perakis, Anastassions N, DI Jaramillo. 1991. Fleet deployment optimization for liner shipping part 1. background, problem formulation and solution approaches. *Maritime Policy and Management* **18**(3) 183–200.

Postek, Krzysztof, Aharon Ben-Tal, Dick den Hertog, Bertrand Melenberg. 2018. Robust optimization with ambiguous stochastic constraints under mean and dispersion information. *Operations Research* **66**(3) 814–833.

Prékopa, András. 2003. Probabilistic programming. *Handbooks in operations research and management science* **10** 267–351.

Shapiro, Alexander. 2001. On duality theory of conic linear problems. *Semi-infinite programming*. Springer, 135–165.

Shapiro, Alexander, Darinka Dentcheva, Andrzej Ruszczyński. 2014. *Lectures on stochastic programming: modeling and theory*. SIAM.

Sirimanne, Shamika N, J Hoffman, W Juan, R Asariotis, M Assaf, G Ayala, A Ayoub, H Benamara, D Chantrel, J Hoffmann, A Larouche-Maltais, A Premti, et al. 2020. Review of maritime transport 2020. *Technical Report*.

Sirimanne, Shamika N, J Hoffman, W Juan, R Asariotis, M Assaf, G Ayala, H Benamara, D Chantrel, J Hoffmann, A Premti, et al. 2019. Review of maritime transport 2019. *Technical Report*.

Taylor, Michael AP. 2012. Network vulnerability in large-scale transport networks. *Transportation Research Part A: Policy and Practice* **46**(5).

van Ackooij, Wim. 2020. A discussion of probability functions and constraints from a variational perspective. *Set-Valued and Variational Analysis* **28**(4) 585–609.

Wang, Shuaian, Qiang Meng. 2017. Container liner fleet deployment: a systematic overview. *Transportation Research Part C: Emerging Technologies* **77** 389–404.

Xie, Weijun. 2019. On distributionally robust chance constrained programs with Wasserstein distance. *Mathematical Programming* 1–41.

Xie, Weijun, Shabbir Ahmed. 2018. On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM Journal on Optimization* **28**(2) 1151–1182.

Xie, Weijun, Shabbir Ahmed, Ruiwei Jiang. 2019. Optimized Bonferroni approximations of distributionally robust joint chance constraints. *Mathematical Programming* 1–34.

Zymler, Steve, Daniel Kuhn, Berç Rustem. 2013. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming* **137**(1-2) 167–198.

Appendix A: Supplementary Figures

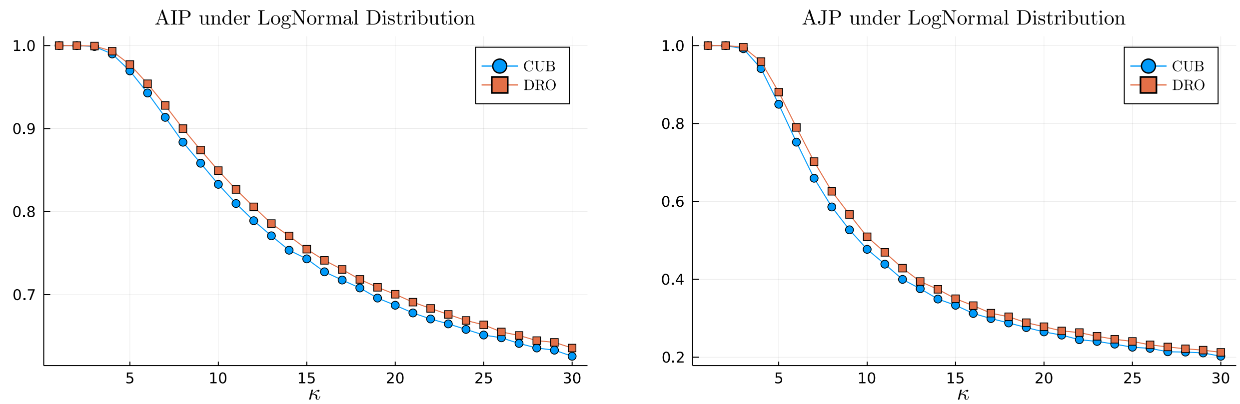

Figure 4 Out-of-sample stress test under lognormal distribution.

