Math. Program., Ser. A (2019) 175:263–305
https://doi.org/10.1007/s10107-018-1243-y

FULL LENGTH PAPER

Routing optimization with time windows under uncertainty

Yu Zhang[1,2] · Roberto Baldacci[3] · Melvyn Sim[4] · Jiafu Tang[5]

Received: 14 June 2017 / Accepted: 29 January 2018 / Published online: 23 February 2018

Abstract We study an a priori Traveling Salesman Problem with Time Windows (TSPTW) in which the travel times along the arcs are uncertain and the goal is to determine within a budget constraint, a route for the service vehicle in order to arrive at the customers' locations within their stipulated time windows as well as possible. In particular, service at customer's location cannot commence before the beginning of the time window and any arrival after the end of the time window is considered late and constitutes to poor customer service. In articulating the service level of the TSPTW under uncertainty, we propose a new decision criterion, called the *essential riskiness index*, which has the computationally attractive feature of convexity that enables us

✉ Roberto Baldacci
r.baldacci@unibo.it

Yu Zhang
waltyuzhang@gmail.com

Melvyn Sim
melvynsim@gmail.com

Jiafu Tang
tangjiafu@dufe.edu.cn

1 School of Business Administration, Southwestern University of Finance and Economics, Chengdu 611130, People's Republic of China

2 Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore 119245, Republic of Singapore

3 Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, 47521 Cesena, Italy

4 Department of Analytics and Operations, National University of Singapore, Singapore 119245, Republic of Singapore

5 College of Management Science and Engineering, Dongbei University of Finance and Economics, Dalian 116000, People's Republic of China

to formulate and solve the problem more effectively. As a decision criterion for articulating service levels, it takes into account both the probability of lateness and its magnitude, and can be applied in contexts where either the distributional information of the uncertain travel times is fully or partially known. We propose a new formulation for the TSPTW, where we explicitly express the service starting time at each customer's location as a convex piecewise affine function of the travel times, which would enable us to obtain the tractable formulation of the corresponding distributionally robust problem. We also show how to optimize the essential riskiness index via Benders decomposition and present cases where we can obtain closed-form solutions to the subproblems. We also illustrate in our numerical studies that this approach scales well with the number of samples used for the sample average approximation. The approach can be extended to a more general setting including Vehicle Routing Problem with Time Windows with uncertain travel times and customers' demands.

Keywords Vehicle routing problem · Uncertain travel time · Time windows · Risk and ambiguity · Distributionally robust optimization

Mathematics Subject Classification 90C15 · 49M27 · 90C11 · 90C22

1 Introduction

In the deterministic setting, Dantzig and Ramser [14] are first to introduce the vehicle routing problem (VRP) as an extension to the Traveling Salesman Problem (TSP) to optimize routes for a fleet of vehicles for the purpose of delivering goods to or collecting them from customers at various locations. Due to its practical importance, despite the computational challenges, the VRP has received much attentions from both industry practitioners and academic researchers (see, for instance, [27,40]). Among the variants of the VRP family, the Capacitated VRP (CVRP) and the VRP with Time Windows (VRPTW) are most prominent and have been studied extensively (see, for instance, [7,16]). In the former problem, a fleet of identical vehicles located at a central depot are optimally routed to fulfill the demands of the customers subject to vehicular capacity constraints. VRPTW generalizes the CVRP by further imposing that each customer has to be served within a specified time interval, known as the time window. Specifically, if a vehicle arrives early, service cannot be rendered until the commencement of the time window. Furthermore, any arrival after the end of the time window or deadline would be prohibited. Notably, the VRP with Deadline (VRPD) is an important special case of VRPTW, where the customers' time windows are only specified by deadlines. When only one vehicle route is involved, we refer to the VRPTW and the VRPD respectively as the TSP with Time Windows (TSPTW) and the TSP with Deadlines (TSPD).

Since TSP, the simplest form of VRP, is already $\mathcal{NP}$-hard, incorporating uncertainty in VRP would further elevate the computational difficulties of solving the problem. Hence, most routing optimization problems considered in the VRP literature assume deterministic travel times. For the state-of-the-art exact algorithms for deterministic VRP, we refer readers to Baldacci et al. [6], Toth and Vigo [40], Pecin et al. [31,32], among others. Nevertheless, from the modeling perspective, ignoring uncertainty at the planning level can potentially result in poor routing decisions leading to, among other things, poor customer services that could adversely impact on the reputation of the company. Hence, approaches that could mitigate uncertainty in VRP are highly desired.

Several models and algorithms have been proposed in the literature to mitigate uncertain travel times in TSPD and VRPD. Kao [25] develops a preference ordered dynamic programming approach to address a stochastic TSPD that minimizes the probability of late completions. However, as noted by Sniedovich [38], the approach may yield suboptimal solutions. Laporte et al. [28] propose a chance-constrained model (see, for instance, [12]) for a stochastic VRPD, which minimizes the operating cost while limiting the probability of late returns of vehicles to the depot. They present instances, such as having normally distributed travel times, under which this model can be transformed to a deterministic optimization problem. They also propose models that minimize the operating and expected tardiness costs, and provide a branch-and-cut procedure to solve these models via sample average approximations. Kenyon and Morton [26] propose two branch-and-cut procedures for addressing a stochastic VRPD that minimizes the expected completion time of the vehicle fleet or its tardiness probability. The first approach is capable of obtaining exact solutions when the sample size is small, while the second approach could obtain approximate solutions for larger problem instances. Verweij et al. [42] provide a computational study of the decomposition and branch-and-cut techniques for solving a stochastic TSPD via sample average approximation, in which the objective is to minimize the operating and expected tardiness costs. Taş et al. [39] propose a column generation procedure for solving the stochastic VRPTW that minimizes sum of the operating costs and expected penalty costs of time window violations. Note that the enforcement of time windows in this model is "soft" in the sense that service commencement is permitted even if the vehicle arrives early at a customer's location.

The literature on stochastic VRP with "hard" time windows is rather limited. The resulting optimization problem is considerably more difficult than the case with soft time windows, since the probability distributions of arrival times at customers locations have to be truncated because of the hard time windows. Errico et al. [18] consider the VRP with hard time windows and stochastic service times and formulated it as a chance-constrained optimization model that includes a probabilistic constraint. The same problem has been recently addressed by Errico et al. [17] and modelled as a two-stage stochastic program using two recourse policies to recover operations feasibility when the first stage plan turns out to be infeasible. The problem has been formulated as a set partitioning problem and solved by an exact branch-cut-and-price algorithm.

A review of the literature on the stochastic VRP including stochastic demands, customers and travel times, and a concise description of relevant solution concepts is found in Gendreau et al. [20].

Robust optimization techniques have also been applied to address VRP with uncertain travel times. Based on the so called budgeted uncertainty sets of Bertsimas and Sim [9], Lee et al. [29] study a robust VRPD which ensures feasible schedules for all uncertainty arising from the uncertainty sets, while minimizing the total travel costs. Agra et al. [3] extend this approach to investigate a robust VRPTW.

Our approach to address the VRPTW under uncertainty is inspired by a recent work of Jaillet et al. [24], who propose a new decision criterion that can be applied to solve a variety of VRPs under uncertainty. In particular, they motivate and adopt a relatively new targeted oriented decision criterion that is based on the *riskiness index* of Aumann and Serrano [5] and has been generalized by Brown and Sim [11] from the perspective of satisficing in decsion making. The decision criterion is a convex function that penalizes the risks of constraints' violations by accounting for both their infeasibility probabilities and magnitudes of violations. Apart from the coherency in decision making, the convexity of the decision criterion has important ramifications in solving the TSP, that would lead to better performance in computational studies against an approach that maximizes the feasibility probability (see also computational results in Adulyasak and Jaillet [1]). However, although the riskiness index can be adopted for distributionally robustness, where probability distributions of the uncertain parameters are not fully characterized, to achieve computationally tractable models, it requires the uncertain outcomes to be expressed as affine summations of independently random variables. Hence, while the approach can be applied to address uncertain VRP in various settings including those with deadlines, soft time windows and uncertain demands (see [21]), as we will further explain, it is however, incapable of extending to VRPTW where service commencement is not permitted if the vehicle arrives earlier at a customer's location.

Inspired by this work, we focus on a new decision criterion that has similar properties to the riskiness index and develop new solution approach that can be used to address an uncertainty VRPTW. For notational simplicity and clarity of the exposition, we will focus on a TSPTW in the context of uncertain travel times and the goal is to determine within a budget constraint, a route for the service vehicle to traverse in order to arrive at the customers' locations within their stipulated time windows as well as possible. We can generalize the approach to address VRPTW under uncertainty including those with uncertain demands. For such extensions, we will refer interested readers to Jaillet et al. [24]. Our distinct contributions in this paper are as follows:

– We propose and motivate a new decision criterion known as the *essential riskiness index* that can be applied to address TSPTW with uncertain travel times. Similar to the riskiness index of Aumann and Serrano [5], the essential riskiness index takes into account of both the probability of deadline violations and its magnitude of such violations whenever they occur. It can also be applied in contexts where either the distributional information of the uncertain travel times is fully or partially known. Quite apart from the Aumann and Serrano [5] riskiness index, the essential riskiness enables us to model correlation naturally, without having to impose stochastic independence in the underlying uncertain parameters. Moreover, an important feature of the new decision criterion is the ability to provide computationally amenable reformations when the risk is in the form of convex piecewise affine functions of the underlying random variables, which enables us to formulate and solve the TSPTW more effectively.
– To incorporate the essential riskiness index in the TSPTW formulation, we adopt a multi-commodity flow formulation that enables us to explicitly characterize the

time for service commencement at each node in the form of a convex piecewise linear function of the travel times. Among other benefits, this approach also enables us to formulate the distributionally robust aspect of the TSPTW explicitly and compactly.

– We propose Benders decomposition methods for solving the stochastic and the distributionally robust versions of the TSPTW. For solving the stochastic TSPTW via sample average approximation and the distributionally robust TSPD with known mean and covariance of travel times, we exploit the problem structures and obtain closed form solutions for the Benders subproblems. Our computational studies suggest that the decomposition approach scales well computationally with the problem size.

The remainder of this paper is organized as follows. In Sect. 2, we introduce the multi-commodity flow formulation that can be applied to model a TSPTW with uncertain travel times. In Sect. 3, we introduce the essential riskiness index as a coherent decision criterion that penalizes late arrivals by accounting for both their tardiness probabilities and magnitudes. In Sect. 4, we apply the essential riskiness index to address a stochastic TSPTW via sample average approximation and a distributionally robust TSPTW with mean and covariance ambiguity set. In Sect. 5 we provide an efficient and scalable Benders decomposition method to solve the problems computationally and show that in some cases closed-form solutions to the subproblems can be obtained. We provide an extension in Sect. 6 and the computational studies in Sect. 7. Finally, we conclude the paper and indicate future research directions in Sect. 8.

Notation

We adopt the following notations throughout the paper. We denote by $|\mathcal{N}|$ the cardinality of a set $\mathcal{N}$. We use boldface lowercase letters to represent vectors and $\boldsymbol{x}'$ to represent the transpose of a vector $\boldsymbol{x}$, for example, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$. We use tilde ($\tilde{\cdot}$) to denote uncertain parameters. We model uncertainty by a state-space Ω and a $\sigma-$algebra $\mathcal{F}$ of events in Ω. We define $\mathcal{V}$ as the space of real-valued random variables and $\mathbb{R}$ the space of real numbers. In the distributionally robust optimization model, instead of specifying the true distribution $\mathbb{P}$ on $(\Omega, \mathcal{F})$, we assume that it belongs to a distributional uncertainty set $\mathbb{F}$, such that $\mathbb{P} \in \mathbb{F}$. We denote by $\mathbb{E}_{\mathbb{P}}\left(\tilde{t}\right)$ the expectation of $\tilde{t}$ under probability distribution $\mathbb{P}$. The inequality between two uncertain parameters $\tilde{t} \geq \tilde{v}$ describes state-wise dominance, i.e., $\tilde{t}(\omega) \geq \tilde{v}(\omega)$ for all $\omega \in \Omega$. The inequality between two vectors $\boldsymbol{x} \geq \boldsymbol{y}$ corresponds to the element-wise comparison.

2 A model for TSPTW with travel times uncertainty

To formulate the TSPTW, we consider a directed network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N} = \{1, 2, \ldots, n\}$ represents the set of nodes and $\mathcal{A}$ the set of arcs. Let node 1 be the origin depot, node n be the destination depot and the rest of the nodes i, $i \in \mathcal{N}\backslash\{1, n\}$ represent customers at various locations. We use (i, j) and a interchangeably to represent an arc in $\mathcal{A}$. Let c_a denote the cost for traversing arc $a \in \mathcal{A}$ and let $\tilde{z}_{ij}$, $(i, j) \in \mathcal{A}$, be the consolidated random variable associated with the random service time at node i and random travel time for traversing arc (i, j). For each node $i \in \mathcal{N}$, we denote, respectively, the set of its incoming arcs and the set of its outgoing arcs by

$$\delta^-(i) = \{(j,i) \in \mathcal{A} \mid j \in \mathcal{N}\backslash\{i\}\} \quad \text{and} \quad \delta^+(i) = \{(i,j) \in \mathcal{A} \mid j \in \mathcal{N}\backslash\{i\}\}.$$

We assume that $\delta^-(1) = \delta^+(n) = \emptyset$ and that $\mathcal{A}$ does not contain arc $(1, n)$.

The service vehicle departs at the origin depot at time zero. At each customer's node i, $i \in \mathcal{N}$, the *time window* for service commencement is denoted by $[\underline{\tau}_i, \overline{\tau}_i]$, where $\underline{\tau}_i \in \mathbb{R}_+$ is the *earliest time* for service commencement and $\overline{\tau}_i \in \mathbb{R}_+ \cup \{\infty\}$ represents the *latest time* or *deadline*. Not all customers have clearly specified service time windows. In the absence of earliest time or deadline, we would let $\underline{\tau}_i = 0$ or $\overline{\tau}_i = \infty$, respectively. At the origin node, we have $\underline{\tau}_1 = 0$ and $\overline{\tau}_1 = \infty$. We denote the set of nodes with positive earliest times by

$$\underline{\mathcal{N}} = \{i \in \mathcal{N} \mid \underline{\tau}_i > 0\}$$

and the set of nodes with finite deadlines by

$$\overline{\mathcal{N}} = \{i \in \mathcal{N} \mid \overline{\tau}_i < \infty\}.$$

In the TSPTW model, if the vehicle arrives early at node i, $i \in \underline{\mathcal{N}}$, we do not allow the service to commence until the time $\underline{\tau}_i$. However, service would still be rendered if the vehicle arrives late after the deadline, $\overline{\tau}_i$. Although late services are inevitable, they could be mitigated through appropriate choice of route by solving an optimization problem that takes into account of the network uncertainty.

To formulate the model, we let $\mathcal{X} \subseteq \{0,1\}^{|\mathcal{A}|}$ represent the set of feasible routes, where each feasible route is a Hamiltonian path that starts from node 1, visits each node i, $i \in \mathcal{N}\backslash\{1, n\}$ exactly once, and ends at node n. Given a feasible route, $\boldsymbol{x} = (x_a)_{a\in\mathcal{A}} \in \mathcal{X}$, we denote $x_a = 1$ if arc a is in the route, and $x_a = 0$, otherwise. Next, we characterize the feasible set of routes for the TSPTW that would enable us to determine the service starting time at each node i, $i \in \mathcal{N}$. To do so, we adopt the multi-commodity flow formulation proposed by Claus [13] for the Asymmetric Traveling Salesman Problem by defining the set

$$\mathcal{S} = \left\{ \begin{array}{l|ll} & \sum_{a\in\delta^+(i)} x_a = 1, & i \in \mathcal{N}\backslash\{n\}, \\ & \sum_{a\in\delta^-(i)} x_a = 1, & i \in \mathcal{N}\backslash\{1\}, \\ & s_a^1 = 0, & a \in \mathcal{A}, \\ \boldsymbol{x} \in \{0,1\}^{|\mathcal{A}|} & s_a^l \le x_a, & l \in \mathcal{N}\backslash\{1\}, a \in \mathcal{A}, \\ \boldsymbol{s} \in \mathbb{R}_+^{|\mathcal{A}|\times|\mathcal{N}|} & \sum_{a\in\delta^+(1)} s_a^l = 1, & l \in \mathcal{N}\backslash\{1\}, \\ & \sum_{a\in\delta^+(i)} s_a^l - \sum_{a\in\delta^-(i)} s_a^l = 0, & l \in \mathcal{N}\backslash\{1\}, i \in \mathcal{N}\backslash\{1, l\}, \\ & \sum_{a\in\delta^+(l)} s_a^l - \sum_{a\in\delta^-(l)} s_a^l = -1, & l \in \mathcal{N}\backslash\{1\}, \end{array} \right\}.$$

where we denote $\boldsymbol{s} = (\boldsymbol{s}^l)_{l \in \mathcal{N}}$ and $\boldsymbol{s}^l \in \mathbb{R}_+^{|\mathcal{A}|}$, $l \in \mathcal{N}$. For extension to VRPTW and to the case with uncertain demand, we refer interested readers to Adulyasak and Jaillet [1] and Jaillet et al. [24].

Proposition 1 *The set of feasible routes is given by*

$$\mathcal{X} = \left\{ \boldsymbol{x} \;\middle|\; (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S} \text{ for some } \boldsymbol{s} \in \mathbb{R}_+^{|\mathcal{A}| \times |\mathcal{N}|} \right\}.$$

Moreover, for any feasible $\boldsymbol{x} \in \mathcal{X}$, there is an unique $\boldsymbol{s} \in \{0, 1\}^{|\mathcal{A}| \times |\mathcal{N}|}$ such that $(\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S}$. In particular, for all $l \in \mathcal{N}$, $\boldsymbol{s}^l$ corresponds to the path on route $\boldsymbol{x}$ that starts at node 1 and ends at node l.

Proof See, for instance, Jaillet et al. [24]. □

The vehicle departs the origin node at $t_1 = 0$. Let $\boldsymbol{z} \in \mathbb{R}_+^{|\mathcal{A}|}$ denote a realization of $(\tilde{z}_{ij})_{(i,j) \in \mathcal{A}}$. For a given solution $\boldsymbol{x} \in \mathcal{X}$, we can determine the service starting time recursively as

$$t_j = \max\{t_i + z_{ij}, \underline{\tau}_j\} \tag{1}$$

for every arc (i, j) along the path $\boldsymbol{x}$, i.e., $x_{ij} = 1$. Hence, if the arrival time at node j, is earlier than $\underline{\tau}_j$, the service starting time would be $\underline{\tau}_j$. Otherwise, the service starting time would coincide with the arrival time given by $t_i + z_{ij}$. Next, we show that the service starting time can be represented as a convex piecewise linear function of $\boldsymbol{z}$ as follows.

