The Distributionally Robust Chance Constrained Vehicle Routing Problem

Shubhechyya Ghosal[1] and Wolfram Wiesemann[1]

[1] *Imperial College Business School, Imperial College London, United Kingdom*

April 12, 2019

Abstract

We study a variant of the capacitated vehicle routing problem (CVRP), which asks for the cost-optimal delivery of a single product to geographically dispersed customers through a fleet of capacity-constrained vehicles. Contrary to the classical CVRP, which assumes that the customer demands are deterministic, we model the demands as a random vector whose distribution is only known to belong to an ambiguity set. We then require the delivery schedule to be feasible with a probability of at least $1 - \epsilon$, where ϵ characterizes the risk tolerance of the decision maker. We show that the emerging distributionally robust CVRP can be solved with standard branch-and-cut algorithms whenever the ambiguity set satisfies a subadditivity condition. We then argue that this subadditivity condition holds for a large class of moment ambiguity sets. We derive cut generation schemes for ambiguity sets that specify the support as well as (bounds on) the first and second moments of the customer demands. Our numerical results indicate that the distributionally robust CVRP has favorable scaling properties and can often be solved in runtimes comparable to those of the deterministic CVRP.

Keywords: vehicle routing, distributionally robust optimization, chance constraints.

1 Introduction

A fundamental problem in logistics concerns the cost-optimal delivery of a product from a depot to a set of geographically dispersed customers through a number of capacity-constrained vehicles. This problem, which is known as the capacitated vehicle routing problem (CVRP), has been studied since the 1950s (Dantzig and Ramser, 1959), and it has found wide-spread application in waste collection, dial-a-ride services, courier delivery and the routing of snow plow trucks, school buses as well as maintenance engineers. For a review of the vast literature on the CVRP, we refer the reader to Cordeau et al. (2006), Golden et al. (2008), Laporte (2009) and Toth and Vigo (2014).

The classical CVRP assumes that the customer demands are known precisely. This assumption is frequently violated in pickup problems such as residential waste collection, where the amount of waste to be collected is only known when the vehicle has arrived at an individual household. Perhaps surprisingly, customer demands are also often uncertain in delivery problems. Internet retailers, online groceries and delivery companies tend to use simplified models to estimate the vehicle space consumed by each customer order (which can itself consist of multiple heterogeneous products). The cumulative space consumed by all customer orders assigned to a vehicle thus becomes an uncertain quantity which depends on the shapes of the involved products, the employed stacking configuration, operational loading constraints as well as the packing skills of the staff involved.

The CVRP with uncertain customer demands is typically solved as a two-stage stochastic program or as a chance constrained program (Gendreau et al., 1996; Cordeau et al., 2006; Toth and Vigo, 2014). In the two-stage version, a tentative delivery schedule is selected here-and-now, that is, before the uncertain customer demands are known, and the routes can be modified through a recourse decision once the customer demands have been observed (*e.g.*, penalty payments for unsatisfied demands, Stewart and Golden 1983, detours to the depot, Dror and Trudeau 1986 and Bertsimas and Simchi-Levi 1996, or preventative restocking, Yang et al. 2000). The chance constrained CVRP, which we focus on in this paper, does not allow for any modification of the selected vehicles routes. Instead, it requires the vehicle routes to be feasible with a high, pre-specified probability. While being more restrictive than the two-stage model, the chance constrained CVRP can lead to simpler optimization problems, and it may be favored due to its planning stability.

Although the chance constrained CVRP reduces to a deterministic CVRP in special cases, *e.g*, when the demands are independent and identically distributed (Golden and Yee, 1979; Dror et al., 1993), the problem is typically solved with tailored branch-and-cut methods (Laporte et al., 1992). The vast majority of exact solution methods for the chance constrained CVRP assume that the customer demands are independent. A notable exception is Dinh et al. (2017), who develop a branch-and-cut-and-price scheme for the chance constrained CVRP under the assumption that the customer demands follow either a joint normal distribution or a given discrete distribution.