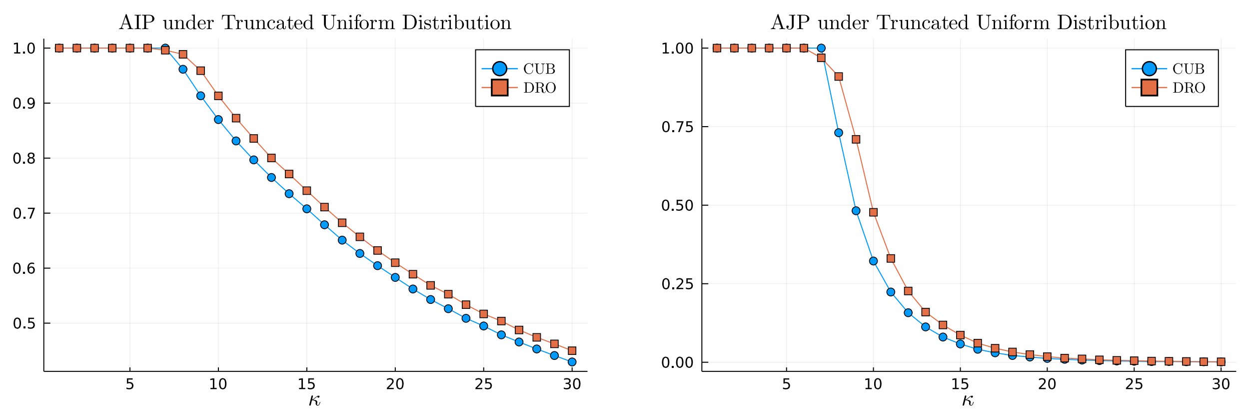

Figure 5 Out-of-sample stress test under truncated uniform distribution.

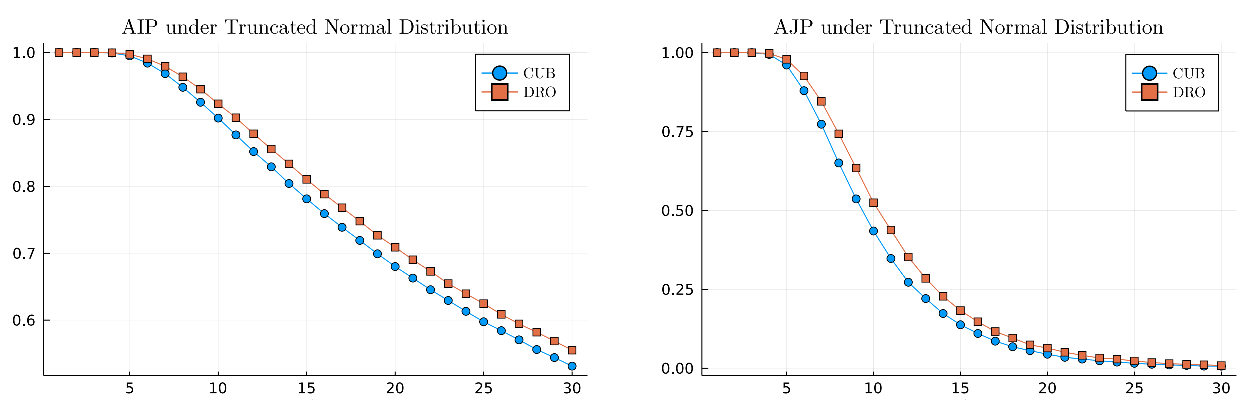

Figure 6 Out-of-sample stress test under truncated normal distribution.