Proposition 2 *Given $(\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S}$ and a realization of $\tilde{\boldsymbol{z}}$, denoted by $\boldsymbol{z}$, the service starting time for each node $l \in \mathcal{N}$ is determined by the function*

$$t_l(\boldsymbol{s}, \boldsymbol{z}) = \max_{k \in \underline{\mathcal{N}} \cup \{1\}} \left\{ \sum_{a \in \delta^-(k)} s_a^l \underline{\tau}_k + \boldsymbol{z}' \left(\boldsymbol{s}^l - \boldsymbol{s}^k \right) \right\}. \tag{2}$$

Proof See Appendix A.1. □

Corollary 1 *The* TSPTW *with deterministic travel times, $\boldsymbol{z}$ can be formulated as*

$$\begin{aligned} &\min \ \boldsymbol{c}'\boldsymbol{x} \\ &\text{s.t.} \sum_{a \in \delta^-(k)} s_a^l \underline{\tau}_k + \boldsymbol{z}' \left(\boldsymbol{s}^l - \boldsymbol{s}^k \right) \le \overline{\tau}_l, \ \forall k \in \underline{\mathcal{N}} \cup \{1\}, l \in \overline{\mathcal{N}}, \\ &\quad (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S}. \end{aligned} \tag{3}$$

Remark As far as we know, this formulation of the deterministic TSPTW is new. It is polynomial in size, $O(|\mathcal{N}|^3)$, and the decision variables $(\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S}$ are associated with the choice of route but not the travel times along it. It avoids "big-M" and does not require decision variables associated with the departure and arrival times as in Fisher et al. [19] nor discrete travel times as in Dash et al. [15]. More importantly, as we will show in Sect. 4, by expressing the service starting time at each customer's location as a convex piecewise affine function of the travel times, we are able to obtain tractable formulation for the corresponding distributionally robust problem. Indeed, an alternative idea to enable a convex piecewise affine expression of the service starting time, and thus a tractable distributionally robust formulation, has been proposed by Agra et al. [2]; they express the underlying graph as a layered graph and present a tractable formulation for the robust VRPTW, which is also of size $O(|\mathcal{N}|^3)$. We leave the in-depth comparison of our formulation and theirs for future research. Although there are other proposed TSPTW models in the literature (see, for instance, [7,15,19]), we are unaware of any method that can adopt or extend them to obtain tractable formulations when addressing uncertain travel times.

We also define the *delay function* at node i, $i \in \overline{\mathcal{N}}$ as

$$\xi_i(\boldsymbol{s}, \boldsymbol{z}) = t_i(\boldsymbol{s}, \boldsymbol{z}) - \overline{\tau}_i$$

so that an arrival at node i is late if and only if $\xi_i(\boldsymbol{s}, \boldsymbol{z}) > 0$. Observe that the delay function is a convex piecewise affine function of $\boldsymbol{z}$. Correspondingly, we denote the function map of uncertain delays by $\boldsymbol{\xi}(\boldsymbol{s}, \tilde{\boldsymbol{z}}) = (\xi_i(\boldsymbol{s}, \tilde{\boldsymbol{z}}))_{i \in \overline{\mathcal{N}}}$. To improve customer service, the optimization problem should penalize tardiness and ensure that the service vehicle could arrive at the customers' locations within their stipulated time windows as well as possible. Hence, a natural approach would be to determine the route within the budget constraint that maximizes the joint probability of punctual arrivals as follows.

$$\begin{aligned} \max \ & \mathbb{P}(\boldsymbol{\xi}(\boldsymbol{s}, \tilde{\boldsymbol{z}}) \leq \boldsymbol{0}) \\ \text{s.t.} \ & \boldsymbol{c}'\boldsymbol{x} \leq B, \\ & (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S}. \end{aligned} \tag{4}$$

However, as articulated in Jaillet et al. [24], the decision criterion associated with Problem (4) may not necessarily be well justified. Among other things, the decision criterion captures only the frequency of tardiness and completely ignores the magnitude of delays. Conceivably, if the tardiness probabilities associated with two random arrivals are the same, the one with a potential delay of 5 min may have the same preference as the alternative with a delay of 50 min. Moreover, since the objective to be maximized is not a concave function, solving Problem (4), even via sampling average approximations, can be computationally excruciating.

3 Essential riskiness index

To quantify the risk associated with the violation of deadlines, Jaillet et al. [24] adopt a different objective function that is based on the riskiness index of Aumann and Serrano [5].

Definition 1 Given a random delay denoted by the random variable $\tilde{\xi} \in \mathcal{V}$ with probability distribution $\mathbb{P}$, the riskiness index $\rho_R\left(\tilde{\xi}\right) : \mathcal{V} \to [0, \infty]$ is defined as

$$\rho_R\left(\tilde{\xi}\right) = \inf\left\{\alpha > 0 \,\middle|\, C_\alpha\left(\tilde{\xi}\right) \leq 0\right\},$$

where $\inf \emptyset = \infty$ and $C_\alpha\left(\tilde{\xi}\right)$, $\alpha > 0$, is the certainty equivalent of the $\tilde{\xi}$ under exponential disutility given by

$$C_\alpha\left(\tilde{\xi}\right) = \alpha \ln \mathbb{E}_{\mathbb{P}}\left(\exp\left(\frac{\tilde{\xi}}{\alpha}\right)\right).$$

In particular, Jaillet et al. [24] propose solving an optimization problem that minimizes the sum of riskiness indexes for all nodes with finite deadlines as follows:

$$\begin{aligned} \min \; & \sum_{i \in \overline{\mathcal{N}}} \rho_R(\xi_i(\boldsymbol{s}, \tilde{\boldsymbol{z}})) \\ \text{s.t. } & \boldsymbol{c}'\boldsymbol{x} \leq B, \\ & (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S}. \end{aligned} \tag{5}$$

Note that the certainty equivalent of the random delay is nonincreasing in the risk tolerant parameter α. Hence, we may interpret the riskiness index as the lowest risk tolerant parameter such that the corresponding certainty equivalent of the random delay under exponential disutility remains nonpositive. With regards to mitigating delay risks, Jaillet et al. [24] motivate the riskiness index by highlighting its salient properties as follows. For all $\tilde{\xi}, \tilde{\xi}_1, \tilde{\xi}_2 \in \mathcal{V}$:

i) **Satisficing**: $\rho_R\left(\tilde{\xi}\right) = 0$ if and only if $\mathbb{P}\left(\tilde{\xi} \leq 0\right) = 1$;
ii) **Infeasibility**: If $\mathbb{E}_{\mathbb{P}}\left(\tilde{\xi}\right) > 0$, then $\rho_R\left(\tilde{\xi}\right) = \infty$;
iii) **Convexity**: For any $\lambda \in [0, 1]$, $\rho_R\left(\lambda\tilde{\xi}_1 + (1-\lambda)\tilde{\xi}_2\right) \leq \lambda\rho_R\left(\tilde{\xi}_1\right) + (1-\lambda)\rho_R\left(\tilde{\xi}_2\right)$;
iv) **Delay bounds:**

$$\mathbb{P}\left(\tilde{\xi} > \rho_R\left(\tilde{\xi}\right)\theta\right) \leq \exp\left(-\theta\right), \quad \forall\, \theta > 0.$$

As a decision criterion to be minimized, the satsificing property implies arrivals that meet deadlines almost surely are most preferred. Moreover, solutions that result in arrivals with positive expected delays are considered infeasible. The convexity property, apart from being synonymous with risk pooling, also leads to more tractable formulations of Problem (5) over Problem (4). The last property ensures that the probability of a delay diminishes exponentially as the magnitude of the delay increases in multiples of the riskiness index.

To obtain tractable results, the riskiness index requires the random delay $\tilde{\xi}$ to be affinely dependent on independently distributed random variables. Indeed, based on Theorem 2, for the case of TSPD where $\underline{\mathcal{N}} = \emptyset$, the delay at node l, $l \in \overline{\mathcal{N}}$ becomes

$$\xi_l(\boldsymbol{s}, \tilde{\boldsymbol{z}}) = t_l(\boldsymbol{s}, \tilde{\boldsymbol{z}}) - \overline{\tau}_l = \tilde{\boldsymbol{z}}'\boldsymbol{s}^l - \overline{\tau}_l.$$

Hence, if $\tilde{z}$ comprises independently distributed random variables, we can evaluate the certainty equivalent without resorting to numerical integration as follows [24].

$$C_\alpha\left(\xi_l(\boldsymbol{s},\tilde{z})\right) = \sum_{a\in\mathcal{A}} C_\alpha(s_a^l \tilde{z}_a) - \overline{\tau}_l.$$

In particular, if $\tilde{z}$ is normally distributed with mean, $\boldsymbol{\mu}$ and covariance, $\boldsymbol{\Sigma}$, then $\xi_l(\boldsymbol{s},\tilde{z})$ would also be normally distributed and

$$C_\alpha\left(\xi_l(\boldsymbol{s},\tilde{z})\right) = \boldsymbol{\mu}'\boldsymbol{s}^l + \frac{\boldsymbol{s}^{l'}\boldsymbol{\Sigma}\boldsymbol{s}^l}{2\alpha} - \overline{\tau}_l,$$

and Problem (5) would be simplified to

$$\begin{array}{lll} \min & \displaystyle\sum_{l\in\overline{\mathcal{N}}} \alpha_l & \\ \text{s.t.} & \boldsymbol{s}^{l'}\boldsymbol{\Sigma}\boldsymbol{s}^l \le 2\alpha_l(\overline{\tau}_l - \boldsymbol{\mu}'\boldsymbol{s}^l), & \forall l \in \overline{\mathcal{N}}, \\ & \boldsymbol{c}'\boldsymbol{x} \le B & \\ & (\boldsymbol{x},\boldsymbol{s}) \in \mathcal{S} & \\ & \alpha_l \ge 0, & \forall l \in \overline{\mathcal{N}}. \end{array} \tag{6}$$

However, for the case of TSPTW, the delay function is not an affine, but rather a piecewise convex function of $\tilde{z}$, which prohibits us from obtaining an explicit and tractable formulation as in the case of TSPD. Although we could evaluate the riskiness index via sampling average approximation, the resultant model would involve optimization over exponential functions, which are nonlinear and incompatible with discrete optimization framework desired for solving the TSPTW. Hence, this motivates us to explore a similar but less computationally demanding decision criterion which, in our opinion, is more computationally scalable and better suited for solving TSPTW under uncertainty.

We propose the essential riskiness index, which has similar properties as the riskiness index of Aumann and Serrano [5].

Definition 2 Given a random delay denoted by the random variable $\tilde{\xi} \in \mathcal{V}$ with probability distribution $\mathbb{P}$, we define the essential riskiness index $\rho_E\left(\tilde{\xi}\right) : \mathcal{V} \to [0,\infty]$ as follows:

$$\rho_E\left(\tilde{\xi}\right) = \min\left\{\alpha \ge 0 \;\middle|\; \mathbb{E}_{\mathbb{P}}\left(\max\left\{\tilde{\xi}, -\alpha\right\}\right) \le 0\right\}.$$

We interpret the essential riskiness index as the lowest risk tolerant parameter such that the corresponding certainty equivalent of the random delay under a ramp disutility remains nonpositive. The essential riskiness index has similar salient properties as follows:

Proposition 3 *For all $\tilde{\xi}, \tilde{\xi}_1, \tilde{\xi}_2 \in \mathcal{V}$:*

i) ***Satisficing***: $\rho_E\left(\tilde{\xi}\right) = 0$ *if and only if* $\mathbb{P}\left(\tilde{\xi} \leq 0\right) = 1$;

ii) ***Infeasibility***: *If* $\mathbb{E}_{\mathbb{P}}\left(\tilde{\xi}\right) > 0$, *then* $\rho_E\left(\tilde{\xi}\right) = \infty$;

iii) ***Convexity***: *For any* $\lambda \in [0, 1]$, $\rho_E\left(\lambda\tilde{\xi}_1 + (1-\lambda)\tilde{\xi}_2\right) \leq \lambda\rho_E\left(\tilde{\xi}_1\right) + (1-\lambda)\rho_E\left(\tilde{\xi}_2\right)$;

iv) ***Delay bounds***:

$$\mathbb{P}\left(\tilde{\xi} > \rho_E\left(\tilde{\xi}\right)\theta\right) \leq \frac{1}{1+\theta}, \quad \forall\, \theta > 0.$$

Proof See Appendix A.2. □

We use the prefix "essential" in the sense that the ramp function is the simplest form of disutility that we could use to obtain the salient properties. Recall that the classical on-time arrival probability criterion captures only the probability of tardiness and completely ignores the magnitude of delays. In contrast, as expounded in the property of delay bounds, the essential riskiness index ensures that the probability of a delay diminishes reciprocally as the magnitude of the delay increases in multiples of the essential riskiness index. Hence, the essential riskiness index accounts for both the probability of tardiness and its magnitude. Although the delay bounds are not as sharp as the Aumann and Serrano [5] riskiness index, as we will demonstrate, the key advantage of having a ramp over exponential disutility is that it provides more tractable formulation for addressing the TSPTW under uncertainty. We will also show in the next section that when applying the essential riskiness index on the TSPD, i.e., $\underline{\mathcal{N}} = \emptyset$, there is an important situation in which the solutions of minimizing the riskiness index coincide with those that minimize the essential riskiness index under distributional ambiguity.

To obtain tractable formulations in routing optimization problems, the use of Aumann and Serrano [5] riskiness index in the decision criterion would require such formulations to have random delays being affinely dependent on a set of independently distributed random variables. As in the case of the TSPTW model of Jaillet et al. [24], it poses a serious modeling issue, which requires the mean arrival times to fall within the "soft" time windows as follows:

Example 1 Consider the network in Fig. 1, which comprises 3 nodes $\{1, 2, 3\}$ and 2 arcs $\{(1, 2), (2, 3)\}$, with travel times such that $\mathbb{P}(\tilde{z}_{12} = 1) = 1$ and $\mathbb{P}(\tilde{z}_{23} = 1) = 1$, respectively. The customer prescribes a time window $[\underline{\tau}_2, \overline{\tau}_2]$ to be $[2, 3]$. There is only

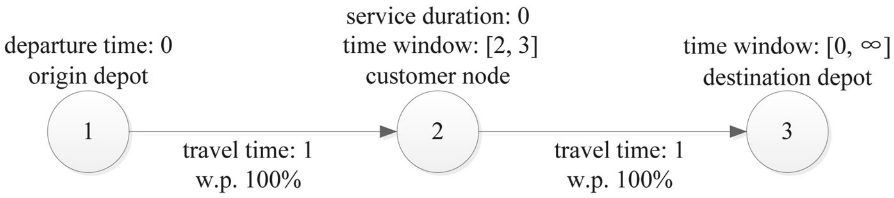

Fig. 1 A network example

one feasible route: $\{(1,2),(2,3)\}$. In the hard time window case, the vehicle departs from the origin depot at time 0, arrives at the customer node at time 1, waits for a duration of 1, serves the customer node and departs at time 2, and finally returns to the destination depot at time 3. The delay function at the customer node, $\max\{\tilde{z}_{12}, \underline{\tau}_2\} - \overline{\tau}_2$, is convex piecewise affine in $\tilde{z}$. The corresponding essential riskiness index is $\min\left\{\alpha \geq 0 \;\middle|\; \mathbb{E}_{\mathbb{P}}\left(\max\left\{\max\{\tilde{z}_{12}, \underline{\tau}_2\} - \overline{\tau}_2, -\alpha\right\}\right) \leq 0\right\} = 0$, implying that the route is feasible, even fully satisficing. However, if we use the approach proposed by Jaillet et al. [24] for soft time windows, the riskiness index would be $\inf\{\alpha > 0 \mid C_\alpha(\tilde{z}_{12}) - \overline{\tau}_2 \leq 0, C_\alpha(-\tilde{z}_{12}) + \underline{\tau}_2 \leq 0\} = \infty$, implying that the route is infeasible.

Moreover, Adulyasak and Jaillet [1] have demonstrated via numerical experiments that applying the Aumann and Serrano [5] riskiness index for TSPTW would result in poor performance in mitigating the time window violation.

As a decision criterion for target oriented stochastic or distributional robust optimization problems, the use of the essential riskiness index has greater computational advantage over Aumann and Serrano [5] riskiness index. For instance, under the essential riskiness index, we can formulate a two-stage stochastic optimization problem as a large scale linear optimization problem via sample average approximation (SAA). In contrast, under the Aumann and Serrano [5] riskiness index, it would become a large scale convex optimization problems involving exponential functions, an optimization format that is not as scalable and efficiently solvable as linear optimization problems. Moreover, due to the convex piecewise nature of the essential riskiness index, it is compatible with the format of adaptive distributionally robust optimization models for which approximation techniques such as linear decision rules [10] and Fourier–Motzkin's elimination are available to solve the problem [45]. However, these approximation techniques would not be possible if the decision criterion is based on the Aumann and Serrano [5] riskiness index. Other applications of the essential riskiness index beyond VRPTW include, among others, inventory management [37,44] and medical appointment scheduling [22,34]. The only computational advantage of the Aumann and Serrano [5] riskiness index over the essential riskiness index occurs when the underlying risk is affinely dependent on a set of independently distributed random variables, which has limited scope for application.

4 Routing optimization over essential riskiness index

Similar to Jaillet et al. [24], we propose solving an optimization problem that minimizes the sum of essential riskiness indexes as follows:

$$\begin{aligned} \min \ & \sum_{i \in \overline{\mathcal{N}}} \rho_E(\xi_i(\boldsymbol{s}, \tilde{\boldsymbol{z}})) \\ \text{s.t. } & \boldsymbol{c}'\boldsymbol{x} \leq B, \\ & (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S} \end{aligned} \tag{7}$$

or equivalently, based on Theorem 2, we have

$$
\begin{aligned}
\min\ & \sum_{l\in\overline{\mathcal{N}}} \alpha_l \\
\text{s.t. } & \mathbb{E}_{\mathbb{P}}\left(\max\left\{\max_{k\in\underline{\mathcal{N}}\cup\{1\}}\left\{\sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k + \tilde{z}'(\boldsymbol{s}^l-\boldsymbol{s}^k)\right\} - \overline{\tau}_l, -\alpha_l\right\}\right) \le 0,\ \forall l\in\overline{\mathcal{N}}, \\
& \boldsymbol{c}'\boldsymbol{x} \le B, \\
& (\boldsymbol{x},\boldsymbol{s}) \in \mathcal{S}, \\
& \alpha_l \ge 0, \qquad \forall l \in \overline{\mathcal{N}}.
\end{aligned}
\tag{8}
$$

4.1 Sample average approximation

For a given distribution, $\mathbb{P}$, evaluating the expectation requires high dimensional integration, which is generally a computationally expensive procedure. Nevertheless, using sample average approximation method, we can reformulate Problem (8) as follows. Let Ω denote the set of sample indexes. In the ω-th sample, $\omega \in \Omega$, the realization of $\tilde{z}$ is denoted by $\boldsymbol{z}(\omega)$. We introduce auxiliary variable $y_l^\omega = \max\left\{\max_{k\in\underline{\mathcal{N}}\cup\{1\}}\left\{\sum_{a\in\delta^-(k)} s_a^l\underline{\tau}_k + \boldsymbol{z}(\omega)'(\boldsymbol{s}^l-\boldsymbol{s}^k)\right\} - \overline{\tau}_l, -\alpha_l\right\}$ for each $l \in \overline{\mathcal{N}}$ in order to linearize the reformulation. Problem (8) can then be approximated by the following mixed integer optimization problem:

$$
\begin{aligned}
\min\ & \sum_{l\in\overline{\mathcal{N}}} \alpha_l \\
\text{s.t. } & \sum_{\omega\in\Omega} y_l^\omega \le 0, && \forall l\in\overline{\mathcal{N}}, \\
& y_l^\omega \ge \sum_{a\in\delta^-(k)} s_a^l\underline{\tau}_k + \boldsymbol{z}(\omega)'(\boldsymbol{s}^l-\boldsymbol{s}^k) - \overline{\tau}_l, && \forall l\in\overline{\mathcal{N}}, k\in\underline{\mathcal{N}}\cup\{1\}, \omega\in\Omega, \\
& y_l^\omega \ge -\alpha_l, && \forall l\in\overline{\mathcal{N}}, \omega\in\Omega, \\
& \boldsymbol{c}'\boldsymbol{x} \le B, \\
& (\boldsymbol{x},\boldsymbol{s}) \in \mathcal{S}, \\
& \alpha_l \ge 0, && \forall l\in\overline{\mathcal{N}}.
\end{aligned}
\tag{9}
$$

Here, each sample $\omega \in \Omega$ of travel times occurs with an equal probability. If, in general, it occurs with a probability $p^\omega \in [0, 1]$ subject to $\sum_{\omega\in\Omega} p^\omega = 1$, we simply revise the first constraint as $\sum_{\omega\in\Omega} p^\omega y_l^\omega \le 0, \forall l \in \overline{\mathcal{N}}$.