Despite its conceptual simplicity and its intuitive appeal, the chance constrained CVRP suffers from several statistical and computational shortcomings that may have contributed to its limited adoption in practice. First and foremost, the assumption of independent customer demands is likely to be violated in practice, and it can lead to severe misjudgment of the probability of satisfying a vehicle's capacity. Secondly, verifying whether a vehicle's capacity is met with high probability—a fundamental building block in exact solution methods for the CVRP—is itself typically a hard optimization problem. Finally, an unrealistically large amount of demand data may be required to estimate the probability distribution governing the customer demands with sufficient accuracy. We will revisit these points in further detail in Section 2 of this paper.

The aforementioned shortcomings of the chance constrained CVRP can to some degree be addressed by the robust CVRP, which abandons probability distributions and instead requires the vehicle routes to be feasible for all customer demands within a pre-specified uncertainty set (*e.g.*, a box, polyhedron or ellipsoid). Branch-and-cut schemes for the exact solution of the robust CVRP have been proposed by Sungur and Ordonez (2008) and Gounaris et al. (2013). The robust CVRP is amenable to solution schemes that appear to scale better than those for the chance constrained CVRP. However, the solutions obtained from the robust CVRP can be overly conservative since all demand scenarios within the uncertainty set are treated as equally likely, and the routes are selected solely in view of the worst demand scenario from the uncertainty set. Furthermore, the shape of the uncertainty set is often selected *ad hoc*, and it remains unclear how this set should be calibrated to historical demand data that may be available in practice.

In this paper, we study the distributionally robust chance constrained CVRP, which assumes that the customer demands follow a probability distribution that is only partially known, and it imposes chance constraints on the vehicles' capacities for all distributions that are deemed plausible in view of the available information. We argue that this formulation can offer an attractive trade-off between the properties of the classical chance constrained CVRP and the robust CVRP. By replacing a single probability distribution with a set of plausible distributions, the distributionally robust chance constrained CVRP relieves the decision maker of estimating the entire joint demand distribution for all customers, and it replaces computationally intractable operations on probability distributions with efficiently solvable optimization problems. Likewise, since the distributionally robust chance constrained CVRP does not abandon probability distributions altogether, it can determine delivery schedules that are less conservative than those of the robust CVRP.

While distributionally robust chance constraints have been considered for other problem classes (see, *e.g.*, the reviews by Ben-Tal and Nemirovski 2001, Nemirovski 2012 and Hanasusanto et al. 2015) and other classes of distributionally robust models have been proposed for variants of the vehicle routing problem (see, *e.g.*, Carlsson and Delage 2013, Adulyasak and Jaillet 2016, Jaillet et al. 2016, Carlsson et al. 2017, Flajolet et al. 2018 and Zhang et al. 2018), the only treatment of the distributionally robust chance constrained CVRP appears to be in the electronic companion of Gounaris et al. (2013) and in Section 4 of Dinh et al. (2017). Gounaris et al. (2013) approximate a particular class of distributionally robust chance constrained CVRPs by a robust CVRP and solve instances with up to 23 customers using a standard branch-and-bound scheme. Dinh et al. (2017) adapt their branch-and-cut-and-price scheme for the classical chance constrained CVRP to a distributionally robust chance constrained CVRP where the uncertain customer demands are characterized by their means and covariances. Under this assumption, the probability of satisfying a vehicle's capacity can be derived by replacing the unknown demand distribution with a normal distribution of the same mean and covariances if the risk tolerance ϵ is adjusted accordingly.