Appendix B: Extension to The Data-Driven Approach

Despite the merits of the distributionally robust joint chance constrained model in Section 3, some issues are worth pointing out. First, for real-world applications, the corresponding inputs such as mean and dispersion usually need to be estimated from some first-hand operational data, *e.g.*, the United Nations International Trade Statistics Database (see a report by Sirimanne et al. 2020), and their estimates might be vulnerable. Second, the mean and dispersion ambiguity set has its fundamental limitation: it does not converge to a singleton set that contains only the true data-generating distribution, however large the historical data can be. Therefore, the adoption of data-driven approaches could be valuable for decision-makers in the maritime industry, guiding them to take advantage of the data source at hand and come up with better decisions. Yet, to the best of our knowledge, no data-driven model has been considered for the vessel deployment problem. This motivates us to introduce to the vessel deployment problem a data-driven distributionally robust chance constrained model, which theoretically captures the SAA approach as its special case.

In this section, we assume the decision-maker has a collection of historical demand samples $\hat{\boldsymbol{d}}_1, \hat{\boldsymbol{d}}_2, \cdots, \hat{\boldsymbol{d}}_N$, which give an empirical distribution $\hat{\mathbb{P}} = \frac{1}{N}\sum_{n=1}^{N} \delta_{\hat{\boldsymbol{d}}_n}$ that is uniformly supported on these samples. In particular, we adopt the data-driven approach proposed by Chen et al. (2018) and Xie (2019), based on the popular Wasserstein ambiguity set in the recent literature of distributionally robust optimization (see, *e.g.*, Gao and Kleywegt 2016, Mohajerin Esfahani and Kuhn 2018, Blanchet and Murthy 2019). Formally, the Wasserstein ambiguity set is given by

$$\mathcal{F}(\theta) = \left\{ \mathbb{P} \in \mathcal{P}(\mathbb{R}^n) \;\middle|\; \Delta(\mathbb{P}, \hat{\mathbb{P}}) \leq \theta \right\}.$$

Here, Δ is the type-1 Wasserstein distance between two distributions, defined as

$$\Delta(\mathbb{P}_1, \mathbb{P}_2) = \inf_{\mathbb{P} \in \mathcal{P}(\mathbb{P}_1, \mathbb{P}_2)} \mathbb{E}_{\mathbb{P}}[\|\tilde{\boldsymbol{d}}_1 - \tilde{\boldsymbol{d}}_2\|],$$

where $\|\cdot\|$ is a general norm, $\tilde{\boldsymbol{d}}_1 \sim \mathbb{P}_1, \tilde{\boldsymbol{d}}_2 \sim \mathbb{P}_2$, and $\mathcal{P}(\mathbb{P}_1, \mathbb{P}_2)$ denotes the set of all distributions on $\mathbb{R}^K \times \mathbb{R}^K$ with marginal distributions $\mathbb{P}_1$ and $\mathbb{P}_2$. Intuitively, the Wasserstein ambiguity set can be viewed as a ball of radius $\theta \geq 0$ in the space of probability distributions, which is centered around the empirical distribution.

Revisiting model (3) and replacing the ambiguity set with $\mathcal{F}(\theta)$, we obtain the following data-driven distributionally robust chance constrained model for the vessel deployment problem:

$$\begin{aligned} \min \quad & F(\boldsymbol{x}, \boldsymbol{y}) \\ \text{s.t.} \quad & \inf_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\left[\tilde{d}_k \leq \sum_{s \in [S]} q_s x_{sk} \quad \forall k \in [K] \right] \geq 1 - \varepsilon \\ & (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}. \end{aligned} \tag{16}$$

On the one hand, an attractive property of this data-driven approach is the theoretical finite-sample guarantee, which states that with a high confidence level $\mathcal{F}(\theta)$ will contain the underlying true probability distribution; see detailed discussions in Mohajerin Esfahani and Kuhn (2018). On the other hand, this data-driven model admits a nice mixed-integer reformulation: by proposition 2 of Chen et al. (2018), problem (16) is equivalent to a mixed-integer linear program as follows.