Note that to solve Problem (4) via sample average approximation, we would require to introduce as many new binary decision variables as the number of samples (see, for instance, [1]). In contrast, the decision variables y_l^ω introduced in Problem (9) are all continuous, which are generally easier to optimize compared to discrete ones. Nevertheless, while we may formulate Problem (9) and solve directly using state-of-the-art commercial solvers such as CPLEX and Gurobi, we will show in Sect. 5 that for a large sample size, it would be more computationally efficient to solve the problem via Benders decomposition.

4.2 A distributionally robust model

We can extend and define the essential riskiness index to encompasses distributional ambiguity as follows

$$\rho_E\left(\tilde{\xi}\right) = \min\left\{\alpha \geq 0 \;\middle|\; \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}\left(\max\left\{\tilde{\xi}, -\alpha\right\}\right) \leq 0\right\},$$

where $\mathbb{F}$ is an *ambiguity set* of probability distributions. Note that this indeed generalizes the previous definition because when the distribution $\mathbb{P}$ is exactly known, the ambiguity set is a singleton, i.e. $\mathbb{F} = \{\mathbb{P}\}$. Under this definition, when the probability distribution is not uniquely specified, the index would be evaluated on the worst case distribution, which reflects the attitude of ambiguity aversion. Correspondingly, we formulate the distributionally robust TSPTW as

$$\begin{aligned}
\min \; & \sum_{l\in\overline{\mathcal{N}}} \alpha_l \\
\text{s.t. } & \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}\left(\max\left\{\max_{k\in\underline{\mathcal{N}}\cup\{1\}}\left\{\sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k + \tilde{\boldsymbol{z}}'(\boldsymbol{s}^l - \boldsymbol{s}^k)\right\} - \overline{\tau}_l, -\alpha_l\right\}\right) \leq 0, \; \forall l \in \overline{\mathcal{N}}, \\
& \boldsymbol{c}'\boldsymbol{x} \leq B, \\
& (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S}, \\
& \alpha_l \geq 0, \qquad \forall l \in \overline{\mathcal{N}}.
\end{aligned} \tag{10}$$

To derive an explicit formulation, we consider the *cross moment ambiguity set*, which is characterized by the mean values and covariance of $\tilde{\boldsymbol{z}}$ as follows

$$\mathbb{F} = \left\{\mathbb{P} \in \mathcal{P} \;\middle|\; \begin{array}{l} \mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left((\tilde{\boldsymbol{z}} - \boldsymbol{\mu})(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})'\right) = \boldsymbol{\Sigma} \end{array}\right\},$$

where $\boldsymbol{\mu} > \boldsymbol{0}$ and $\boldsymbol{\Sigma}$ is a positive definite matrix and $\mathcal{P}$ is the set of all probability distributions on $\mathbb{R}^{|\mathcal{A}|}$. We refer interested readers to Wiesemann et al. [43] for more general forms of ambiguity sets that would also lead to tractable formulations.

Theorem 1 *Problem* (10) *under the cross moments ambiguity set is equivalent to the following optimization problem.*

$$
\begin{aligned}
\min\ & \sum_{l\in\overline{\mathcal{N}}} \alpha_l \\
\text{s.t. } & v_{l0} + \operatorname{tr}(\boldsymbol{\Sigma V}_l) \le \alpha_l, && \forall l \in \overline{\mathcal{N}}, \\
& \begin{bmatrix} v_{l0} & \frac{\boldsymbol{v}_l'}{2} \\ \frac{\boldsymbol{v}_l}{2} & \boldsymbol{V}_l \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1}, && \forall l \in \overline{\mathcal{N}}, \\
& \begin{bmatrix} v_{l0} - \sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k - \boldsymbol{\mu}'(\boldsymbol{s}^l - \boldsymbol{s}^k) + \overline{\tau}_l - \alpha_l & \frac{(\boldsymbol{v}_l - \boldsymbol{s}^l + \boldsymbol{s}^k)'}{2} \\ \frac{\boldsymbol{v}_l - \boldsymbol{s}^l + \boldsymbol{s}^k}{2} & \boldsymbol{V}_l \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1}, && \forall l \in \overline{\mathcal{N}}, \\
& && k \in \underline{\mathcal{N}} \cup \{1\} \\
& \boldsymbol{c}'\boldsymbol{x} \le B, \\
& (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S}, \\
& \alpha_l \ge 0, && \forall l \in \overline{\mathcal{N}}.
\end{aligned}
\tag{11}
$$

where $\operatorname{tr}(\boldsymbol{A})$ *denotes the trace of matrix* $\boldsymbol{A}$ *and* $\mathbb{S}_+^{|\mathcal{A}|+1}$ *denotes the set of symmetric positive semidefinite matrices in* $\mathbb{R}^{(|\mathcal{A}|+1)\times(|\mathcal{A}|+1)}$.

Proof See Appendix A.3. □

For the case of TSPD, which is a special case of Problem (10) with $\underline{\mathcal{N}} = \emptyset$, Problem (11) becomes

$$
\begin{aligned}
\min\ & \sum_{l\in\overline{\mathcal{N}}} \alpha_l \\
\text{s.t. } & v_{l0} + \operatorname{tr}(\boldsymbol{\Sigma V}_l) \le \alpha_l, && \forall l \in \overline{\mathcal{N}}, \\
& \begin{bmatrix} v_{l0} & \frac{\boldsymbol{v}_l'}{2} \\ \frac{\boldsymbol{v}_l}{2} & \boldsymbol{V}_l \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1}, && \forall l \in \overline{\mathcal{N}}, \\
& \begin{bmatrix} v_{l0} - \boldsymbol{\mu}'\boldsymbol{s}^l + \overline{\tau}_l - \alpha_l & \frac{(\boldsymbol{v}_l - \boldsymbol{s}^l)'}{2} \\ \frac{\boldsymbol{v}_l - \boldsymbol{s}^l}{2} & \boldsymbol{V}_l \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1}, && \forall l \in \overline{\mathcal{N}}, \\
& \boldsymbol{c}'\boldsymbol{x} \le B, \\
& (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S}, \\
& \alpha_l \ge 0, && \forall l \in \overline{\mathcal{N}}.
\end{aligned}
\tag{12}
$$

However, we can further eliminate the the positive semidefinite constraints and replace them with second-order conic constraints as follows:

Theorem 2 *When* $\underline{\mathcal{N}} = \emptyset$, *Problem* (10) *under the cross moments ambiguity set is equivalent to the following optimization problem.*

$$
\begin{aligned}
\min \; & \sum_{l \in \overline{\mathcal{N}}} \alpha_l \\
\text{s.t. } & \boldsymbol{s}^{l\prime} \boldsymbol{\Sigma} \boldsymbol{s}^l \leq 4\alpha_l (\overline{\tau}_l - \boldsymbol{\mu}' \boldsymbol{s}^l), \; \forall l \in \overline{\mathcal{N}}, \\
& \boldsymbol{c}' \boldsymbol{x} \leq B, \\
& (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S}, \\
& \alpha_l \geq 0, \qquad \forall l \in \overline{\mathcal{N}}.
\end{aligned} \tag{13}
$$

Proof We can derive the result using the projection theorem of Popescu [33] and the worst-case expectation result for a newsvendor problem of Scarf et al. [36]. In Appendix A.4, we present a different proof that demonstrates directly the equivalence of Problems (12) and (13). □

Incidentally, it is interesting to note that despite the difference in objective values, the optimal routes obtained from solving Problems (13) and (6) are the same. We also note that Problems (11) and (13) have both discrete and nonlinear conic constraints. Although such a format is generally not supported by discrete optimization software packages, by exploiting conic duality, we can still adopt Benders decomposition techniques to solve these problems.

5 Benders decomposition

We develop a Benders decomposition approach for solving the routing optimization problem (7), which in the most general form, can be expressed as follows:

$$
\begin{aligned}
\min \; & \sum_{l \in \overline{\mathcal{N}}} \eta_l \\
\text{s.t. } & \eta_l \geq F_l(\boldsymbol{s}), \forall l \in \overline{\mathcal{N}}, \\
& \boldsymbol{c}' \boldsymbol{x} \leq B, \\
& (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S},
\end{aligned}
$$

where the function $F_l(\boldsymbol{s}) : \mathbb{R}^{|\mathcal{A}| \times |\mathcal{N}|} \mapsto [0, \infty]$ corresponds to the essential riskiness index at the node l, $l \in \overline{\mathcal{N}}$ given by

$$
\begin{aligned}
F_l(\boldsymbol{s}) = \min \; & \alpha_l \\
\text{s.t. } & \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(\max \left\{ \max_{k \in \underline{\mathcal{N}} \cup \{1\}} \left\{ \sum_{a \in \delta^-(k)} s_a^l \underline{\tau}_k + \tilde{\boldsymbol{z}}'(\boldsymbol{s}^l - \boldsymbol{s}^k) \right\} - \overline{\tau}_l, -\alpha_l \right\} \right) \leq 0, \\
& \alpha_l \geq 0.
\end{aligned} \tag{14}
$$

The following proposition shows the convexity of $F_l(\boldsymbol{s})$, which is crucial in developing Benders decomposition approach.

Proposition 4 *Function $F_l(\boldsymbol{s})$ is convex in $\boldsymbol{s}$.*

Proof See Appendix A.5. □

We further assume that Problem (14) has a tractable dual formulation and that the conditions of strong duality holds so that

$$F_l(\boldsymbol{s}) = \max_{(\zeta_{l0},\zeta_l)\in\mathcal{Z}_l} \zeta_{l0} + \boldsymbol{\zeta}_l'\boldsymbol{s}, \tag{15}$$

where $\zeta_{l0} \in \mathbb{R}$, $\boldsymbol{\zeta}_l = (\boldsymbol{\zeta}_l^j)_{j\in\mathcal{N}}$, $\boldsymbol{\zeta}_l^j \in \mathbb{R}_+^{|\mathcal{A}|}$ and $\mathcal{Z}_l \subseteq \mathbb{R} \times \mathbb{R}^{|\mathcal{A}|\times|\mathcal{N}|}$ is a non-empty convex set. Note that when Problem (15) is unbounded, it corresponds to the case when the essential riskiness index evaluated at node l is infinite and implies that the route is infeasible. Hence, we define the set $\mathcal{R}_l$ as the recession cone of $\mathcal{Z}_l$, i.e.,

$$\mathcal{R}_l = \left\{(\overline{\zeta}_{l0}, \overline{\boldsymbol{\zeta}}_l) \in \mathbb{R} \times \mathbb{R}^{|\mathcal{A}|\times|\mathcal{N}|} \mid (\zeta_{l0}, \boldsymbol{\zeta}_l) + \lambda\left(\overline{\zeta}_{l0}, \overline{\boldsymbol{\zeta}}_l\right) \in \mathcal{Z}_l, \forall(\zeta_{l0}, \boldsymbol{\zeta}_l) \in \mathcal{Z}_l, \lambda \geq 0\right\},$$

so that $F_l(\boldsymbol{s})$ is finite if and only if

$$\overline{\zeta}_{l0} + \overline{\boldsymbol{\zeta}}_l'\boldsymbol{s} \leq 0, \quad \forall(\overline{\zeta}_{l0}, \overline{\boldsymbol{\zeta}}_l) \in \mathcal{R}_l.$$

We assume that there exists an efficient algorithm to obtain the optimal solution of Problem (15) if it is finite. Otherwise, we can also efficiently determine a recession direction $(\overline{\zeta}_{l0}, \overline{\boldsymbol{\zeta}}_l) \in \mathcal{R}_l$ such that $\overline{\zeta}_{l0} + \overline{\boldsymbol{\zeta}}_l'\boldsymbol{s} > 0$.

To solve Problem (7) using the Benders decomposition approach, we define the restricted master problem at the t-th iteration as

$$\begin{aligned}
\min \ & \sum_{l\in\overline{\mathcal{N}}} \eta_l \\
\text{s.t. } & \zeta_{l0} + \boldsymbol{\zeta}_l'\boldsymbol{s} \leq \eta_l, \ \forall(\zeta_{l0}, \boldsymbol{\zeta}_l) \in \mathcal{Z}_l^t, \forall l \in \overline{\mathcal{N}}, \\
& \overline{\zeta}_{l0} + \overline{\boldsymbol{\zeta}}_l'\boldsymbol{s} \leq 0, \ \forall(\overline{\zeta}_{l0}, \overline{\boldsymbol{\zeta}}_l) \in \mathcal{R}_l^t, \forall l \in \overline{\mathcal{N}}, \\
& \boldsymbol{c}'\boldsymbol{x} \leq B, \\
& (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S},
\end{aligned} \tag{16}$$

where $\mathcal{Z}_l^t$ and $\mathcal{R}_l^t$ are finite subsets of $\mathcal{Z}_l$ and $\mathcal{R}_l$ respectively, for all $l \in \overline{\mathcal{N}}$.

Algorithm 1 (Benders decomposition)
Initialization: *Set* $t := 1$, $\mathcal{Z}_l^1 := \emptyset$ *and* $\mathcal{R}_l^1 := \emptyset$, $\forall l \in \overline{\mathcal{N}}$.

1. *Solve Problem* (16) *and let* $(\boldsymbol{x}^*, \boldsymbol{s}^*, \boldsymbol{\eta}^*)$ *be an optimal solution.*
2. *For all* $l \in \overline{\mathcal{N}}$, *solve Problem* (15). *If* $F_l(\boldsymbol{s}^*)$ *is finite, let the optimal solution be* $(\zeta_{l0}, \boldsymbol{\zeta}_l) \in \mathcal{Z}_l$. *Otherwise, let* $(\overline{\zeta}_{l0}, \overline{\boldsymbol{\zeta}}_l) \in \mathcal{R}_l$ *be a recession direction such that* $\overline{\zeta}_{l0} + \overline{\boldsymbol{\zeta}}_l{}'\boldsymbol{s}^* > 0$.
3. *If* $F_l(\boldsymbol{s}^*) = \eta_l^*$ *for all* $l \in \overline{\mathcal{N}}$ *then terminate algorithm and return optimal route* $\boldsymbol{x}^*$.
4. *Set*

$$\mathcal{R}_l^{t+1} := \mathcal{R}_l^t \cup \{(\overline{\zeta}_{l0}, \overline{\boldsymbol{\zeta}}_l)\} \qquad \forall l \in \overline{\mathcal{N}} : F_l(\boldsymbol{s}^*) = \infty$$

and

$$\mathcal{Z}_l^{t+1} := \mathcal{Z}_l^t \cup \{(\zeta_{l0}, \boldsymbol{\zeta}_l)\} \qquad \forall l \in \overline{\mathcal{N}} : F_l(\boldsymbol{s}^*) \in (\eta_l^*, \infty).$$

5. *Set* $t := t + 1$. *Go to Step 1.*

Since the domain of variables $(\boldsymbol{x}, \boldsymbol{s})$ is finite, the algorithm must terminate because only finitely many subproblems can be defined.

Algorithm 1 presents a cutting-plane implementation, in which Benders dual subproblems (15) would be solved only after the restricted master problem (16) is solved to optimality. However, it is well-known that Benders decomposition can benefit in computational speed from using branch-and-cut implementations [41], in which Benders dual subproblems (15) are solved at the restricted master problem (16)'s integer nodes and possibly at the fractional nodes of low depth, but not just at its optimal node. The branch-and-cut variant of Algorithm 1 can be easily implemented in modern integer programming solvers. In our computational study in Sect. 7, we will use the IBM CPLEX general purpose integer programming solver to solve the restricted master problem in a branch-and-cut fashion, where the Benders cuts are added using the function `ILOLAZYCONSTRAINTCALLBACK`.

We next elaborate on how we can form and solve the dual problem (15) for several concrete cases of the primal problem (14). For sample average approximation, we write Problem (14) as

$$\begin{aligned} F_l(\boldsymbol{s}) = \min\ & \alpha_l \\ \text{s.t.}\ & \sum_{\omega \in \Omega} y^\omega \le 0, \\ & y^\omega \ge \xi_{lk}^\omega(\boldsymbol{s}),\ \forall k \in \underline{\mathcal{N}} \cup \{1\}, \omega \in \Omega, \\ & y^\omega \ge -\alpha_l, \quad \forall \omega \in \Omega, \\ & \alpha_l \ge 0, \end{aligned} \tag{17}$$

where $\xi_{lk}^\omega(\boldsymbol{s}) = \sum_{a \in \delta^-(k)} s_a^l \underline{\tau}_k + \boldsymbol{z}(\omega)'(\boldsymbol{s}^l - \boldsymbol{s}^k) - \overline{\tau}_l$.