The present paper aims to contribute to a deeper understanding of both the structural properties and the solution of the distributionally robust chance constrained CVRP. We show that the rounded capacity inequalities (RCIs), a popular class of cutting planes for the deterministic CVRP, can be adapted to the distributionally robust chance constrained CVRP whenever the underlying ambiguity set satisfies a subadditivity property. While several classes of popular ambiguity sets, such as ϕ-divergence (Hu and Hong, 2013; Jiang and Guan, 2015) and Wasserstein (Esfahani and Kuhn, 2017; Zhao and Guan, 2018) ambiguity sets, violate this subadditivity property, the condition holds for a wide class of moment ambiguity sets (El Ghaoui et al., 2003; Delage and Ye, 2010). Motivated by this insight, we study marginal moment ambiguity sets, which characterize each customer demand individually, and generic moment ambiguity sets, which also describe the interactions between different customer demands. We find that the distributionally robust chance constrained CVRP over a marginal moment ambiguity set reduces to a deterministic CVRP with altered customer demands. The same problem over a generic moment ambiguity set, on the other hand, does not have an equivalent reformulation as a deterministic CVRP in general. We present RCI separation procedures for two classes of generic moment ambiguity sets. Our numerical experiments indicate that contrary to the deterministic CVRP, which appears to be best solved with branch-and-cut-and-price schemes, branch-and-cut algorithms may be competitive for the distributionally robust chance constrained CVRP.

More succinctly, the contributions of this paper may be summarized as follows.

1. We show that whether or not the distributionally robust chance constrained CVRP can be solved with a standard branch-and-cut scheme depends on the presence or absence of a subadditivity property in the employed ambiguity set. We prove that this subadditivity property is present in a wide class of moment ambiguity sets.

2. We show that for marginal moment ambiguity sets, the distributionally robust chance constrained CVRP reduces to a deterministic CVRP with altered customer demands. We derive these demands for various classes of marginal moment ambiguity sets, and we describe the associated worst-case demand distributions in closed form.

3. We develop cut separation schemes for different classes of generic moment ambiguity sets, and we show that the associated worst-case distributions can be determined *a posteriori* through the solution of tractable optimization problems.

The intimate connection between the applicability of branch-and-cut schemes and the subadditivity of ambiguity sets appears to have implications well beyond the CVRP, and we believe that this relationship deserves further study in the wider context of distributionally robust optimization.

The remainder of the paper is organized as follows. Section 2 introduces and motivates the distributionally robust chance constrained CVRP. Section 3 shows that the problem can be solved with branch-and-cut schemes whenever the ambiguity set satisfies a subadditivity condition, and that this subadditivity condition holds for a wide class of moment ambiguity sets. Sections 4 and 5 study the properties of marginal and generic moment ambiguity sets, respectively. We present our numerical results in Section 6, and we offer concluding remarks in Section 7. For ease of exposition, all proofs are relegated to the appendix. The source code of the proposed branch-and-cut algorithm is available as part of the paper's online supplement.

Notation. We denote scalars and vectors by regular and bold lowercase letters, whereas bold uppercase letters are reserved for matrices. The vectors $\mathbf{e}$ and $\mathbf{0}$ refer to the vectors of all ones and all zeros, respectively, while $\mathbf{e}_i$ is the i-th basis vector. For a set $A \subseteq \{1, \ldots, n\}$, the vector $\mathbf{1}_A \in \{0,1\}^n$ satisfies $(\mathbf{1}_A)_i = 1$ if and only if $i \in A$. We define the conjugate of a real-valued function $f : \mathbb{R}^n \mapsto \mathbb{R}$ by $f^\star(\boldsymbol{y}) = \sup \{\boldsymbol{y}^\top \boldsymbol{x} - f(\boldsymbol{x}) \,:\, \boldsymbol{x} \in \mathbb{R}^n\}$.

2 Capacitated Vehicle Routing under Uncertainty

We consider a complete and directed graph $G = (V, A)$ with nodes $V = \{0, \ldots, n\}$ and arcs $A = \{(i,j) \in V \times V \,:\, i \neq j\}$. The node $0 \in V$ represents the unique depot, and the nodes $V_C = \{1, \ldots, n\} \subset V$ denote the customers. The depot is equipped with m homogeneous vehicles of capacity $Q \in \mathbb{R}_+$, which we index through the set $K = \{1, \ldots, m\}$. Each vehicle incurs transportation costs of $c(i,j) \in \mathbb{R}_+$ if it traverses the arc $(i,j) \in A$. Throughout the paper, we allow for asymmetric transportation costs. As is standard in the literature, our models simplify if the transportation costs satisfy $c(i,j) = c(j,i)$ for all $(i,j) \in A$; see, *e.g.*, Semet et al. (2014).