$$\begin{aligned} \min \quad & F(\boldsymbol{x},\boldsymbol{y}) \\ \text{s.t.} \quad & \varepsilon N\varphi - \boldsymbol{e}^\top \boldsymbol{\psi} \geq \theta N \\ & \rho_i + M z_i \geq \varphi - \psi_i && \forall i \in [N] \\ & M(1-z_i) \geq \varphi - \psi_i && \forall i \in [N] \\ & \sum_{s\in[S]} q_s x_{sk} - \hat{d}_{ik} \geq \rho_i && \forall k \in [K],\ \forall i \in [N] \\ & \boldsymbol{z} \in \{0,1\}^N, \boldsymbol{\psi} \in \mathbb{R}_+^N, \varphi \in \mathbb{R}, \boldsymbol{\rho} \in \mathbb{R}^N \\ & (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}, \end{aligned} \tag{17}$$

where M is a sufficiently large positive number.

An interesting observation is that when $\theta = 0$, the Wasserstein ambiguity set shrinks to a singleton, *i.e.*, $\mathcal{F}(\theta) = \{\hat{\mathbb{P}}\}$. One would expect that problem (16) then reduces[7] to the stochastic problem

$$\begin{aligned} \min \quad & F(\boldsymbol{x},\boldsymbol{y}) \\ \text{s.t.} \quad & \hat{\mathbb{P}}\left[\tilde{d}_k \leq \sum_{s\in[S]} q_s x_{sk} \quad \forall k \in [K]\right] \geq 1-\varepsilon \\ & (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}. \end{aligned} \tag{18}$$

Problem (18) can be viewed as an SAA approach to solve the basic stochastic formulation (1) by assuming a known empirical distribution of the uncertain shipping demands and it can be reformulated as a mixed-integer linear program

$$\begin{aligned} \min \quad & F(\boldsymbol{x},\boldsymbol{y}) \\ \text{s.t.} \quad & \boldsymbol{e}^\top \boldsymbol{z} \leq \varepsilon N \\ & \hat{d}_{ik} \leq \sum_{s\in[S]} q_s x_{sk} + M z_i \quad \forall k \in [K],\ \forall i \in [N] \\ & \boldsymbol{z} \in \{0,1\}^N \\ & (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}. \end{aligned} \tag{19}$$

Here, when $z_i = 0$, the condition in chance constraint of problem (18) is satisfied at sample point i, and the first constraint in problem (19) assures that such condition is met at least $(1-\varepsilon)N$ times.

Appendix C: Numerical Experiments in Data-Driven Setting

In this section we consider the data-driven setting. For both CUB and DRO models, we consider the corresponding data-driven counterparts where we estimate their parameters directly from data, while keeping other settings unchanged as in Section 4. We call the data-driven approach introduced in Section B as the

[7] Indeed, this reduction is true and we refer interested readers to Ho-Nguyen et al. (2021) on how problem (17) with $\theta = 0$ is equivalent to problem (19).

DDW model[8] and we denote the special case of $\theta = 0$ as the SAA model. For all the models, we generate $N = 100$ samples from the uniform distributions[9] with the baseline mean $\boldsymbol{\mu}$ and standard deviation $\bar{\boldsymbol{\sigma}}$, *i.e.*, $\tilde{d}_k \sim \mathbb{U}(\mu_k - \sqrt{3}\bar{\sigma}_k, \mu_k + \sqrt{3}\bar{\sigma}_k)\ \forall k \in [K]$. After generating the training samples, we estimate the mean and standard deviation by $\frac{1}{N}\sum_{n=1}^{N} \hat{\boldsymbol{d}}_n$ and $\sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(\hat{d}_{nk} - \hat{\mu}_{nk})^2}\ \forall k \in [K]$, respectively. Instead of using standard deviation as a substitute for mean absolute deviation in Section 4, we directly estimate it by $\frac{1}{N}\sum_{n=1}^{N} |\hat{d}_{nk} - \hat{\mu}_{nk}|\ \forall k \in [K]$, which will give a more decent solution. The above estimation process is straightforward and easy to implement, and we believe more sophisticated methods could be adopted for further studies. We fix $n = 26$ and $\varepsilon = 0.05$ in this section for all the experiments.