Theorem 3 *Let* $\xi_l^\omega(\boldsymbol{s}) = \max_{k \in \underline{\mathcal{N}} \cup \{1\}} \xi_{lk}^\omega(\boldsymbol{s})$ *and let the index function* $\nu : \Omega \mapsto \Omega$ *be a permutation of* Ω *such that*

$$\xi_l^{\nu(1)}(\boldsymbol{s}) \ge \xi_l^{\nu(2)}(\boldsymbol{s}) \ge \cdots \ge \xi_l^{\nu(|\Omega|)}(\boldsymbol{s}).$$

If $\sum_{\omega \in \Omega} \xi_l^\omega(\boldsymbol{s}) > 0$*, then* $F_l(\boldsymbol{s}) = \infty$*. Otherwise,*

$$F_l(\boldsymbol{s}) = \max \left\{ \max_{i \in \{1,2,\ldots,|\Omega|-1\}} \left\{ \sum_{\omega=1}^{i} \frac{\xi_l^{\nu(\omega)}(\boldsymbol{s})}{|\Omega| - i} \right\}, 0 \right\}. \tag{18}$$

Proof See Appendix A.6. □

Theorem 3 indicates that we can solve the Benders subproblems via a sorting algorithm, which is a strongly polynomial time algorithm that scales well computationally with the number of samples, i.e., $O(|\Omega| \log(|\Omega|))$. When implementing the Benders decomposition for a given solution of the restricted master problem, $\boldsymbol{s}^*$, we determine for each $l \in \overline{\mathcal{N}}$, whether $\sum_{\omega\in\Omega} \xi_l^\omega(\boldsymbol{s}^*) > 0$. If so, we extract the index function, $\kappa^*(\omega) \in \arg\max_{k\in\underline{\mathcal{N}}\cup\{1\}} \xi_{lk}^\omega(\boldsymbol{s}^*)$, determine the affine relation such that $\sum_{\omega\in\Omega} \xi_{l\kappa^*(\omega)}^\omega(\boldsymbol{s}^*) = \overline{\zeta}_{l0} + \overline{\boldsymbol{\zeta}}_l' \boldsymbol{s}^*$, and add $(\overline{\zeta}_{l0}, \overline{\boldsymbol{\zeta}}_l)$ to the set $\mathcal{R}_l^t$. Specifically, $\overline{\zeta}_{l0} = -|\Omega|\overline{\tau}_l$ and, for $a \in \mathcal{A}$ and $j \in \mathcal{N}$, $\overline{\zeta}_{la}^j = \sum_{\omega\in\Omega} \zeta_{la\omega}^j$, where

$$\zeta_{la\omega}^j = \begin{cases} z_a(\omega), & \text{if } j = l \neq \kappa^*(\omega), a \in \mathcal{A}\backslash\delta^-(\kappa^*(\omega)), \\ z_a(\omega) + \underline{\tau}_{\kappa^*(\omega)}, & \text{if } j = l \neq \kappa^*(\omega), a \in \delta^-(\kappa^*(\omega)), \\ -z_a(\omega), & \text{if } j = \kappa^*(\omega) \neq l, a \in \mathcal{A}, \\ \underline{\tau}_{\kappa^*(\omega)}, & \text{if } j = l = \kappa^*(\omega), a \in \delta^-(\kappa^*(\omega)), \\ 0, & \text{otherwise.} \end{cases}$$

If $\xi_l^\omega(\boldsymbol{s}^*) \leq 0$ for all $\omega \in \Omega$, we obtain from (18) that $F_l(\boldsymbol{s}^*) = 0$. Otherwise, we have

$$F_l(\boldsymbol{s}^*) = \sum_{\omega=1}^{i^*} \frac{\xi_{l\kappa^*(\nu(\omega))}^{\nu(\omega)}(\boldsymbol{s}^*)}{|\Omega| - i^*}$$

for some $i^* \in \{1, \ldots, |\Omega| - 1\}$. Similarly, we can extract the affine relation such that $F_l(\boldsymbol{s}^*) = \zeta_{l0} + \boldsymbol{\zeta}_l' \boldsymbol{s}^*$ and introduce $(\zeta_{l0}, \boldsymbol{\zeta}_l)$ to the set $\mathcal{Z}_l^t$, where

$$\zeta_{l0} = -\frac{i^*}{|\Omega| - i^*}\overline{\tau}_l,$$

and, for $a \in \mathcal{A}$ and $j \in \mathcal{N}$,

$$\zeta_{la}^j = \sum_{\nu(\omega)=1}^{i^*} \frac{\zeta_{la\nu(\omega)}^j}{|\Omega| - i^*}.$$

For the distributionally robust model with the cross moments ambiguity set, we express Problem (14) as the following semidefinite optimization problem.

$$\begin{aligned} F_l(\boldsymbol{s}) = \min\ & \alpha_l \\ \text{s.t. } & v_{l0} + \operatorname{tr}(\boldsymbol{\Sigma} \boldsymbol{V}_l) \leq \alpha_l, \\ & \begin{bmatrix} v_{l0} & \dfrac{\boldsymbol{v}_l'}{2} \\ \dfrac{\boldsymbol{v}_l}{2} & \boldsymbol{V}_l \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1}, \\ & \begin{bmatrix} v_{l0} - \sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k - \boldsymbol{\mu}'(\boldsymbol{s}^l - \boldsymbol{s}^k) + \overline{\tau}_l - \alpha_l & \dfrac{(\boldsymbol{v}_l - \boldsymbol{s}^l + \boldsymbol{s}^k)'}{2} \\ \dfrac{\boldsymbol{v}_l - \boldsymbol{s}^l + \boldsymbol{s}^k}{2} & \boldsymbol{V}_l \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1}, \\ & \qquad k \in \underline{\mathcal{N}} \cup \{1\}, \\ & \alpha_l \geq 0. \end{aligned} \tag{19}$$

Theorem 4 *The dual problem of Problem (19) is given by*

$$
\begin{aligned}
F_l(\boldsymbol{s}) = \max \ & \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} \left(\left(\sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k + \boldsymbol{\mu}'(\boldsymbol{s}^l - \boldsymbol{s}^k) - \overline{\tau}_l \right) r_{k0} + \left(\boldsymbol{s}^l - \boldsymbol{s}^k\right)' \boldsymbol{r}_k \right) \\
\text{s.t. } & \beta - \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} r_{k0} \le 1, \\
& -\beta + \gamma_0 + \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} r_{k0} = 0, \\
& \boldsymbol{\gamma} + \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} \boldsymbol{r}_k = \boldsymbol{0}, \\
& -\boldsymbol{\Sigma}\beta + \boldsymbol{\Gamma} + \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} \boldsymbol{R}_k = \boldsymbol{0}, \\
& \beta \ge 0, \\
& \begin{bmatrix} \gamma_0 & \boldsymbol{\gamma}' \\ \boldsymbol{\gamma} & \boldsymbol{\Gamma} \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1}, \\
& \begin{bmatrix} r_{k0} & \boldsymbol{r}_k' \\ \boldsymbol{r}_k & \boldsymbol{R}_k \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1}, \qquad \forall k \in \underline{\mathcal{N}} \cup \{1\}.
\end{aligned}
\tag{20}
$$

and their objectives coincide.

Proof See Appendix A.7. □

As before, Theorem 4 indicates that we can solve the Benders subproblems in polynomial time using solvers that support semidefinite optimization. From the optimal objective of Problem(20), we can extract the affine relation with respect to $\boldsymbol{s}^*$ such that $F_l(\boldsymbol{s}^*) = \zeta_{l0} + \boldsymbol{\zeta}_l'\boldsymbol{s}^*$, where $\zeta_{l0} = -\sum_{k\in\underline{\mathcal{N}}\cup\{1\}} \overline{\tau}_l r_{k0}$ and, for $a \in \mathcal{A}$ and $j \in \mathcal{N}$, $\zeta_{la}^j = \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} \zeta_{lak}^j$,

$$
\zeta_{lak}^j = \begin{cases} \mu_a r_{k0} + r_{ak}, & \text{if } j = l \ne k, a \in \mathcal{A}\backslash\delta^-(k), \\ \mu_a r_{k0} + r_{ak} + r_{k0}\underline{\tau}_k, & \text{if } j = l \ne k, a \in \delta^-(k), \\ -\mu_a r_{k0} - r_{ak}, & \text{if } j = k \ne l, a \in \mathcal{A}, \\ r_{k0}\underline{\tau}_k, & \text{if } j = l = k, a \in \delta^-(k), \\ 0, & \text{otherwise.} \end{cases}
$$

Moreover, whenever the objective is unbounded, i.e., $F_l(\boldsymbol{s}^*) = \infty$, we assume that the solver can return the recession direction, that allows us to determine the violating inequality, $\overline{\zeta}_{l0} + \overline{\boldsymbol{\zeta}}_l'\boldsymbol{s}^* > 0$, which is the case of popular semidefinite programming solvers such as MOSEK and SDPT3.

The Benders subproblem would be further simplified for the case of distributionally robust TSPD under the cross moment ambiguity set. Correspondingly, we write Problem (14) as

$$
\begin{aligned}
F_l(\boldsymbol{s}) = \min \ & \alpha_l \\
\text{s.t. } & \boldsymbol{s}^{l\prime}\boldsymbol{\Sigma}\boldsymbol{s}^l \le 4\alpha_l(\overline{\tau}_l - \boldsymbol{\mu}'\boldsymbol{s}^l) \\
& \alpha_l \ge 0.
\end{aligned}
\tag{21}
$$

Theorem 5 *The dual problem of Problem* (21) *is given by*

$$F_l(\boldsymbol{s}) = \max_{\boldsymbol{r}} \left\{ -\boldsymbol{r}'\boldsymbol{r}(\overline{\tau}_l - \boldsymbol{\mu}'\boldsymbol{s}^l) - \left(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{r}\right)'\boldsymbol{s}^l \right\}. \tag{22}$$

Proof See Appendix A.8. □

Note that when $\boldsymbol{s}^{l*} = \boldsymbol{0}$, then $F_l(\boldsymbol{s}^*) = 0$ and a cut would not be introduced in the Benders decomposition. Otherwise, observe that if $\overline{\tau}_l - \boldsymbol{\mu}'\boldsymbol{s}^{l*} > 0$, then by the first-order condition, the optimal solution is

$$\boldsymbol{r} = \frac{\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{s}^{l*}}{2\left(\boldsymbol{\mu}'\boldsymbol{s}^{l*} - \overline{\tau}_l\right)}$$

and correspondingly,

$$F_l(\boldsymbol{s}^*) = \zeta_{l0} + \sum_{j\in\mathcal{N}} \boldsymbol{\zeta}_l^{j\prime}\boldsymbol{s}^{j*}$$

where

$$\zeta_{l0} = -\frac{\boldsymbol{s}^{l*\prime}\boldsymbol{\Sigma}\boldsymbol{s}^{l*}\overline{\tau}_l}{4\left(\boldsymbol{\mu}'\boldsymbol{s}^{l*} - \overline{\tau}_l\right)^2}$$

and

$$\boldsymbol{\zeta}_l^j = \begin{cases} \dfrac{\boldsymbol{s}^{l*\prime}\boldsymbol{\Sigma}\boldsymbol{s}^{l*}}{4\left(\boldsymbol{\mu}'\boldsymbol{s}^{l*} - \overline{\tau}_l\right)^2}\boldsymbol{\mu} - \dfrac{1}{2\left(\boldsymbol{\mu}'\boldsymbol{s}^{l*} - \overline{\tau}_l\right)}\boldsymbol{\Sigma}\boldsymbol{s}^{l*}, & \text{if } j = l, \\ \boldsymbol{0}, & \text{otherwise.} \end{cases}$$

Note also that since $\boldsymbol{\Sigma}$ is positive definite, Problem (22) is unbounded when $\overline{\tau}_l - \boldsymbol{\mu}'\boldsymbol{s}^{l*} \le 0$, in which case, observe that

$$\bar{\boldsymbol{r}} = -\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{s}^{l*}$$

is a recession direction. Correspondingly, we have

$$\overline{\zeta}_{l0} = -\boldsymbol{s}^{l*\prime}\boldsymbol{\Sigma}\boldsymbol{s}^{l*}\overline{\tau}_l$$

and

$$\overline{\boldsymbol{\zeta}}_l^j = \begin{cases} \boldsymbol{s}^{l*\prime}\boldsymbol{\Sigma}\boldsymbol{s}^{l*}\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{s}^{l*}, & \text{if } j = l, \\ \boldsymbol{0}, & \text{otherwise.} \end{cases}$$

6 Extensions

Apart from the essential riskiness index decision criterion, the results developed in the paper can easily be extended to a travel cost minimization problem with constraints that safeguard the risk of late arrivals defined via the popular Conditional Value-at-Risk (CVaR) measure of [35] defined as

$$\text{CVaR}_{1-\epsilon}\,(\tilde{v}) = \min_{\beta\in\mathbb{R}} \left(\beta + \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}}\left((\tilde{v}-\beta)^{+}\right)\right).$$

In particular, the constraint $\text{CVaR}_{1-\epsilon}\,(\tilde{v}) \leq 0$ would imply that $\mathbb{P}(\tilde{v} \leq 0) \geq 1-\epsilon$. As it is well-known that CVaR is the best convex approximation of chance constrained problems (see, for instance, [30]), we can formulate the TSPTW under uncertainty as

$$\begin{aligned} &\min\ \boldsymbol{c}'\boldsymbol{x} \\ &\ \text{s.t. } \text{CVaR}_{1-\epsilon_l}\,(\xi_l(\boldsymbol{s},\tilde{\boldsymbol{z}})) \leq 0, \quad \forall l \in \overline{\mathcal{N}}, \\ &\qquad (\boldsymbol{x},\boldsymbol{s}) \in \mathcal{S}, \end{aligned} \tag{23}$$

which is a safe approximation for the chance constrained TSPTW under uncertainty proposed in Laporte et al. [28] as follows,

$$\begin{aligned} &\min\ \boldsymbol{c}'\boldsymbol{x} \\ &\ \text{s.t. } \mathbb{P}\,(\xi_l(\boldsymbol{s},\tilde{\boldsymbol{z}}) \leq \overline{\tau}_l) \geq 1-\epsilon_l, \quad \forall l \in \overline{\mathcal{N}}, \\ &\qquad (\boldsymbol{x},\boldsymbol{s}) \in \mathcal{S}. \end{aligned} \tag{24}$$

Observe that Problem (23) can be explicitly written as

$$\begin{aligned} &\min\ \boldsymbol{c}'\boldsymbol{x} \\ &\ \text{s.t. } \beta_l + \frac{1}{\epsilon_l}\mathbb{E}_{\mathbb{P}}\left(\max\left\{\max_{k\in\underline{\mathcal{N}}\cup\{1\}}\left\{\sum_{a\in\delta^-(k)} s_a^l\underline{\tau}_k + \tilde{\boldsymbol{z}}'(\boldsymbol{s}^l-\boldsymbol{s}^k)\right\} - \overline{\tau}_l - \beta_l, 0\right\}\right) \leq 0, \\ &\qquad\qquad \forall l \in \overline{\mathcal{N}}, \\ &\qquad (\boldsymbol{x},\boldsymbol{s}) \in \mathcal{S}, \\ &\qquad \beta_l \in \mathbb{R}, \quad \forall l \in \overline{\mathcal{N}}. \end{aligned} \tag{25}$$

which has similar structure as Problem (8) and, as well, its distributioanally robust counterpart is similar to Problem (10). Hence, to address this variant of the TSPTW under uncertainty, we can straightforwardly extend the solution techniques that we have proposed in this paper.

We briefly present the idea for solving the stochastic programming problem (25). Due to the intractability for calculating the sum of random variables, we routinely solve its sample average approximation reformulation (26) as follows.

$$
\begin{aligned}
&\min\ \boldsymbol{c}'\boldsymbol{x} \\
&\text{s.t. } \beta_l + \frac{1}{\epsilon_l|\Omega|}\sum_{\omega\in\Omega} y_l^\omega \le 0, && \forall l \in \overline{\mathcal{N}}, \\
&\quad y_l^\omega \ge \sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k + \boldsymbol{z}(\omega)'(\boldsymbol{s}^l - \boldsymbol{s}^k) - \overline{\tau}_l - \beta_l, && \forall l \in \overline{\mathcal{N}}, k \in \underline{\mathcal{N}} \cup \{1\}, \omega \in \Omega, \\
&\quad (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S}. \\
&\quad y_l^\omega \ge 0, && \forall l \in \overline{\mathcal{N}}, \omega \in \Omega, \\
&\quad \beta_l \in \mathbb{R}, && \forall l \in \overline{\mathcal{N}}.
\end{aligned}
\tag{26}
$$

Here, we use $\boldsymbol{z}(\omega)$, $\omega \in \Omega$, to denote the ω-th sample of travel times $\tilde{\boldsymbol{z}}$. The auxiliary decision variable, $y_l^\omega = \max\{\max_{k\in\underline{\mathcal{N}}\cup\{1\}}\{\sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k + \boldsymbol{z}(\omega)'(\boldsymbol{s}^l - \boldsymbol{s}^k)\} - \overline{\tau}_l - \beta_l, 0\}$, $l \in \overline{\mathcal{N}}$, $\omega \in \Omega$, is introduced to linearize the reformulation. The reformulation can be solved via state-of-the-art commercial solvers such as CPLEX and Gurobi directly, or via a more sophisticated Benders decomposition method. We assume that typical readers are familiar with Benders decomposition and next focus on solving its subproblem. We regard $(\boldsymbol{x}, \boldsymbol{s})$ as the restricted master problem's decision variables. Given the values of $(\boldsymbol{x}, \boldsymbol{s})$, we formulate the Benders subproblem for some $l \in \overline{\mathcal{N}}$ as follows.

$$
\begin{aligned}
F_l(\boldsymbol{s}) = \min\ & \beta_l + \frac{1}{\epsilon_l|\Omega|}\sum_{\omega\in\Omega} y_l^\omega \\
\text{s.t. } & y_l^\omega + \beta_l \ge \xi_l^\omega(\boldsymbol{s}), && \forall \omega \in \Omega, \\
& y_l^\omega \ge 0, && \forall \omega \in \Omega, \\
& \beta_l \in \mathbb{R}, && \forall l \in \overline{\mathcal{N}},
\end{aligned}
\tag{27}
$$

where $\xi_l^\omega(\boldsymbol{s}) = \max_{k\in\underline{\mathcal{N}}\cup\{1\}}\left\{\sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k + \boldsymbol{z}(\omega)'(\boldsymbol{s}^l - \boldsymbol{s}^k) - \overline{\tau}_l\right\}$. Its decision variables are β_l and $(y_l^\omega)_{\omega\in\Omega}$. Problem (27) is a linear programming problem, for which the strong duality holds. We then formulate its dual problem as follows.

$$
\begin{aligned}
F_l(\boldsymbol{s}) = \max\ & \sum_{\omega\in\Omega} \xi_l^\omega(\boldsymbol{s}) p_l^\omega \\
\text{s.t. } & \sum_{\omega\in\Omega} p_l^\omega = 1 \\
& p_l^\omega \le \frac{1}{\epsilon_l|\Omega|}, && \forall \omega \in \Omega, \\
& p_l^\omega \ge 0, && \forall \omega \in \Omega,
\end{aligned}
\tag{28}
$$

in which $(p_l^\omega)_{\omega\in\Omega}$ are the decision variables. Observe that the objective function can be interpreted as an expectation of a discrete random variable taking values at $(\xi_l^\omega(\boldsymbol{s}))_{\omega\in\Omega}$ with masses $(p_l^\omega)_{\omega\in\Omega}$, respectively. To maximize the expectation, we can sort $(\xi_l^\omega(\boldsymbol{s}))_{\omega\in\Omega}$ and greedily determine their masses. Specifically, we let the index function $\nu : \Omega \mapsto \Omega$ be a permutation of Ω such that

$$
\xi_l^{\nu(1)}(\boldsymbol{s}) \ge \xi_l^{\nu(2)}(\boldsymbol{s}) \ge \cdots \ge \xi_l^{\nu(|\Omega|)}(\boldsymbol{s}).
$$

We then assign mass $\frac{1}{\epsilon_l|\Omega|}$ for each of the first $\lfloor\epsilon_l|\Omega|\rfloor$ values, and the remaining mass for the $(\lfloor\epsilon_l|\Omega|\rfloor + 1)$-th value. Therefore, we have a closed-form solution to the dual problem, given as follows.

$$F_l(\boldsymbol{s}) = \sum_{i=1}^{\lfloor\epsilon_l|\Omega|\rfloor} \frac{\xi_l^{\nu(i)}(\boldsymbol{s})}{\epsilon_l|\Omega|} + \left(1 - \sum_{i=1}^{\lfloor\epsilon_l|\Omega|\rfloor} \frac{1}{\epsilon_l|\Omega|}\right) \xi_l^{\nu(\lfloor\epsilon_l|\Omega|\rfloor+1)}(\boldsymbol{s}), \tag{29}$$

where we define $\sum_{i=1}^{0} \xi = 0$ and $\xi_l^{\nu(|\Omega|+1)} = 0$. The result can be rigorously proved in a similar way as the proof of Theorem 3. We leave this and the development of the distributionally robust optimization method as an exercise to the reader.