We denote by $\mathfrak{P}(V_C, m)$ the set of all (ordered) partitions of the customer set V_C into m mutually disjoint and collectively exhaustive (ordered) routes $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m$:

$$\mathfrak{P}(V_C, m) = \left\{ (\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m) \,:\, \boldsymbol{R}_k \neq \emptyset \;\; \forall k, \;\; \boldsymbol{R}_k \cap \boldsymbol{R}_l = \emptyset \;\; \forall k \neq l, \;\; \bigcup_k \boldsymbol{R}_k = V_C \right\}$$

In this definition, each route $\boldsymbol{R}_k = (R_{k,1}, \ldots, R_{k,n_k})$ is an ordered list, where $R_{k,l} \in V_C$ is the l-th customer and n_k the total number of customers visited by vehicle $k \in K$. With a slight abuse of notation, we apply set operations to routes whenever the interpretation is clear. We also refer to the collection of routes $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m$ as the route set $\mathbf{R}$.

The *deterministic CVRP* asks for a route set $\mathbf{R} \in \mathfrak{P}(V_C, m)$ that minimizes the overall transportation costs $c(\mathbf{R}) = \sum_{k \in K} \sum_{l=0}^{n_k} c(R_{k,l}, R_{k,l+1})$ while satisfying all vehicle capacities:

$$\begin{array}{ll} \text{minimize} & c(\mathbf{R}) \\ \text{subject to} & \boldsymbol{R}_k \in \mathcal{R}(\boldsymbol{q}) \qquad \forall k \in K \\ & \mathbf{R} \in \mathfrak{P}(V_C, m) \end{array}$$

Here, we set $R_{k,0} = R_{k,n_k+1} = 0$ for all $k \in K$ so that each route starts and ends at the depot, and we assume that each customer $i \in V_C$ has a known demand $q_i \in \mathbb{R}_+$. The shorthand notation $\boldsymbol{R}_k \in \mathcal{R}(\boldsymbol{q})$ expresses the capacity constraint for the k-th vehicle, that is, $\sum_{i \in \boldsymbol{R}_k} q_i \leq Q$.

The *robust CVRP* seeks for a route set that satisfies the vehicle capacities for all anticipated demand realizations within an uncertainty set $\mathcal{Q}$. Thus, the formulation of the robust CVRP replaces the deterministic capacity constraints $\boldsymbol{R}_k \in \mathcal{R}(\boldsymbol{q})$ with the robust capacity constraints $\boldsymbol{R}_k \in \bigcap_{\boldsymbol{q} \in \mathcal{Q}} \mathcal{R}(\boldsymbol{q})$, $k \in K$. The robust CVRP reduces to a deterministic CVRP if $\mathcal{Q} = \{\boldsymbol{q}\}$.

The *chance constrained CVRP* models the customer demands as a random vector $\tilde{\boldsymbol{q}}$ governed by a known probability distribution $\mathbb{Q}$. The objective is to find a route set that satisfies all vehicle capacities with high probability. Thus, we replace the capacity constraints $\boldsymbol{R}_k \in \mathcal{R}(\boldsymbol{q})$ of the deterministic CVRP with the probabilistic capacity constraints $\mathbb{Q}\left[\boldsymbol{R}_k \in \mathcal{R}(\tilde{\boldsymbol{q}})\right] \geq 1 - \epsilon$, where $\epsilon \in (0, 1)$ represents a prescribed tolerance for capacity violations. Note that the chance constrained CVRP reduces to a deterministic CVRP if $\mathbb{P} = \delta_{\boldsymbol{q}}$, where $\delta_{\boldsymbol{q}}$ denotes the Dirac distribution that places unit mass on the demand realization $\tilde{\boldsymbol{q}} = \boldsymbol{q}$.