C.1. Comparison of Solutions

We replicate the experiments 50 times with different random seeds, *i.e.*, in each replication we run the same model but with different randomly generated training samples. Note that the runtime of CUB and DRO models do not depend on the sample size N because we only need to estimate their parameters from samples. In contrast, both DDW and SAA models would require more computational time as the sample size N grows. Therefore, we set $N = 100$ for all the models and we will provide the runtime analysis of DDW and SAA models separately.

In our experiments, the Wasserstein radius θ in the DDW model is found via cross validation based on the following: (*i*) as θ grows, the optimal cost of DDW model increases; (*ii*) the larger θ is, the more conservative DDW model is, and finally it becomes infeasible when θ goes beyond 8900. These observations show the trade-off between cost and robustness when tuning an appropriate value of θ. Therefore, we run the experiments for $\theta \in \{1000, 2000, \cdots, 8000\}$ in the background and select the one $\theta = 3000$ that performs the best. For more detailed discussions on choosing the Wasserstein radius, we refer to Chen et al. (2018) and Mohajerin Esfahani and Kuhn (2018).

In Table 7, we report for the four models several statistics of the optimal cost and runtime over 50 random instances. Although in terms of the objective value, the SAA model appears to be advantageous over other models, it has a disappointing performance in out-of-sample stress tests as we will show shortly. In general, the objective value of the DWW model is between that of the DRO or CUB model, and it is only larger than that of the CUB model within 3% when compared in terms of the mean or 50^{th} quantile. On the other hand, in terms of the solution stability with respect to training data, the DDW model performs better than DRO and CUB models; see standard deviation of the objective value.

C.2. Out-of-Sample Test of Solutions

We follow the same treatment on the test distributions as in Section 4.3. Figure 7 and Figure 8 show the out-of-sample stress test of all models with certain statistics over 50 random instances under uniform and multi-normal distributions, respectively. Results of other test distributions are omitted since the information conveyed is similar.

[8] The “DD” stands for the term “data-driven” and “W” stands for the word “Wasserstein”.

[9] Indeed, we do not observe any qualitative changes in the comparative results when different distributions are adopted for data generation.

	Objective (million)					Runtime (seconds)				
	5%	50%	95%	Mean	Std	5%	50%	95%	Mean	Std
CUB	1323.41	1343.61	1360.38	1342.85	11.02	1.15	2.95	10.86	4.31	4.90
DRO	1378.62	1388.05	1402.75	1389.00	7.72	92.21	147.90	286.36	163.97	63.55
SAA	911.26	913.52	914.24	912.97	0.95	7.25	9.14	14.31	9.78	2.98
DDW	1366.63	1377.92	1384.84	1376.75	5.28	120.03	120.04	120.06	118.20	15.33

Table 7 We report the 5%, 50%, 95% quantiles, mean and standard deviation over 50 random instances.