7 Computational study

As a proof of concept of our proposed models, we perform numerical studies to understand their computational efficiency and to elucidate their effectiveness in mitigating travel and service times uncertainty in TSPTW and TSPD. In particular, we consider a simple directed network of 12 nodes, $\mathcal{N} = \{1, 2, \dots, 12\}$ with the nodes 1 and 12 being respectively the origin and the destination depots. The set of arcs is given by $\mathcal{A} = \{(i, j) \mid i \in \mathcal{N}\setminus\{n\}, j \in \mathcal{N}\setminus\{1\}, i \neq j, (i, j) \neq (1, 12)\}$ and hence, there are a total of $|\mathcal{A}| = 110$ arcs. Let $\tilde{z}_{ij}^T$ and $\tilde{z}_i^S$ denote respectively the random travel time along arc $(i, j) \in \mathcal{A}$ and the random service time at node $i \in \mathcal{N}$. Note that the service times at nodes 1 and 12 have zero values. By definition, we have $\tilde{z}_{ij} = \tilde{z}_i^S + \tilde{z}_{ij}^T$ for $(i, j) \in \mathcal{A}$. We assume that $\tilde{z}_a^T$, $a \in \mathcal{A}$ is a two-point independently distributed random variable with mean z_a^T so that $\mathbb{P}(\tilde{z}_a^T = (1 - \lambda_a)z_a^T) = \mathbb{P}(\tilde{z}_a^T = (1 + \lambda_a)z_a^T) = 0.5$ for some $\lambda_a > 0$. Likewise, $\tilde{z}_i^S$, $i \in \mathcal{N}\setminus\{1, 12\}$ is also a similar two-point independently distributed random variable with mean z_i^S so that $\mathbb{P}(\tilde{z}_i^S = (1 - \lambda_i)z_i^S) = \mathbb{P}(\tilde{z}_i^S = (1 + \lambda_i)z_i^S) = 0.5$ for some $\lambda_i > 0$.

The parameters z_a^T and z_i^S, $a \in \mathcal{A}, i \in \mathcal{N}$ are given in Table 1 and for each instance of the problem, λ_a and λ_i are randomly and independently selected from the set $\{0.1, 0.2, \dots, 0.8\}$ with equal probability. These parameters of the problem are adopted from the instance `rbg010a` of Ascheuer et al. [4]. For simplicity, we do not impose a cost budget constraint in our experiments. As shown in Table 1, three of the nodes have stipulated earliest arrival times and eight of the nodes have deadlines.

7.1 Experiments with sample average approximation methods on TSPTW

Suppose we have $|\Omega|$ independent samples of travel times $\tilde{z}$, $(z(\omega))_{\omega\in\Omega}$. We solve the stochastic TSPTW through the following methods.

i) *D*: Solving the "deterministic" problem (3) that minimizes the travel cost by a CPLEX solver, in which the travel times $\boldsymbol{z}$ are replaced with their sample means $|\Omega|^{-1} \sum_{\omega\in\Omega} \boldsymbol{z}(\omega)$. Hereafter, we let the travel costs $\boldsymbol{c}$ also be equal to $|\Omega|^{-1} \sum_{\omega\in\Omega} \boldsymbol{z}(\omega)$ in values.

Table 1 The dataset for the TSPTW

$\mathcal{N}$	z_a^T												z_i^S	$\underline{\tau}$	$\overline{\tau}$
	1	2	3	4	5	6	7	8	9	10	11	12			
1	–	0	0	0	0	0	0	0	0	0	0	–	–	–	–
2	–	–	14	14	14	25	14	14	26	14	6	0	71	–	0
3	–	15	–	27	27	12	27	27	12	27	24	0	50	–	400
4	–	10	24	–	24	16	24	24	18	24	23	0	64	–	400
5	–	10	24	24	–	16	24	24	18	24	23	0	44	–	400
6	–	24	0	0	0	–	0	0	29	0	14	0	51	300	600
7	–	11	25	25	25	15	–	25	16	25	23	0	53	300	600
8	–	16	18	18	18	24	18	–	24	18	12	0	51	300	600
9	–	18	28	28	28	0	28	28	–	28	25	0	43	–	–
10	–	11	25	25	25	15	25	25	16	–	27	0	53	–	–
11	–	24	10	10	10	28	10	10	28	10	–	0	42	–	–
12	–	–	–	–	–	–	–	–	–	–	–	–	–	–	700

ii) *S-C*: Solving Problem (26) that minimizes the travel cost subject to the deadlines' CVaR constraints by a CPLEX solver. Based on some preliminary results, we find that the solution performance is affected by the parameters $\epsilon_l \in (0, 1)$, $l \in \overline{\mathcal{N}}$. Small values for them may cause infeasibility for the problem. In our experiments, we let $\epsilon_l = 0.8$ for $l \in \overline{\mathcal{N}}$.

iii) *S-C-B*: Solving Problem (26) via the Benders decomposition algorithm described in Sect. 6, in which the Benders subproblems are solved through Formula (29). We use a CPLEX solver to solve the restricted master problem and invoke function `ILOLAZYCONSTRAINTCALLBACK` to add the Benders cuts in a branch-and-cut fashion. We let $\epsilon_l = 0.8$ for $l \in \overline{\mathcal{N}}$.

iv) *S-P*: Solving Problem (4) that maximizes the joint on-time arrival probability via solving the following sample average approximation reformulation (30) by a CPLEX solver.

$$
\begin{aligned}
\max\ & |\Omega|^{-1} \textstyle\sum_{\omega \in \Omega} I_\omega \\
\text{s.t.}\ & \sum_{a \in \delta^-(k)} s_a^l \underline{\tau}_k + \boldsymbol{z}(\omega)' \left(\boldsymbol{s}^l - \boldsymbol{s}^k\right) - \overline{\tau}_l \le (1 - I_\omega) M_l,\ \forall k \in \underline{\mathcal{N}} \cup \{1\}, l \in \overline{\mathcal{N}}, \omega \in \Omega, \\
& \boldsymbol{c}'\boldsymbol{x} \le B, \\
& (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S}, \\
& I_\omega \in \{0, 1\}, \qquad\qquad \forall \omega \in \Omega.
\end{aligned} \tag{30}
$$

Here, binary decision variable I_ω indicates whether the vehicle arrives at all nodes on time in sample $\omega \in \Omega$. If it does, $I_\omega = 1$, otherwise $I_\omega = 0$. Notation M_l, $l \in \underline{\mathcal{N}}$, represents a big number. We choose $M_l := \max_{k \in \underline{\mathcal{N}}} \underline{\tau}_k + \max_{\omega \in \Omega} \sum_{a \in \mathcal{A}} z_a(\omega) - \overline{\tau}_l$, which is obviously an upper bound of $\sum_{a \in \delta^-(k)} s_a^l \underline{\tau}_k + \boldsymbol{z}(\omega)' \left(\boldsymbol{s}^l - \boldsymbol{s}^k\right) - \overline{\tau}_l$.

v) *S-I*: Solving Problem (9) that minimizes the sum of essential riskiness indexes by a CPLEX solver.
vi) *S-I-B*: Solving Problem (9) by the branch-and-cut implementation of the Benders decomposition algorithm described in Sect. 5, in which we solve the Benders dual subproblems based on Theorem 3.

All these methods are implemented using the C++ language and the IBM CPLEX solver (ver 12.6). We adopt the 8-thread computation provided by CPLEX and impose a time limit of 3 hours, unless otherwise stated, for solving each instance via each method. The experiments are run on a personal computer with a 4-core 3.2GHz CPU and a 8GB RAM.

For each of the above method under evaluation, we vary the the sample size, $|\Omega| \in \{20, 50, 80, 100, 150\}$ to understand its influence on the computational times as well as the quality of the solutions. For a given sample size, we perform 20 set of experiments. In each set, we randomly generate the parameters $(\lambda_a)_{a\in\mathcal{A}}$, $(\lambda_i)_{i\in\mathcal{N}\setminus\{1,12\}}$ to first establish the probability distributions associated with the random variables for the set of experiments. Subsequently, we solve the problem using the method via sample average approximation where we generate $|\Omega|$ independent samples of $\tilde{z}$, $(z(\omega))_{\omega\in\Omega}$. In particular, we use the same set of samples $(z(\omega))_{\omega\in\Omega}$ to compute the optimal solutions for the various methods. Thereafter, to evaluate the performance of these solutions, we perform out-of-sample evaluation by generating another 20,000 independent samples of $\tilde{z}$. After completing the 20 set of experiments, we report the performance by taking the average values of the individual indicators obtained in each set. In particular, we focus on the following indicators:

i) *ObjVal*: the objective value,
ii) *Mean*: the mean travel time,
iii) *LateProb*: the lateness probability, i.e., $\mathbb{P}(\exists i \in \overline{\mathcal{N}} | \xi_i(\boldsymbol{s}, \tilde{z}) > 0)$,
iv) *ExpLate*: the expected lateness, i.e., $\sum_{i\in\overline{\mathcal{N}}} \mathbb{E}((\xi_i(\boldsymbol{s}, \tilde{z}))^+)$, and
v) *ComTime*: the wall-clock computational time (in seconds),
vi) *NoC/ToC*: the number of the Benders cuts added/the "wall-clock time equivalence" (in seconds) for solving the Benders subproblems, estimated using the accumulated CPU time divided by 8 (threads).

In the above definitions, the notation ξ^+ represents $\max\{\xi, 0\}$; the probability $\mathbb{P}$ and the expectation $\mathbb{E}$ are evaluated empirically using the 20,000 out-of-sample data.

We report the solution performances for the various methods in Table 2. The number in the bracket in column *Method* represents the number of instances without feasible solutions out of the 20 instances. When $|\Omega| = 3000$, Methods S-C, S-P and S-I cannot solve each instance due to the time limit imposed. In contrast, we are able to obtain the optimal solution for Methods S-I-B and S-I-C without much computational time.

We conclude from Table 2 that Methods D, S-C, and S-C-I outperform Methods S-P, S-I, and S-I-B in terms of the average travel time *Mean*, while the latter three methods are more effective in mitigating the lateness, as indicated by *LateProb* and *ExpLate*. This phenomenon is a natural result of the objectives of these methods.

Method S-C presents a better performance than Method D in mitigating the lateness, and meanwhile is slightly better in reducing the average travel time expect for the case of $|\Omega| = 50$. Although Method D accounts for the mean values of travel times,

Table 2 Performance comparison of different methods for solving the TSPTW

Method	$\|\Omega\|$	*ObjVal*	*Mean*	*LatePro*	*ExpLate*	*ComTime*	*NoC/ToC*
D	20	630.43	642.70	0.485	53.9	0.39	–
S-C(3)		626.00	641.64	0.388	32.5	5.04	–
S-C-B(3)		626.00	641.64	0.388	32.5	1.83	118/0.19
S-P		0.80	650.78	0.337	31.5	351.02	–
S-I		23.69	646.22	0.309	23.3	394.48	–
S-I-B		23.69	646.22	0.309	23.3	460.31	119/0.46
D	50	635.80	641.79	0.467	50.7	0.34	–
S-C(1)		637.15	642.02	0.359	29.4	12.51	–
S-C-B(1)		637.15	642.02	0.359	29.4	2.26	153/0.42
S-P		0.76	646.13	0.306	24.7	988.92	–
S-I		28.44	644.35	0.294	19.8	1015.54	–
S-I-B		28.44	644.35	0.294	19.8	503.47	139/0.42
D	80	637.44	641.48	0.490	56.4	0.46	–
S-C		637.16	640.96	0.340	27.9	22.11	–
S-C-B		637.16	640.96	0.340	27.9	2.69	173/0.75
S-P		0.74	646.16	0.308	25.8	2539.41	–
S-I		27.68	644.78	0.293	19.2	2126.59	–
S-I-B		27.68	644.78	0.293	19.2	556.62	158/0.53
D	100	635.41	641.63	0.486	59.1	0.35	–
S-C		638.14	641.02	0.347	29.8	25.90	–
S-C-B		638.14	641.02	0.347	29.8	2.44	144/0.64
S-P		0.75	646.86	0.301	25.1	3983.28	–
S-I		26.11	645.02	0.292	19.4	3626.74	–
S-I-B		26.11	645.02	0.292	19.4	431.43	161/0.82
D	150	638.17	641.42	0.485	55.2	0.39	–
S-C		636.53	641.05	0.342	27.6	42.62	–
S-C-B		636.53	641.05	0.342	27.6	4.54	256/0.23
S-P		0.74	646.33	0.293	23.4	7824.18	–
S-I		27.52	644.42	0.288	19.3	6570.30	–
S-I-B		27.52	644.42	0.288	19.3	494.61	192/1.17
D	3000	640.70	641.13	0.479	48.6	0.39	–
S-C-B		639.81	640.96	0.351	23.4	68.27	328/50.25
S-I-B		27.58	643.14	0.286	18.6	526.29	297/14.16

its general poor performance may result from ignoring the travel times' dispersions. Method S-C provides a trade-off between the low average travel time and the high likelihood of on-time arrivals.

We next focus on comparing the latter three methods regarding their performance in mitigating the lateness. We observe that the performance of both indicators *LateProb* and *ExpLate* generally improves with the sample size, $|\Omega|$, which is consistent with our expectations. In particular, Method S-I-B with 3000 samples has the best performance. We also observe that for the same sample size, Method S-I outperforms Method S-P in both performance indicators, which could be rather surprising since we may have expected Method S-P to have better performance on the *LateProb* indicator. Nevertheless, when the sample size is large enough, we would expect Method S-P to yield solutions that are at least as good over those of Method S-I on this indicator. However, on the flip side, it may also be computationally prohibitive to solve these problems to optimality using Method S-P.

Method D consumes the least computational time and the time is insensitive to the sample size, which is consistent with our expectation since the deterministic model has less decision variables and the number of decision variables is independent of the sample size. Method S-C requires more computational time; Methods S-P and S-I need far more. We also observe that the computational times of methods S-C, S-P and S-I are severely affected by the sample size, $|\Omega|$ and are in stark contrast to Methods S-C-B and S-I-B, where the total computational times, the numbers of Benders cuts added, and the computational times for solving Benders subproblems are only slightly impacted. Benders decomposition method S-C-B/S-I-B greatly improves the computational speed upon its counterpart S-C/S-I when $|\Omega| \geq 50$ in our experiment. As the sample size increases, the computational times of Methods S-P and S-I increase super-linearly, with the former increasing at a faster rate than the latter. We note that the superior computational performance of Method S-I-B underscores the effectiveness of the Benders decomposition approach.

In summary, we suggest the decision makers who attempt to minimize the travel costs using Method S-C-B. Those who are seeking to mitigate the lateness are suggested choosing Method S-I-B. We encourage them to use a large number of samples (e.g., $|\Omega| = 3000$) in solving the problems.

7.2 Experiments with distributionally robust TSPD

We had initially tried to extend the computational study to distributionally robust TSPTW by implementing the corresponding branch-and-cut variant of the Benders decomposition algorithm described in Sect. 5, in which we used a MOSEK solver (ver 8.1) to obtain the solutions of the semidefinite programming subproblem based on Theorem 4. Preliminary results showed that it could take 4–8 min to obtain the optimal solution for each subproblem, or if unbounded, obtain its recession direction, and we were unable to obtain the optimal solution for any of the instances of the distributionally robust TSPTW within two days. The sluggish computational time might result from long time to solve each Benders subproblem, a large number of Benders cuts to add, and the inability to get MOSEK to solve subproblems in parallel. Hence, we abandon the numerical study for TSPTW and focus on distributionally robust TSPD, where we can obtain the solutions to the subproblems easily.

We repeat the aforementioned computational study without the earliest arrival time $\underline{\tau}$ in Table 1 so that the TSPTW would be reduced to a TSPD. To concentrate on comparing the methods for mitigating the lateness, we omit Methods D, S-C, and S-C-B, and indicator *Mean*. Apart from Methods *S-P*, *S-I*, and *S-I-B*, we introduce two more methods based on distributionally robust optimization.

i) *R-I-B*: Solving the distributionally robust TSPD (13) by the branch-and-cut implementation of the Benders decomposition algorithm described in Sect. 5, where the closed form solutions of the subproblems are computed via Theorem 5. In particular, the means $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ in the cross moment ambiguity set are estimated empirically from the samples of travel times, that is, we let $\boldsymbol{\mu} = |\Omega|^{-1} \sum_{\omega \in \Omega} \boldsymbol{z}(\omega)$ and each element of $\boldsymbol{\Sigma}$, the covariance of $\tilde{z}_a$ and $\tilde{z}_{\hat{a}}$, $a, \hat{a} \in \mathcal{A}$, be $|\Omega|^{-1} \sum_{\omega \in \Omega} (z_a(\omega) - \mu_a)(z_{\hat{a}}(\omega) - \mu_{\hat{a}})$.

ii) *R-RI-B*: Solving the riskiness-index-based distributionally robust TSPD (31) by using a branch-and-cut implementation of the Benders decomposition algorithm proposed by Jaillet et al. [24].

$$
\begin{aligned}
\inf \ & \sum_{l \in \overline{\mathcal{N}}} \alpha_l \\
\text{s.t.} \ & \alpha_l \ln \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(\exp \left(\frac{\tilde{\boldsymbol{z}}' \boldsymbol{s}^l - \overline{\tau}_l}{\alpha_l} \right) \right) \leq 0, \ \forall l \in \overline{\mathcal{N}}, \\
& \boldsymbol{c}' \boldsymbol{x} \leq B, \\
& (\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{S}, \\
& \alpha_l \geq 0, \qquad \forall l \in \overline{\mathcal{N}}.
\end{aligned}
\tag{31}
$$

In this problem, the ambiguity set $\mathbb{F}$ is given as

$$
\mathbb{F} = \left\{ \mathbb{P} \in \mathcal{P} \ \middle| \ \begin{array}{l} \mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{z}}) = \boldsymbol{\mu}, \\ \mathbb{P}\left(\tilde{\boldsymbol{z}} \in [\underline{\boldsymbol{z}}, \overline{\boldsymbol{z}}] \right) = 1, \end{array} \right\},
$$

where the means $\boldsymbol{\mu}$ and supports $[\underline{\boldsymbol{z}}, \overline{\boldsymbol{z}}]$ of travel times $\tilde{\boldsymbol{z}}$ are estimated empirically from the samples. In particular, we let $\underline{z}_a = \min_{\omega \in \Omega} z_a(\omega)$ and $\overline{z}_a = \max_{\omega \in \Omega} z_a(\omega)$ for $a \in \mathcal{A}$. We regard $(\boldsymbol{x}, \boldsymbol{s})$ as the restricted master problem's decision variables and solve the Benders subproblem (32) for each $l \in \overline{\mathcal{N}}$ using the technique proposed by Jaillet et al. [24].

$$
\begin{aligned}
\inf \ & \alpha_l \\
\text{s.t.} \ & \alpha_l \ln \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(\exp \left(\frac{\tilde{\boldsymbol{z}}' \boldsymbol{s}^l - \overline{\tau}_l}{\alpha_l} \right) \right) \leq 0, \\
& \alpha_l \geq 0.
\end{aligned}
\tag{32}
$$

We report the results in Table 3. When the number of samples is small, $|\Omega| \leq 100$, Method R-RI-B has the best performance in terms of *LatePro* and *ExpLate*, but it requires significantly longer computational time than the others. We expect Method R-RI-B to perform well because the travel times are independently distributed, which can be exploited by the Aumann and Serrano [5] riskiness index. Methods R-I-B and S-I-B performs reasonably well against R-RI-B. Nonetheless, when we have 3000 samples of travel times, Method S-I-B has the best overall performance.