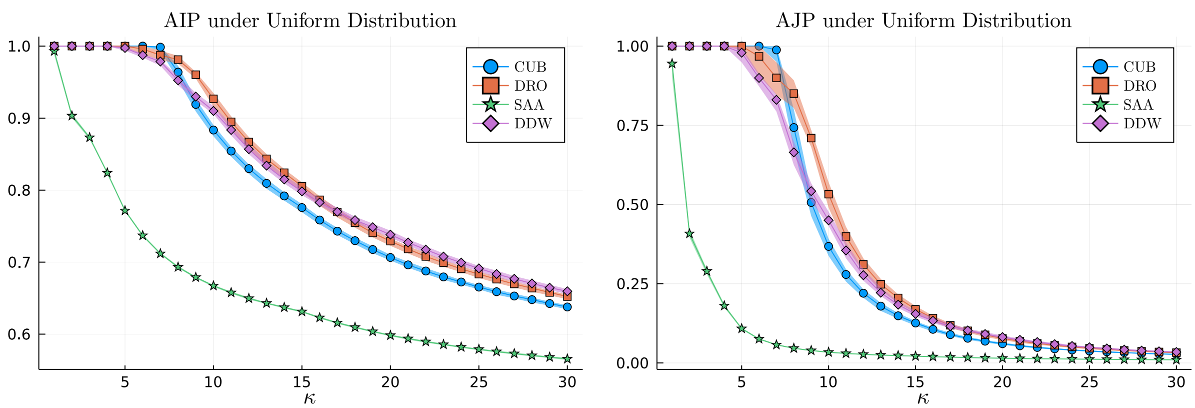

Figure 7 Out-of-sample stress test under uniform distribution. The solid line represents the median cost over 50 random instances and the shaded regions cover the 5% to 95% quantiles.

It can be seen that performances of CUB and DRO models are similar to those in Section 4.3. Moreover, the narrow shaded regions imply that even for a small sample size $N = 100$ the data-driven CUB and DRO models still yield a stable performance in out-of-sample stress tests when κ is not too large. On the contrary, the SAA model quickly becomes inferior after $\kappa > 1$ and works consistently worse than the other approaches. Interestingly, the DDW model does not outperform the CUB and DRO models at the beginning but it could catch up when κ increases.

C.3. Runtime Analysis

We first provide the runtime of CUB and DRO models over 50 random instances in Figure 9. Since the estimated parameters are different for various random seeds, the result is sufficient to serve as evidence for the general runtime analysis of both CUB and DRO models. It is shown that the optimal solution of the DRO model could be found within a few minutes, which we believe is acceptable for many real-world applications.

One of the major obstacles for using DDW model is the increasing runtime when the sample size becomes larger, which is also an interesting open research question. As mentioned in Chen et al. (2018), it is observed that the standard solver (*e.g.*, CPLEX) could find a decent solution with a small optimality gap in a short period of time, while it would spend plenty of effort in tightening the optimality gap or certifying the optimality. Such phenomena also occur in our numerical experiments, we therefore follow the strategy in

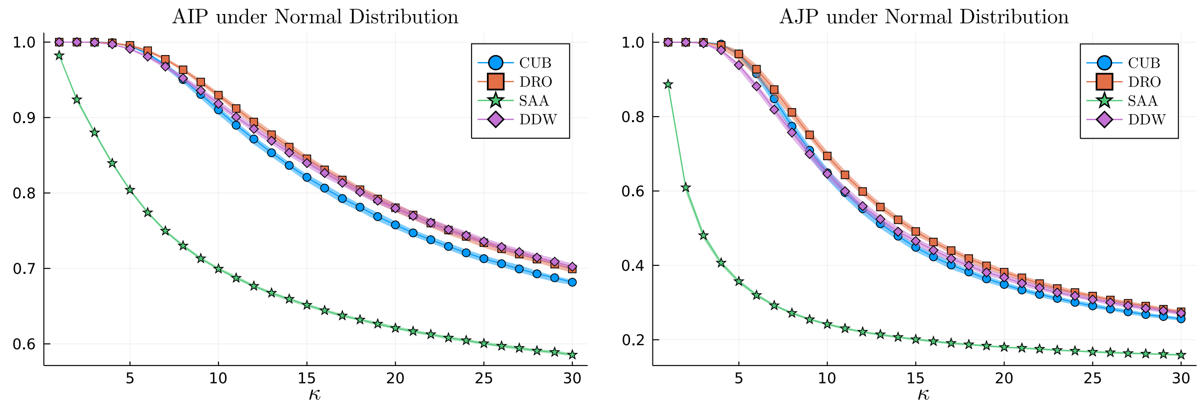

Figure 8 Out-of-sample stress test under the normal distribution. The solid line represents the median cost over 50 random instances and the shaded regions cover the 5% to 95% quantiles.