It is also interesting to note that despite the limited use of distributional information, the performance of the distributionally robust TSPD is relatively close to the stochastic TSPD. Hence, since the actual distribution may not be available in many practical

Table 3 Performance comparison of different methods for solving the TSPD

Method	$\|\Omega\|$	*ObjVal*	*LatePro*	*ExpLate*	*ComTime*	*NoC/ToC*
S-P	20	0.81	0.337	28.3	89.26	–
S-I		21.17	0.309	23.4	85.73	–
S-I-B		21.17	0.309	23.4	173.82	137/0.55
R-RI-B		136.17	0.302	20.6	586.87	2524/11.48
R-I-B		64.71	0.318	23.4	216.25	441/0.49
S-P	50	0.77	0.318	25.7	194.10	–
S-I		26.65	0.291	19.6	176.39	–
S-I-B		26.65	0.291	19.6	195.70	178/0.68
R-RI-B		123.95	0.290	18.8	627.40	2452/9.83
R-I-B		73.08	0.301	19.9	213.95	438/0.51
S-P	80	0.75	0.302	23.6	340.23	–
S-I		27.46	0.291	19.6	228.79	–
S-I-B		27.46	0.291	19.6	202.01	183/0.66
R-RI-B		122.52	0.286	18.4	693.77	2845/12.95
R-I-B		75.05	0.298	19.5	211.74	483/0.53
S-P	100	0.76	0.298	23.1	482.48	–
S-I		24.22	0.289	19.3	335.92	–
S-I-B		24.22	0.289	19.3	208.76	185/0.77
R-RI-B		119.11	0.287	18.3	689.77	2956/12.05
R-I-B		69.67	0.296	19.5	201.22	456/0.49
S-P	150	0.74	0.286	20.3	902.31	–
S-I		25.50	0.283	18.7	488.53	–
S-I-B		25.50	0.283	18.7	210.90	212/0.95
R-RI-B		124.79	0.284	18.2	611.10	2675/10.06
R-I-B		71.48	0.295	19.2	201.44	462/0.54
S-I-B	3000	26.40	0.279	18.2	216.23	332/10.16
R-RI-B		123.14	0.285	18.2	637.37	2647/10.72
R-I-B		73.26	0.287	18.6	189.98	479/0.46

situations, it is reassuring that the solutions to the distributioanlly robust TSPD are near optimal. Moreover, to exemplify the benefits of the distributional robust solutions, there are already computational studies suggesting that if the assumed distribution used in the sample average approximation deviates from the true distribution, the solutions may be inferior to those obtained from distributionally robust models (see, for instance, [1]).

8 Conclusions and future research

We study a TSPTW with hard time windows under uncertain travel and service times. To quantify the risk associated with deadline violations, we propose the essential riskiness index as the decision criterion to be minimized, which has the salient properties such as convexity for coherent decision making and computational needs. We also propose algorithms to minimize the index for the TSPTW via Benders decomposition technique based on sample average approximation and distributionally robust formuations. We demonstrate through a computational study that that our approach can outperform the approach that maximizes punctuality probability via sample average approximations. Apart from TSPTW under uncertainty, the application of the essential riskiness index is quite broad and can be applied in other contexts that may arise involving multiple agents and the criterion will help them collectively attain their targets as well as possible under uncertainty. Nevertheless, we have yet to adequately address the computational efficiency of our approach as we are unable to solve larger sized problems within reasonable time. Hence, further work will be needed to investigate how we can address larger sized problems, perhaps by leveraging on the state-of-the-art vehicle routing techniques in the literature.

Acknowledgements The authors would like to thank the anonymous reviewers and associate editor for their extremely helpful suggestions and very thorough review of the paper. The first author is financially supported by the Chinese Scholarship Council, the National Research Foundation Singapore (NRF-RSS2016-004), and the National Natural Science Foundation of China (71725001). The third author is financially supported by Singapore Ministry of Education Social Science Research Thematic Grant (MOE2016-SSRTG-059). The last author is funded by the National Natural Science Foundation of China (71420107028, 91747105).

Disclaimer Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Singapore Ministry of Education or the Singapore Government.

A Proofs of analytical results

A.1 Proof of Proposition 2

It has been established that the variable $\boldsymbol{s}^l, l \in \mathcal{N}$, represents the partial route from node 1 to node l along the route determined by $\boldsymbol{x}$, say $\{(1, i_2), (i_2, i_3), \ldots, (i_{\check{\kappa}-1}, \underbrace{i_{\check{\kappa}}}_{=l}),$ $\ldots, (i_{n-1}, \underbrace{i_n}_{=n})\}$ so that

$$s_a^l = \begin{cases} 1, & \text{if } a \in \{(1, i_2), (i_2, i_3), \ldots, (i_{\check{\kappa}-1}, l)\}, \\ 0, & \text{otherwise} \end{cases} \tag{33}$$

(see, for instance, [24]). We have from Eq. (1) that the service commencement time at the node l can be determined recursively as follows,

$$\begin{aligned} t_1 &= \underline{\tau}_1 = 0, \\ t_{i_2} &= \max\{t_1 + z_{1i_2}, \underline{\tau}_{i_2}\}, \\ t_{i_3} &= \max\{t_{i_2} + z_{i_2 i_3}, \underline{\tau}_{i_3}\}, \\ &\ \ \vdots \\ t_l &= \max\{t_{i_{\check{\kappa}-1}} + z_{i_{\check{\kappa}-1}l}, \underline{\tau}_l\}. \end{aligned}$$

Hence, we have

$$t_l = \max\left\{\max\left\{\cdots\max\left\{\max\left\{t_1 + z_{1i_2}, \underline{\tau}_{i_2}\right\} + z_{i_2 i_3}, \underline{\tau}_{i_3}\right\}\cdots, \underline{\tau}_{i_{\check{\kappa}-1}}\right\} + z_{i_{\check{\kappa}-1}l}, \underline{\tau}_l\right\}$$
$$= \max_{k\in\{1,i_2,i_3,\ldots,i_{\check{\kappa}-1},l\}}\left\{\underline{\tau}_k + \sum_{a\in\{(k,i_{\hat{\kappa}+1}),(i_{\hat{\kappa}+1},i_{\hat{\kappa}+2}),\ldots,(i_{\check{\kappa}-1},l)\}} z_a\right\}, \tag{34}$$

where $k = i_{\hat{\kappa}}$ represents a node along the partial route from node 1 to node l.

Next we show that the service commencement time at node l can also determined by Eq. (2). For notational convenience, we define

$$t_l^k = \sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k + z'\left(\boldsymbol{s}^l - \boldsymbol{s}^k\right),$$

for all $k \in \underline{\mathcal{N}} \cup \{1\}$ and consider three cases as follows:

i) When $k = 1$, we have $\underline{\tau}_1 = 0$ and $\boldsymbol{s}^1 = \boldsymbol{0}$ and hence, we have

$$t_l^1 = \sum_{a\in\delta^-(1)} s_a^l \underline{\tau}_1 + \boldsymbol{z}'(\boldsymbol{s}^l - \boldsymbol{s}^1) = \underline{\tau}_1 + \sum_{a\in\{(1,i_2),(i_2,i_3),\ldots,(i_{\check{\kappa}-1},l)\}} z_a.$$

Moreover, $t_l^1 \geq 0$ because $\boldsymbol{z} \geq \boldsymbol{0}$.

ii) When $k \in \{i_2, i_3, \ldots, i_{\check{\kappa}-1}, l\} \cap \underline{\mathcal{N}}$, we have $\sum_{a\in\delta^-(k)} s_a^l = 1$ and

$$t_l^k = \sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k + \boldsymbol{z}'(\boldsymbol{s}^l - \boldsymbol{s}^k) = \underline{\tau}_k + \sum_{a\in\{(k,i_{\hat{\kappa}+1}),(i_{\hat{\kappa}+1},i_{\hat{\kappa}+2}),\ldots,(i_{\check{\kappa}-1},l)\}} z_a.$$

iii) When $k \in \{i_{\check{\kappa}+1}, i_{\check{\kappa}+2}, \ldots, n\} \cap \underline{\mathcal{N}}$, we have $\sum_{a\in\delta^-(k)} s_a^l = 0$ and $(\boldsymbol{s}^l - \boldsymbol{s}^k) \leq \boldsymbol{0}$. Since $\boldsymbol{z} \geq \boldsymbol{0}$, we have

$$t_l^k = \sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k + \boldsymbol{z}'(\boldsymbol{s}^l - \boldsymbol{s}^k) \leq 0 \leq t_l^1.$$

Hence, $\max\{t_l^1, t_l^k\} = t_l^1$ for all $k \in \{i_{\check{\kappa}+1}, i_{\check{\kappa}+2}, \ldots, n\} \cap \underline{\mathcal{N}}$.

Observe that we can express Eq. (2) as

$$t_l = \max_{k\in\underline{\mathcal{N}}\cup\{1\}} t_l^k = \max_{k\in\{1\}\cup(\{i_2,i_3,\ldots,i_{\check{\kappa}-1},l\}\cap\underline{\mathcal{N}})}\left\{\underline{\tau}_k + \sum_{a\in\{(k,i_{\hat{\kappa}+1}),(i_{\hat{\kappa}+1},i_{\hat{\kappa}+2}),\ldots,(i_{\check{\kappa}-1},l)\}} z_a\right\}. \tag{35}$$

We note that for all $k \in \{i_2, i_3, \ldots, i_{\hat{\kappa}-1}, l\}\backslash\underline{\mathcal{N}}$ we have $\underline{\tau}_k = 0$ and since $\boldsymbol{z} \geq \boldsymbol{0}$, we also have

$$\underline{\tau}_k + \sum_{a\in\{(k,i_{\hat{\kappa}+1}),(i_{\hat{\kappa}+1},i_{\hat{\kappa}+2}),\ldots,(i_{\check{\kappa}-1},l)\}} z_a \leq \underline{\tau}_1 + \sum_{a\in\{(1,i_2),(i_2,i_3),\ldots,(k,i_{\hat{\kappa}+1}),\ldots,(i_{\check{\kappa}-1},l)\}} z_a = t_l^1.$$

Hence, $\max\{t_l^1, t_l^k\} = t_l^1$ for all $k \in \{i_2, i_3, \ldots, i_{\check{\kappa}-1}, l\}\backslash\underline{\mathcal{N}}$. Therefore, taking these conditions into account, we have shown the equivalence of Eqs. (34), (35) and (2).

A.2 Proof of Proposition 3

i) *Satisficing:* If $\mathbb{P}\left(\tilde{\xi} \leq 0\right) = 1$, then $\mathbb{E}_{\mathbb{P}}\left(\max\{\tilde{\xi}, -\alpha\}\right) \leq 0$ for all $\alpha \geq 0$ and $\rho_E(\tilde{\xi}) = 0$. Conversely, if $\rho_E(\tilde{\xi}) = 0$, we have $\mathbb{E}_{\mathbb{P}}\left(\max\{\tilde{\xi}, 0\}\right) \leq 0$, which implies $\mathbb{P}\left(\tilde{\xi} \leq 0\right) = 1$.

ii) *Infeasibility:* For any $\alpha \in [0, \infty)$, if $\mathbb{E}_{\mathbb{P}}\left(\tilde{\xi}\right) > 0$, then $\mathbb{E}_{\mathbb{P}}\left(\max\{\tilde{\xi}, -\alpha\}\right) > 0$. However, the definition requires

$$\mathbb{E}_{\mathbb{P}}\left(\max\{\tilde{\xi}, -\alpha\}\right) \leq 0.$$

We conclude that $\rho_E(\tilde{\xi}) = \min \emptyset = \infty$.

iii) *Convexity:* We denote $\alpha_1^* = \rho_E\left(\tilde{\xi}_1\right)$ and $\alpha_2^* = \rho_E\left(\tilde{\xi}_2\right)$. Hence, $\mathbb{E}_{\mathbb{P}}\left(\max\{\tilde{\xi}_1, -\alpha_1^*\}\right) \leq 0$ and $\mathbb{E}_{\mathbb{P}}\left(\max\{\tilde{\xi}_2, -\alpha_2^*\}\right) \leq 0$. Since the function $\mathbb{E}_{\mathbb{P}}\left(\max\{\tilde{\xi}, -\alpha\}\right)$ is jointly convex in $\tilde{\xi}$ and α, we have for $\lambda \in [0, 1]$,

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\left(\max\{\lambda\tilde{\xi}_1 + (1-\lambda)\tilde{\xi}_2, -\left(\lambda\alpha_1^* + (1-\lambda)\alpha_2^*\right)\}\right)\\
&\quad \leq \lambda\mathbb{E}_{\mathbb{P}}\left(\max\{\tilde{\xi}_1, -\alpha_1^*\}\right) + (1-\lambda)\mathbb{E}_{\mathbb{P}}\left(\max\{\tilde{\xi}_2, -\alpha_2^*\}\right)\\
&\quad \leq 0.
\end{aligned}$$

Hence, $\alpha_\lambda = \lambda\alpha_1^* + (1-\lambda)\alpha_2^*$ satisfies $\mathbb{E}_{\mathbb{P}}\left(\max\{\lambda\tilde{\xi}_1 + (1-\lambda)\tilde{\xi}_2, -\alpha_\lambda\}\right) \leq 0$ and $\rho_E\left(\lambda\tilde{\xi}_1 + (1-\lambda)\tilde{\xi}_2\right) \leq \alpha_\lambda = \lambda\rho_E\left(\tilde{\xi}_1\right) + (1-\lambda)\rho_E\left(\tilde{\xi}_2\right)$.

iv) *Delay bounds:* The bound is true for $\rho_E\left(\tilde{\xi}\right) = \infty$ and $\rho_E\left(\tilde{\xi}\right) = 0$, since the latter would imply $\mathbb{P}\left(\tilde{\xi} > 0\right) = 0$. For $\rho_E\left(\tilde{\xi}\right) \in (0, \infty)$, we let $\alpha^* = \rho_E\left(\tilde{\xi}\right)$ and hence we have

$$\begin{aligned}
\mathbb{P}\left(\tilde{\xi} > \alpha^*\theta\right) &= \mathbb{P}\left(\tilde{\xi} + \alpha^* > \alpha^*\theta + \alpha^*\right)\\
&\leq \mathbb{P}\left(\left(\tilde{\xi} + \alpha^*\right)^+ > \alpha^*(1+\theta)\right)\\
&\leq \frac{\mathbb{E}\left((\tilde{\xi} + \alpha^*)^+\right)}{\alpha^*(1+\theta)}\\
&\leq \frac{\alpha^*}{\alpha^*(1+\theta)}\\
&= \frac{1}{1+\theta}.
\end{aligned}$$

The second inequality holds because of Markov inequality; the third inequality is due to $\mathbb{E}_{\mathbb{P}}\left(\max\{\tilde{\xi}+\alpha^*, 0\}\right) \le \alpha^*$, implied by the definition.

A.3 Proof of Theorem 1

We denote $\tilde{\boldsymbol{u}} = \tilde{\boldsymbol{z}} - \boldsymbol{\mu}$ and rewrite the ambiguity set as

$$\mathbb{G} = \left\{ \mathbb{Q} \in \mathcal{P} \;\middle|\; \begin{array}{l} \mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{u}}) = \boldsymbol{0} \\ \mathbb{E}_{\mathbb{Q}}\left(\tilde{\boldsymbol{u}}\tilde{\boldsymbol{u}}'\right) = \boldsymbol{\Sigma} \end{array} \right\}.$$

For each $l \in \overline{\mathcal{N}}$, we determine the worse-case expectation

$$\sup_{\mathbb{Q}\in\mathbb{G}} \mathbb{E}_{\mathbb{Q}}\left(\max\left\{\max_{k\in\underline{\mathcal{N}}\cup\{1\}}\left\{\sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k + (\tilde{\boldsymbol{u}}+\boldsymbol{\mu})'(\boldsymbol{s}^l - \boldsymbol{s}^k)\right\} - \overline{\tau}_l + \alpha_l, 0\right\}\right)$$

by formulating the following optimization problem:

$$\begin{aligned}
\sup\ & \int_{\mathbb{R}^{|\mathcal{A}|}} \left(\max\left\{\max_{k\in\underline{\mathcal{N}}\cup\{1\}}\left\{\sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k + (\boldsymbol{u}+\boldsymbol{\mu})'\left(\boldsymbol{s}^l - \boldsymbol{s}^k\right)\right\} - \overline{\tau}_l + \alpha_l, 0\right\}\right) d\mathbb{Q} \\
\text{s.t.}\ & \int_{\mathbb{R}^{|\mathcal{A}|}} \boldsymbol{u}\, d\mathbb{Q} = \boldsymbol{0}, \\
& \int_{\mathbb{R}^{|\mathcal{A}|}} \boldsymbol{u}\boldsymbol{u}'\, d\mathbb{Q} = \boldsymbol{\Sigma}, \\
& \int_{\mathbb{R}^{|\mathcal{A}|}} d\mathbb{Q} = 1, \\
& d\mathbb{Q} \ge 0.
\end{aligned}$$

Its dual problem is given as

$$\begin{aligned}
\inf\ & v_{l0} + \operatorname{tr}(\boldsymbol{\Sigma}\boldsymbol{V}_l) \\
\text{s.t.}\ & v_{l0} + \boldsymbol{u}'\boldsymbol{v}_l + \boldsymbol{u}'\boldsymbol{V}_l\boldsymbol{u} \ge 0, && \forall \boldsymbol{u} \in \mathbb{R}^{|\mathcal{A}|}, \\
& v_{l0} + \boldsymbol{u}'\boldsymbol{v}_l + \boldsymbol{u}'\boldsymbol{V}_l\boldsymbol{u} \ge \sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k + (\boldsymbol{u}+\boldsymbol{\mu})'(\boldsymbol{s}^l - \boldsymbol{s}^k) - \overline{\tau}_l + \alpha_l, && \forall k \in \underline{\mathcal{N}} \cup \{1\}, \\
& && \boldsymbol{u} \in \mathbb{R}^{|\mathcal{A}|},
\end{aligned} \tag{36}$$

for which strong duality holds and their objectives coincide (see, for instance, [23]). We express the first constraint in Problem (36) equivalently as

$$\begin{bmatrix} 1 \\ \boldsymbol{u} \end{bmatrix}' \begin{bmatrix} v_{l0} & \dfrac{\boldsymbol{v}_l'}{2} \\ \dfrac{\boldsymbol{v}_l}{2} & \boldsymbol{V}_l \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{u} \end{bmatrix} \ge 0, \quad \forall \boldsymbol{u} \in \mathbb{R}^{|\mathcal{A}|},$$

or further as

$$\begin{bmatrix} v_{l0} & \frac{\boldsymbol{v}_l'}{2} \\ \frac{\boldsymbol{v}_l}{2} & \boldsymbol{V}_l \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1}.$$

Similarly, we express the second constraint in Problem (36) equivalently as

$$\begin{bmatrix} v_{l0} - \sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k - \boldsymbol{\mu}'(\boldsymbol{s}^l - \boldsymbol{s}^k) + \overline{\tau}_l - \alpha_l & \frac{(\boldsymbol{v}_l - \boldsymbol{s}^l + \boldsymbol{s}^k)'}{2} \\ \frac{\boldsymbol{v}_l - \boldsymbol{s}^l + \boldsymbol{s}^k}{2} & \boldsymbol{V}_l \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1}.$$

Hence, the first constraint of Problem (10) is satisfied if and only if there exists $v_{l0} \in \mathbb{R}$, $\boldsymbol{v}_l \in \mathbb{R}^{|\mathcal{A}|}$ and $\boldsymbol{V}_l \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$ for $l \in \overline{\mathcal{N}}$ that are feasible in the following system of conic constraints,

$$\begin{aligned}
& v_{l0} + \mathrm{tr}(\boldsymbol{\Sigma} \boldsymbol{V}_l) \le \alpha_l, && \forall l \in \overline{\mathcal{N}}, \\
& \begin{bmatrix} v_{l0} & \frac{\boldsymbol{v}_l'}{2} \\ \frac{\boldsymbol{v}_l}{2} & \boldsymbol{V}_l \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1}, && \forall l \in \overline{\mathcal{N}}, \\
& \begin{bmatrix} v_{l0} - \sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k - \boldsymbol{\mu}'(\boldsymbol{s}^l - \boldsymbol{s}^k) + \overline{\tau}_l - \alpha_l & \frac{(\boldsymbol{v}_l - \boldsymbol{s}^l + \boldsymbol{s}^k)'}{2} \\ \frac{\boldsymbol{v}_l - \boldsymbol{s}^l + \boldsymbol{s}^k}{2} & \boldsymbol{V}_l \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1}, && \forall l \in \overline{\mathcal{N}}, \\
& && k \in \underline{\mathcal{N}} \cup \{1\}.
\end{aligned}$$

Substituting them in Problem (10), we obtain Problem (11).

A.4 Proof of Theorem 2

It suffices to show that for a given $s_0 \in \mathbb{R}$, $\boldsymbol{s} \in \mathbb{R}^{|\mathcal{A}|}$, $\boldsymbol{s} \neq \boldsymbol{0}$, $\alpha \ge 0$, the following constraints are feasible

$$\begin{aligned}
& v_0 + \mathrm{tr}(\boldsymbol{\Sigma} \boldsymbol{V}) \le \alpha \\
& \begin{bmatrix} v_0 - s_0 & \frac{(\boldsymbol{v} - \boldsymbol{s})'}{2} \\ \frac{\boldsymbol{v} - \boldsymbol{s}}{2} & \boldsymbol{V} \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1}, \\
& \begin{bmatrix} v_0 & \frac{\boldsymbol{v}'}{2} \\ \frac{\boldsymbol{v}}{2} & \boldsymbol{V} \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|+1},
\end{aligned} \tag{37}$$

for some $v_0 \in \mathbb{R}$, $\boldsymbol{v} \in \mathbb{R}^{|\mathcal{A}|}$, $\boldsymbol{V} \in \mathbb{S}_+^{|\mathcal{A}|}$ if and only if

$$\boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s} \leq 4\alpha(\alpha - s_0). \tag{38}$$

For the "if" direction, suppose the constraint (38) is feasible, since $\boldsymbol{s}'\boldsymbol{\Sigma}'\boldsymbol{s} > 0$, we would have $\alpha > 0$ and $\alpha - s_0 > 0$. Let

$$\begin{aligned}
v_0 &= \frac{\left(s_0 + \sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}\right)^2}{4\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}}, \\
\boldsymbol{v} &= \frac{\left(s_0 + \sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}\right)\boldsymbol{s}}{2\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}}, \\
\boldsymbol{V} &= \frac{\boldsymbol{s}\boldsymbol{s}'}{4\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}}.
\end{aligned}$$

Observe that for all $\boldsymbol{u} \in \mathbb{R}^{|\mathcal{A}|}$,

$$\begin{aligned}
&\begin{bmatrix} 1 \\ \boldsymbol{u} \end{bmatrix}' \begin{bmatrix} v_0 & \frac{\boldsymbol{v}'}{2} \\ \frac{\boldsymbol{v}}{2} & \boldsymbol{V} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{u} \end{bmatrix} \\
&= \frac{\left(s_0 + \sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}\right)^2}{4\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}} + \frac{\left(s_0 + \sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}\right)}{2\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}}\boldsymbol{s}'\boldsymbol{u} + \frac{(\boldsymbol{s}'\boldsymbol{u})^2}{4\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}} \\
&= \frac{\left(s_0 + \sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}} + \boldsymbol{s}'\boldsymbol{u}\right)^2}{4\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}} \geq 0,
\end{aligned}$$

and

$$\begin{aligned}
&\begin{bmatrix} 1 \\ \boldsymbol{u} \end{bmatrix}' \begin{bmatrix} v_0 - s_0 & \frac{(\boldsymbol{v} - \boldsymbol{s})'}{2} \\ \frac{(\boldsymbol{v} - \boldsymbol{s})}{2} & \boldsymbol{V} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{u} \end{bmatrix} \\
&= \frac{\left(s_0 + \sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}\right)^2}{4\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}} - s_0 + \left(\frac{\left(s_0 + \sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}\right)}{2\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}} - 1\right)\boldsymbol{s}'\boldsymbol{u} + \frac{(\boldsymbol{s}'\boldsymbol{u})^2}{4\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}}
\end{aligned}$$

$$
= \frac{\left(s_0 - \sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}\right)^2}{4\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}} + \frac{\left(s_0 - \sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}\right)}{2\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}}\boldsymbol{s}'\boldsymbol{u} + \frac{(\boldsymbol{s}'\boldsymbol{u})^2}{4\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}}
$$

$$
= \frac{\left(s_0 - \sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}} + \boldsymbol{s}'\boldsymbol{u}\right)^2}{4\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}} \geq 0.
$$

Moreover,

$$
\begin{aligned}
&\boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s} \leq 4\alpha(\alpha - s_0) \\
&\quad\Rightarrow s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s} \leq (2\alpha - s_0)^2 \\
&\quad\Rightarrow \sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}} \leq 2\alpha - s_0 && \text{since } 2\alpha - s_0 > \alpha - s_0 > 0 \\
&\quad\Rightarrow \frac{1}{2}s_0 + \frac{1}{2}\sqrt{s^0 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}} \leq \alpha \\
&\quad\Rightarrow \frac{2s_0\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}} + 2(s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s})}{4\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}} \leq \alpha && \text{since } \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s} > 0 \\
&\quad\Rightarrow \frac{\left(s_0 + \sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}\right)^2}{4\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}} + \frac{\boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}{4\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}} \leq \alpha \\
&\quad\Rightarrow v_0 + \operatorname{tr}(\boldsymbol{\Sigma}\boldsymbol{V}) \leq \alpha.
\end{aligned}
$$

Hence, the constraints in (37) are also feasible.

Conversely, suppose the constraints in (37) are feasible for some $v_0 \in \mathbb{R}$, $\boldsymbol{v} \in \mathbb{R}^{|\mathcal{A}|}$, $\boldsymbol{V} \in \mathbb{S}_+^{|\mathcal{A}|}$. Let

$$
\begin{aligned}
r_0 &= \frac{1}{2} + \frac{s_0}{2\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}}, \\
\boldsymbol{r} &= \frac{\boldsymbol{\Sigma}\boldsymbol{s}}{2\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}},
\end{aligned}
$$

Note that since $\boldsymbol{s}\boldsymbol{\Sigma}\boldsymbol{s} > 0$, we have $r_0 \in (0, 1)$. Observe that

$$
\begin{bmatrix} r_0 & \boldsymbol{r}' \\ \boldsymbol{r} & \frac{1}{r_0}\boldsymbol{r}\boldsymbol{r}' \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|} \quad \text{and} \quad \begin{bmatrix} 1 - r_0 & -\boldsymbol{r}' \\ -\boldsymbol{r} & \boldsymbol{\Sigma} - \frac{1}{r_0}\boldsymbol{r}\boldsymbol{r}' \end{bmatrix} \in \mathbb{S}_+^{|\mathcal{A}|}
$$

since for all $\boldsymbol{w} \in \mathbb{R}^{|\mathcal{A}|}$,

$$
\begin{bmatrix} 1 \\ \boldsymbol{w} \end{bmatrix}' \begin{bmatrix} r_0 & \boldsymbol{r}' \\ \boldsymbol{r} & \frac{1}{r_0}\boldsymbol{r}\boldsymbol{r}' \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{w} \end{bmatrix} = r_0 + 2\boldsymbol{r}'\boldsymbol{w} + \frac{1}{r_0}(\boldsymbol{r}'\boldsymbol{w})^2 = \left(\sqrt{r_0} + \frac{1}{\sqrt{r_0}}\boldsymbol{r}'\boldsymbol{w}\right)^2 \geq 0
$$

and

$$
\begin{aligned}
&\begin{bmatrix} 1 \\ \boldsymbol{w} \end{bmatrix}' \begin{bmatrix} 1-r_0 & -\boldsymbol{r}' \\ -\boldsymbol{r} & \boldsymbol{\Sigma} - \frac{1}{r_0}\boldsymbol{r}\boldsymbol{r}' \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{w} \end{bmatrix} \\
&= 1 - r_0 - 2\boldsymbol{r}'\boldsymbol{w} - \frac{1}{r_0}(\boldsymbol{r}'\boldsymbol{w})^2 + \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \\
&= \left(\sqrt{1-r_0} - \frac{1}{\sqrt{1-r_0}}\boldsymbol{r}'\boldsymbol{w}\right)^2 - \left(\frac{1}{r_0} + \frac{1}{1-r_0}\right)(\boldsymbol{r}'\boldsymbol{w})^2 + \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \\
&\geq -\frac{1}{r_0(1-r_0)}(\boldsymbol{r}'\boldsymbol{w})^2 + \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \\
&= -\frac{4\left(s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}\right)}{\boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}(\boldsymbol{r}'\boldsymbol{w})^2 + \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \\
&= -\frac{1}{\boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}(\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{s})^2 + \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \\
&= -\frac{1}{\boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}((\boldsymbol{\Sigma}^{1/2}\boldsymbol{w})'\boldsymbol{\Sigma}^{1/2}\boldsymbol{s})^2 + \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \\
&\geq -\frac{1}{\boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}(\boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s})(\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}) + \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \qquad \text{Cauchy–Schwarz inequality} \\
&= 0.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\operatorname{tr}\left(\begin{bmatrix} v_0 & \frac{\boldsymbol{v}'}{2} \\ \frac{\boldsymbol{v}}{2} & \boldsymbol{V} \end{bmatrix} \begin{bmatrix} 1-r_0 & -\boldsymbol{r}' \\ -\boldsymbol{r} & \boldsymbol{\Sigma} - \frac{1}{r_0}\boldsymbol{r}\boldsymbol{r}' \end{bmatrix}\right) \\
&\quad + \operatorname{tr}\left(\begin{bmatrix} v_0 - s_0 & \frac{(\boldsymbol{v}-\boldsymbol{s})'}{2} \\ \frac{\boldsymbol{v}-\boldsymbol{s}}{2} & \boldsymbol{V} \end{bmatrix} \begin{bmatrix} r_0 & \boldsymbol{r}' \\ \boldsymbol{r} & \frac{1}{r_0}\boldsymbol{r}\boldsymbol{r}' \end{bmatrix}\right) \geq 0
\end{aligned}
$$

or equivalently

$$
v_0 + \operatorname{tr}(\boldsymbol{\Sigma}\boldsymbol{V}) \geq s_0 r_0 + \boldsymbol{r}'\boldsymbol{s} = \frac{1}{2}s_0 + \frac{1}{2}\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}}.
$$

Hence,

$$
\frac{1}{2}s_0 + \frac{1}{2}\sqrt{s_0^2 + \boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}} \leq v_0 + \operatorname{tr}(\boldsymbol{\Sigma}\boldsymbol{V}) \leq \alpha,
$$

which implies the feasibility of the constraint (38).

A.5 Proof of Proposition 4

Denote the optimal solutions of Problem (14) with $\boldsymbol{s}$ being $\check{\boldsymbol{s}}$ and $\hat{\boldsymbol{s}}$ by $\check{\alpha}_l$ and $\hat{\alpha}_l$, respectively, so that $F_l(\check{\boldsymbol{s}}) = \check{\alpha}_l$ and $F_l(\hat{\boldsymbol{s}}) = \hat{\alpha}_l$. To prove the convexity of $F_l(\boldsymbol{s})$ by definition, we shall prove $F_l(\lambda\check{\boldsymbol{s}} + (1-\lambda)\hat{\boldsymbol{s}}) \leq \lambda F_l(\check{\boldsymbol{s}}) + (1-\lambda)F_l(\hat{\boldsymbol{s}})$ for any $\lambda \in [0, 1]$.

Observe that the left-hand side function of the first constraint in Problem (14), $g(\boldsymbol{s}, \alpha_l) = \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}\left(\max\left\{\max_{k\in\underline{\mathcal{N}}\cup\{1\}}\left\{\sum_{a\in\delta^-(k)} s_a^l \underline{\tau}_k + \tilde{z}'(\boldsymbol{s}^l - \boldsymbol{s}^k)\right\} - \overline{\tau}_l, -\alpha_l\right\}\right)$, is convex piecewise affine in $(\boldsymbol{s}, \alpha_l)$. We then have, for any $\lambda \in [0, 1]$,

$$
\begin{aligned}
g\left(\lambda\check{\boldsymbol{s}} + (1-\lambda)\hat{\boldsymbol{s}}, \lambda\check{\alpha} + (1-\lambda)\hat{\alpha}\right) &\le \lambda g(\check{\boldsymbol{s}}, \check{\alpha}) + (1-\lambda) g(\hat{\boldsymbol{s}}, \hat{\alpha}),\\
&\le \lambda 0 + (1-\lambda)0.\\
&= 0.
\end{aligned}
$$

Also note that $\lambda\check{\alpha} + (1-\lambda)\hat{\alpha} \ge 0$. Hence, $\lambda\check{\alpha} + (1-\lambda)\hat{\alpha}$ is a feasible solution to Problem (14) with $\boldsymbol{s}$ being $\lambda\check{\boldsymbol{s}} + (1-\lambda)\hat{\boldsymbol{s}}$. As a result,

$$
F_l(\lambda\check{\boldsymbol{s}} + (1-\lambda)\hat{\boldsymbol{s}}) \le \lambda\check{\alpha} + (1-\lambda)\hat{\alpha} = \lambda F_l(\check{\boldsymbol{s}}) + (1-\lambda) F_l(\hat{\boldsymbol{s}}).
$$

This completes the proof.

A.6 Proof of Theorem 3

We specify the dual problem of Problem (17) as follows,

$$
\begin{aligned}
\hat{F}_l(\boldsymbol{s}) = \max & \sum_{\omega\in\Omega}\sum_{k\in\underline{\mathcal{N}}\cup\{1\}} \xi_{lk}^{\omega}(\boldsymbol{s}) r_k^{\omega} \\
\text{s.t.} & -\psi + \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} r_k^{\omega} + q^{\omega} = 0, \ \forall\omega\in\Omega,\\
& \sum_{\omega\in\Omega} q^{\omega} \le 1, \\
& \psi \ge 0,\\
& r_k^{\omega} \ge 0, \qquad k \in \underline{\mathcal{N}}\cup\{1\}, \omega\in\Omega,\\
& q^{\omega} \ge 0, \qquad \forall\omega\in\Omega,
\end{aligned}
\tag{39}
$$

which we claim is equivalent to the following problem,

$$
\begin{aligned}
\check{F}_l(\boldsymbol{s}) = \max & \sum_{\omega\in\Omega} \xi_l^{\omega}(\boldsymbol{s}) p^{\omega}\\
\text{s.t.} & \ |\Omega| p^{\omega} - \sum_{i\in\Omega} p^i \le 1, \ \forall\omega\in\Omega,\\
& p^{\omega} \ge 0, \qquad \forall\omega\in\Omega.
\end{aligned}
\tag{40}
$$

We next prove the claim. For an optimal solution $(\psi, \boldsymbol{r}, \boldsymbol{q})$ to Problem (39), we infer from the first and the fifth constraints that $\psi \ge \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} r_k^{\omega}$, $\forall\omega\in\Omega$, and therefore, $q^i = \psi - \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} r_k^i \ge \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} r_k^{\omega} - \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} r_k^i$, $\forall i, \omega \in \Omega$. This inequality and the second constraint in Problem (39) imply that $|\Omega| \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} r_k^{\omega} - \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} \sum_{i\in\Omega} r_k^i \le 1$, $\forall\omega\in\Omega$. Hence, the solution $p^{\omega} := \sum_{k\in\underline{\mathcal{N}}\cup\{1\}} r_k^{\omega}$, $\forall\omega\in\Omega$, is clearly feasible in Problem (40). Since, by definition $\xi_l^\omega(\boldsymbol{s}) = \max_{k\in\underline{\mathcal{N}}\cup\{1\}} \xi_{lk}^\omega(\boldsymbol{s})$, it follows that $\check{F}_l(\boldsymbol{s}) \geq \hat{F}_l(\boldsymbol{s})$. Conversely, for an optimal solution $\boldsymbol{p}$ to Problem (40), we construct a solution to Problem (39) as follows. Let $r^\omega_{\kappa^*(\omega)} := p^\omega$ for all $\omega \in \Omega$ where $\kappa^*(\omega) \in \arg\max_{k\in\underline{\mathcal{N}}\cup\{1\}} \xi_{lk}^\omega(\boldsymbol{s})$, $r_k^\omega := 0$ for all $\omega \in \Omega$ and $k \in \underline{\mathcal{N}} \cup \{1\}\backslash\{\kappa^*(\omega)\}$, $\psi := \max_{\omega\in\Omega} p^\omega$, and $q^\omega := \max_{i\in\Omega} p^i - p^\omega$ for all $\omega \in \Omega$. We can verify that the solution is feasible and the objective value is equal to $\check{F}_l(\boldsymbol{s})$. In particular, Constraint $\sum_{\omega\in\Omega} q^\omega \leq 1$ is feasible since $|\Omega|p^\omega - \sum_{i\in\Omega} p^i \leq 1$ for all $\omega \in \arg\max_{i\in\Omega} p^i$. Hence, $\hat{F}_l(\boldsymbol{s}) \geq \check{F}_l(\boldsymbol{s})$. We then conclude that $\hat{F}_l(\boldsymbol{s}) = \check{F}_l(\boldsymbol{s})$ and the two problems are equivalent. Observe that $\boldsymbol{0}$ is a feasible solution to Problem (40); the problem then can be bounded optimal or unbounded. Since strong duality holds for linear programming problems, we also have $\hat{F}_l(\boldsymbol{s}) = F_l(\boldsymbol{s})$.