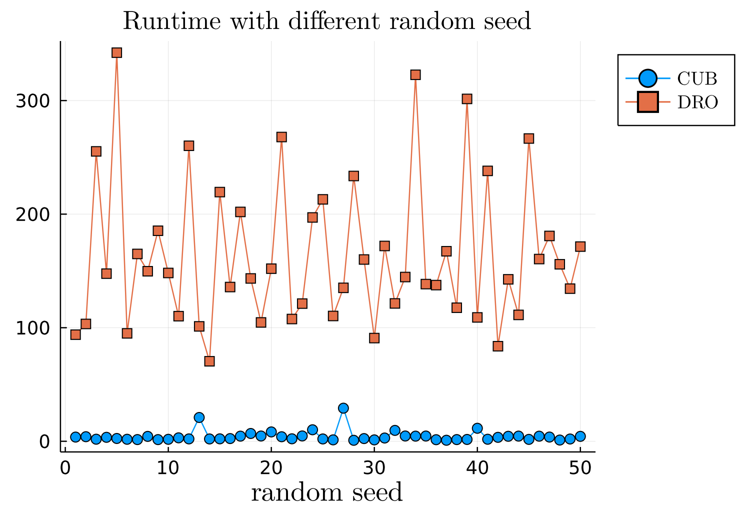

Figure 9 Runtime of CUB and DRO models with different random seeds.

Chen et al. (2018) by setting the time limit to 120 seconds with 80 seconds for branch-and-bound and 40 seconds for solution "polishing". It is worth noting that a recent work (Ho-Nguyen et al. 2021) provides efficient approaches to solve the general distributionally robust chance constrained models where the runtime for some problems could be largely reduced.

For the SAA model, the runtime will increase but the optimal cost is less affected as the number of samples N grows. We set $N \in \{100, 150, \cdots, 400\}$ and provide the numerical evidence in Figure 10. As N increases, the runtime increases as expected. Although the shaded region shrinks, the median of optimal cost does not change too much in our problem. Hence, we choose $N = 100$ for the experiments presented in Table 7 as it is sufficient to obtain a decent solution and much more economical runtime.

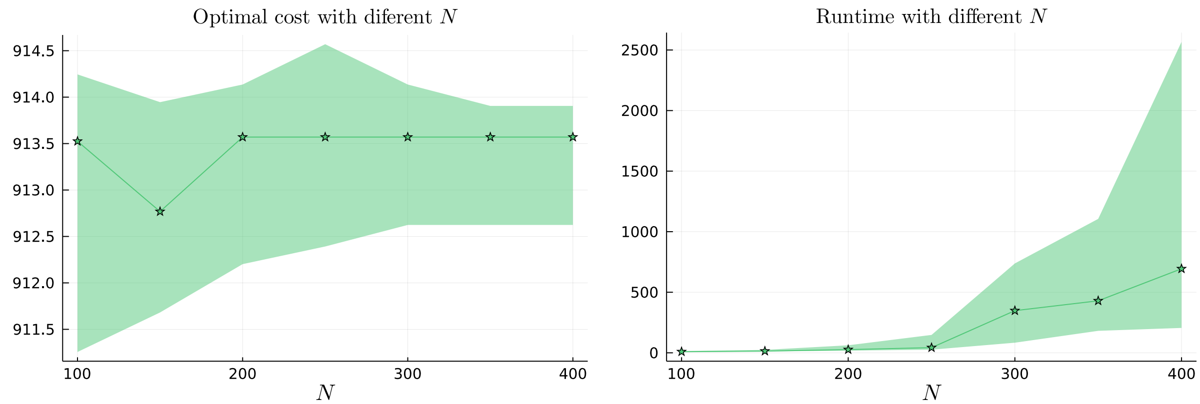

Figure 10 The optimal cost (million) and runtime (seconds) of the SAA model with different sample sizes. The solid line represents the median cost over 50 random instances and the shaded regions cover the 5% to 95% quantiles.