We next solve Problem (40). The problem is unbounded if and only if there exists a recession direction $\boldsymbol{p} \neq \boldsymbol{0}$, $|\Omega|p^\omega - \sum_{i\in\Omega} p^i \leq 0$ and $p^\omega \geq 0$ for $\omega \in \Omega$, such that $\sum_{\omega\in\Omega} \xi_l^\omega(\boldsymbol{s})p^\omega > 0$. The recession direction should satisfy $p^1 = p^2 = \cdots = p^{|\Omega|} > 0$, or otherwise the constraint $|\Omega|p^\omega - \sum_{i\in\Omega} p^i \leq 0$ would be violated for all $\omega \in \arg\max_{i\in\Omega} p^i$. Hence, the condition $\sum_{\omega\in\Omega} \xi_l^\omega(\boldsymbol{s})p^\omega > 0$ is equivalent to $\sum_{\omega\in\Omega} \xi_l^\omega(\boldsymbol{s}) > 0$, and in which case, the problem is unbounded and $F_l(\boldsymbol{s}) = \infty$.

If the problem is bounded and optimal, there exists an optimal extreme point solution, which is determined by selecting $|\Omega|$ binding constraints in Problem (40). Observe that for a given $\omega \in \Omega$ the constraints $p^\omega \geq 0$ and $|\Omega|p^\omega - \sum_{i\in\Omega} p^i \leq 1$ cannot be simultaneously binding because it would imply that $-\sum_{i\in\Omega\backslash\{\omega\}} p^i = 1$, which contradicts with the nonnegativity of $\boldsymbol{p}$. Hence, partition Ω into two subsets Ω_1 and Ω_2, such that $|\Omega|p^\omega - \sum_{i\in\Omega} p^i = 1$, $\forall\omega \in \Omega_1$ and $p^\omega = 0$, $\forall\omega \in \Omega_2$. When $\Omega_1 = \emptyset$, we have $\boldsymbol{p} = \boldsymbol{0}$ being an extreme point with objective value of zero. We also note that $|\Omega_1| = |\Omega|$ is inadmissible because we would have $\sum_{\omega\in\Omega}(|\Omega|p^\omega - \sum_{i\in\Omega} p^i) = \sum_{\omega\in\Omega} 1$, which is a contradiction unless $|\Omega| = 0$. When $1 \leq |\Omega_1| \leq |\Omega| - 1$, we solve for the system of linear equations, $|\Omega|p^\omega - \sum_{i\in\Omega} p^i = 1$, $\forall\omega \in \Omega_1$, or equivalently, $(|\Omega|\boldsymbol{I} - \boldsymbol{e}\boldsymbol{e}')\boldsymbol{p} = \boldsymbol{e}$, where $\boldsymbol{I}$ is the unit diagonal matrix in $\mathbb{R}^{|\Omega_1|\times|\Omega_1|}$, $\boldsymbol{p} = (p^\omega)'_{\omega\in\Omega_1}$, and $\boldsymbol{e}$ is the vector of ones in $\mathbb{R}^{|\Omega_1|}$. Observe that $\boldsymbol{p} = \frac{1}{|\Omega|-|\Omega_1|}\boldsymbol{e}$ is a feasible solution. Moreover, it is the unique solution because $(|\Omega|\boldsymbol{I} - \boldsymbol{e}\boldsymbol{e}')$ is invertible and we can verify that $(|\Omega|\boldsymbol{I} - \boldsymbol{e}\boldsymbol{e}')(\frac{1}{|\Omega|}\boldsymbol{I} + \frac{1}{|\Omega|(|\Omega|-|\Omega_1|)}\boldsymbol{e}\boldsymbol{e}') = \boldsymbol{I}$. Hence, $p^\omega = \frac{1}{|\Omega|-|\Omega_1|}$, $\forall\omega \in \Omega_1$ and $p^\omega = 0$, $\forall\omega \in \Omega_2$ is an extreme point solution. Observe that at optimality, Ω_1 would correspond to the set $\{\nu(1), \nu(2), \ldots, \nu(|\Omega_1|)\}$ and the objective value would be $\sum_{\omega=1}^{|\Omega_1|} \frac{\xi_l^{\nu(\omega)}(\boldsymbol{s})}{|\Omega|-|\Omega_1|}$. Finally, we conclude that

$$\check{F}_l(\boldsymbol{s}) = \max\left\{\max_{i\in\{1,2,\ldots,|\Omega|-1\}}\left\{\sum_{\omega=1}^{i} \frac{\xi_l^{\nu(\omega)}(\boldsymbol{s})}{|\Omega| - i}\right\}, 0\right\},$$

and the result follows.

A.7 Proof of Theorem 4

Let β, $\begin{bmatrix} \gamma_0, & \boldsymbol{\gamma}' \\ \boldsymbol{\gamma}, & \boldsymbol{\Gamma} \end{bmatrix}$, and $\begin{bmatrix} r_{k0}, & \boldsymbol{r}'_k \\ \boldsymbol{r}_k, & \boldsymbol{R}_k \end{bmatrix}$, $k \in \underline{\mathcal{N}} \cup \{1\}$, be the dual variables corresponding to the first three constraints in Problem (19), respectively, where $\boldsymbol{\gamma} \in \mathbb{R}^{|\mathcal{A}|}$, $\boldsymbol{\Gamma} \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{A}|}$, $\boldsymbol{r}_k \in \mathbb{R}^{|\mathcal{A}|}$, $\boldsymbol{R}_k \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{A}|}$. Based on the theory of conic duality (see, for instance, [8]), we obtain its dual problem (20). Moreover, their objectives coincide because the Slater's condition holds, i.e., there exists a strictly relative interior point for Problem (20), defined as $\beta = 1$, $\gamma_0 = r_{k0} = \frac{1}{|\underline{\mathcal{N}}|+2}$, $\boldsymbol{\gamma} = \boldsymbol{r}_k = \boldsymbol{0}$, and $\boldsymbol{\Gamma} = \boldsymbol{R}_k = \frac{\boldsymbol{\Sigma}}{|\underline{\mathcal{N}}|+2}$, $k \in \underline{\mathcal{N}} \cup \{1\}$, such that

$$
\begin{aligned}
&\beta - \sum_{k \in \underline{\mathcal{N}} \cup \{1\}} r_{k0} < 1, \\
&-\beta + \gamma_0 + \sum_{k \in \underline{\mathcal{N}} \cup \{1\}} r_{k0} = 0, \\
&\boldsymbol{\gamma} + \sum_{k \in \underline{\mathcal{N}} \cup \{1\}} \boldsymbol{r}_k = \boldsymbol{0}, \\
&-\boldsymbol{\Sigma}\beta + \boldsymbol{\Gamma} + \sum_{k \in \underline{\mathcal{N}} \cup \{1\}} \boldsymbol{R}_k = \boldsymbol{0}, \\
&\beta > 0, \\
&\begin{bmatrix} \gamma_0 & \boldsymbol{\gamma}' \\ \boldsymbol{\gamma} & \boldsymbol{\Gamma} \end{bmatrix} \in \mathbb{S}_{++}^{|\mathcal{A}|+1}, \\
&\begin{bmatrix} r_{k0} & \boldsymbol{r}'_k \\ \boldsymbol{r}_k & \boldsymbol{R}_k \end{bmatrix} \in \mathbb{S}_{++}^{|\mathcal{A}|+1}, \qquad \forall k \in \underline{\mathcal{N}} \cup \{1\},
\end{aligned}
$$

where $\mathbb{S}_{++}^{|\mathcal{A}|+1}$ represents the set of symmetric positive definite matrices in $\mathbb{R}^{(|\mathcal{A}|+1)\times(|\mathcal{A}|+1)}$. The last two constraints hold because

$$
\boldsymbol{\Sigma} \in \mathbb{S}_{++}^{|\mathcal{A}|} \Leftrightarrow \begin{bmatrix} \dfrac{1}{|\underline{\mathcal{N}}|+2} & \boldsymbol{0}' \\ \boldsymbol{0} & \dfrac{\boldsymbol{\Sigma}}{|\underline{\mathcal{N}}|+2} \end{bmatrix} \in \mathbb{S}_{++}^{|\mathcal{A}|+1}.
$$

A.8 Proof of Theorem 5

Given a restricted master problem's solution $\boldsymbol{s}^*$, we denote the objective values of Problem (21) and (22) by $\hat{F}(\boldsymbol{s}^*)$ and $\check{F}(\boldsymbol{s}^*)$, respectively. When $\boldsymbol{s}^{l*} = \boldsymbol{0}$, we observe that $\hat{F}(\boldsymbol{s}^*) = \check{F}(\boldsymbol{s}^*) = 0$. Otherwise, if $\overline{\tau}_l - \boldsymbol{\mu}'\boldsymbol{s}^{l*} > 0$, it is easy to see that

$$
\hat{F}(\boldsymbol{s}^*) = \frac{\boldsymbol{s}^{l*\prime}\boldsymbol{\Sigma}\boldsymbol{s}^{l*}}{4(\overline{\tau}_l - \boldsymbol{\mu}'\boldsymbol{s}^{l*})}.
$$

Based on the first-order condition, the optimal solution of Problem (22) is

$$\boldsymbol{r} = \frac{\boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{s}^{l*}}{2\left(\boldsymbol{\mu}' \boldsymbol{s}^{l*} - \overline{\tau}_l\right)}$$

and correspondingly,

$$\check{F}(\boldsymbol{s}^*) = \frac{\boldsymbol{s}^{l*\prime} \boldsymbol{\Sigma} \boldsymbol{s}^{l*}}{4(\overline{\tau}_l - \boldsymbol{\mu}' \boldsymbol{s}^{l*})} = \hat{F}(\boldsymbol{s}^*).$$

If $\overline{\tau}_l - \boldsymbol{\mu}' \boldsymbol{s}^{l*} \leq 0$, since $\boldsymbol{\Sigma}$ is positive definite, we observe that Problem (21) is infeasible and $\hat{F}(\boldsymbol{s}^*) = \infty$. Observe also that

$$\bar{\boldsymbol{r}} = -\boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{s}^{l*}$$

is an recession direction to Problem (22) such that $\check{F}(\boldsymbol{s}^*) = \infty = \hat{F}(\boldsymbol{s}^*)$. Hence, the result follows.

References

1. Adulyasak, Y., Jaillet, P.: Models and algorithms for stochastic and robust vehicle routing with deadlines. Transp. Sci. **50**(2), 608–626 (2015)
2. Agra, A., Christiansen, M., Figueiredo, R., Hvattum, L.M., Poss, M., Requejo, C.: Layered formulation for the robust vehicle routing problem with time windows. In: Lecture Notes in Computer Science. Springer, pp. 249–260 (2012)
3. Agra, A., Christiansen, M., Figueiredo, R., Hvattum, L.M., Poss, M., Requejo, C.: The robust vehicle routing problem with time windows. Comput. Oper. Res. **40**(3), 856–866 (2013)
4. Ascheuer, N., Fischetti, M., Grötschel, M.: Solving the asymmetric travelling salesman problem with time windows by branch-and-cut. Math. Program. **90**(3), 475–506 (2001)
5. Aumann, R.J., Serrano, R.: An economic index of riskiness. J. Polit. Econ. **116**(5), 810–836 (2008)
6. Baldacci, R., Bartolini, E., Mingozzi, A., Roberti, R.: An exact solution framework for a broad class of vehicle routing problems. Comput. Manag. Sci. **7**(3), 229–268 (2010)
7. Baldacci, R., Mingozzi, A., Roberti, R.: Recent exact algorithms for solving the vehicle routing problem under capacity and time window constraints. Eur. J. Oper. Res. **218**(1), 1–6 (2012)
8. Ben-Tal, A., Nemirovski, A.: Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications, vol. 2. SIAM, Philadelphia (2001)
9. Bertsimas, D., Sim, M.: Robust discrete optimization and network flows. Math. Program. **98**(1–3), 49–71 (2003)
10. Bertsimas, D., Sim, M., Zhang, M.: A practically efficient approach for solving adaptive distributionally robust linear optimization problems. Manag. Sci. (2017, Forthcoming)
11. Brown, D.B., Sim, M.: Satisficing measures for analysis of risky positions. Manag. Sci. **55**(1), 71–84 (2009)
12. Charnes, A., Cooper, W.W.: Chance-constrained programming. Manag. Sci. **6**(1), 73–79 (1959)
13. Claus, A.: A new formulation for the travelling salesman problem. SIAM J. Algebr. Discrete Methods **5**(1), 21–25 (1984)
14. Dantzig, G.B., Ramser, J.H.: The truck dispatching problem. Manag. Sci. **6**(1), 80–91 (1959)
15. Dash, S., Günlük, O., Lodi, A., Tramontani, A.: A time bucket formulation for the traveling salesman problem with time windows. INFORMS J. Comput. **24**(1), 132–147 (2012)
16. Desaulniers, G., Madsen, O.B., Ropke, S.: Chapter 5: The vehicle routing problem with time windows. In: Vehicle Routing: Problems, Methods, and Applications, 2nd edn. Society for Industrial & Applied Mathematics (SIAM), pp. 119–159 (2014)

17. Errico, F., Desaulniers, G., Gendreau, M., Rei, W., Rousseau, L.M.: A priori optimization with recourse for the vehicle routing problem with hard time windows and stochastic service times. Eur. J. Oper. Res. **249**(1), 55–66 (2016a)
18. Errico, F., Desaulniers, G., Gendreau, M., Rei, W., Rousseau, L.M.: The vehicle routing problem with hard time windows and stochastic service times. Eur. J. Transp. Logist. 1–29 (2016)
19. Fisher, M.L., Jörnsten, K.O., Madsen, O.B.: Vehicle routing with time windows: two optimization algorithms. Oper. Res. **45**(3), 488–492 (1997)
20. Gendreau, M., Jabali, O., Rei, W.: Chapter 8: Stochastic vehicle routing problems. In: MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics, pp. 213–239 (2014)
21. Gounaris, C.E., Wiesemann, W., Floudas, C.A.: The robust capacitated vehicle routing problem under demand uncertainty. Oper. Res. **61**(3), 677–693 (2013)
22. Gupta, D., Denton, B.: Appointment scheduling in health care: challenges and opportunities. IIE Trans. **40**(9), 800–819 (2008)
23. Isii, K.: On sharpness of tchebycheff-type inequalities. Ann. Inst. Stat. Math. **14**(1), 185–197 (1962)
24. Jaillet, P., Qi, J., Sim, M.: Routing optimization under uncertainty. Oper. Res. **64**(1), 186–200 (2016)
25. Kao, E.P.C.: A preference order dynamic program for a stochastic traveling salesman problem. Oper. Res. **26**(6), 1033–1045 (1978)
26. Kenyon, A.S., Morton, D.P.: Stochastic vehicle routing with random travel times. Transp. Sci. **37**(1), 69–82 (2003)
27. Laporte, G.: Fifty years of vehicle routing. Transp. Sci. **43**(4), 408–416 (2009)
28. Laporte, G., Louveaux, F., Mercure, H.: The vehicle routing problem with stochastic travel times. Transp. Sci. **26**(3), 161–170 (1992)
29. Lee, C., Lee, K., Park, S.: Robust vehicle routing problem with deadlines and travel time/demand uncertainty. J. Oper. Res. Soc. **63**(9), 1294–1306 (2011)
30. Nemirovski, A., Shapiro, A.: Convex approximations of chance constrained programs. SIAM J. Optim. **17**(4), 969–996 (2006)
31. Pecin, D., Contardo, C., Desaulniers, G., Uchoa, E.: New enhancements for the exact solution of the vehicle routing problem with time windows. INFORMS J. Comput. **29**(3), 489–502 (2017a)
32. Pecin, D., Pessoa, A., Poggi, M., Uchoa, E.: Improved branch-cut-and-price for capacitated vehicle routing. Math. Program. Comput. **9**(1), 61–100 (2017b)
33. Popescu, I.: Robust mean-covariance solutions for stochastic optimization. Oper. Res. **55**(1), 98–112 (2007)
34. Qi, J.: Mitigating delays and unfairness in appointment systems. Manag. Sci. **63**(2), 566–583 (2016)
35. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**, 21–42 (2000)
36. Scarf, H., Arrow, K., Karlin, S.: A min–max solution of an inventory problem. In: Studies in the Mathematical Theory of Inventory and Production, vol. 10, pp. 201–209 (1958)
37. See, C.T., Sim, M.: Robust approximation to multiperiod inventory management. Oper. Res. **58**(3), 583–594 (2010)
38. Sniedovich, M.: Technical note—analysis of a preference order traveling salesman problem. Oper. Res. **29**(6), 1234–1237 (1981)
39. Taş, D., Gendreau, M., Dellaert, N., van Woensel, T., de Kok, A.: Vehicle routing with soft time windows and stochastic travel times: a column generation and branch-and-price solution approach. Eur. J. Oper. Res. **236**(3), 789–799 (2014)
40. Toth, P., Vigo, D. (eds.): Vehicle Routing: Problems, Methods, and Applications, 2nd edn. Society for Industrial & Applied Mathematics (SIAM), Philadelphia (2014)
41. Vanderbeck, F., Wolsey, L.A.: Reformulation and decomposition of integer programs. In: 50 Years of Integer Programming 1958–2008, pp. 431–502 (2010)
42. Verweij, B., Ahmed, S., Kleywegt, A.J., Nemhauser, G., Shapiro, A.: The sample average approximation method applied to stochastic routing problems: a computational study. Comput. Optim. Appl. **24**(2–3), 289–333 (2003)
43. Wiesemann, W., Kuhn, D., Sim, M.: Distributionally robust convex optimization. Oper. Res. **62**(6), 1358–1376 (2014)
44. Xin, L., Goldberg, D.A., Shapiro, A.: Time (in) consistency of multistage distributionally robust inventory models with moment constraints. arXiv preprint arXiv:1304.3074 (2013)
45. Zhen, J., den Hertog, D., Sim, M.: Adjustable robust optimization via Fourier–Motzkin elimination. Oper. Res. (2017, Forthcoming)