Although modeling the customer demands as a random vector that is governed by a known distribution is intuitively appealing, the practicability of the chance constrained CVRP is challenged in three ways: *(i)* most of the solution schemes for chance constrained CVRPs require the customer demands to be independent; *(ii)* merely establishing the (in-)feasibility of a fixed route plan can already be challenging from a computational perspective; and *(iii)* estimating the customer demand distribution from historical records may require unrealistically large amounts of data. We elaborate on these shortcomings in Section EC.1 of the electronic companion.

In this paper, we study the *distributionally robust chance constrained CVRP*:

$$\begin{aligned} \text{minimize} \quad & c(\mathbf{R}) \\ \text{subject to} \quad & \mathbb{P}\left[\boldsymbol{R}_k \in \mathcal{R}(\tilde{\boldsymbol{q}})\right] \geq 1 - \epsilon \qquad \forall \mathbb{P} \in \mathcal{P},\ \forall k \in K \\ & \mathbf{R} \in \mathfrak{P}(V_C, m) \end{aligned} \tag{RVRP($\mathcal{P}$)}$$

In this problem, the ambiguity set $\mathcal{P}$ contains all distributions that are deemed plausible for governing the random demand vector $\tilde{\boldsymbol{q}}$. In particular, if the unknown true distribution $\mathbb{Q}$ is contained in $\mathcal{P}$, then any feasible solution to RVRP($\mathcal{P}$) is guaranteed to satisfy each vehicle's capacity constraint with a probability of at least $1 - \epsilon$ under $\mathbb{Q}$, that is, the corresponding route set is feasible in the chance constrained CVRP with the unknown true distribution $\mathbb{Q}$. Note that RVRP($\mathcal{P}$) reduces to a deterministic CVRP if $\mathcal{P} = \{\delta_{\boldsymbol{q}}\}$, to a robust CVRP if $\mathcal{P} = \{\delta_{\boldsymbol{q}} : \boldsymbol{q} \in \mathcal{Q}\}$ and to a chance constrained CVRP if $\mathcal{P} = \{\mathbb{Q}\}$. In the remainder of the paper, we use the terms 'distributionally robust chance constrained CVRP', 'distributionally robust CVRP' and 'RVRP($\mathcal{P}$)' interchangeably.

As we will see in the following, the distributionally robust CVRP simultaneously addresses all three of the aforementioned challenges: *(i)* it caters for dependent customer demands through ambiguity sets that contain both independent and dependent demand distributions; *(ii)* for large classes of ambiguity sets, the (in-)feasibility of a fixed route set can be established in polynomial time; and *(iii)* since an ambiguity set only characterizes certain properties of the unknown true distribution $\mathbb{Q}$, its estimation requires less data and can often be done using historical records.

At this stage it is worth pointing out the potential *shortcomings* of the distributionally robust CVRP. Firstly, the tractability of RVRP($\mathcal{P}$) crucially depends on the shape of the ambiguity set $\mathcal{P}$. As we will see in the remainder of the paper, some intuitively appealing ambiguity sets lead to tractable reformulations, whereas others do not. Secondly, since the ambiguity set $\mathcal{P}$ only characterizes certain properties of the unknown true distribution $\mathbb{Q}$, it may contain other distributions that are unlikely to govern the customer demands $\tilde{\boldsymbol{q}}$ but that still need to be considered in the vehicles' capacity constraints in RVRP($\mathcal{P}$). Finally, and closely related, the worst-case distribution $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\big[\boldsymbol{R}_k \in \mathcal{R}(\tilde{\boldsymbol{q}})\big]$ minimizing the probability of the k-th vehicle's capacity constraint being satisfied is typically a pathological distribution that is unlikely to be encountered in practice. In fact, we will see that for the classes of ambiguity sets considered in this paper, one can construct worst-case distributions that are supported on two demand realizations only. The aforementioned shortcomings are intrinsic to the distributionally robust optimization methodology and are not specific to the distributionally robust CVRP. We emphasize that despite these weaknesses, distributionally robust optimization has been successfully applied in many diverse application areas, ranging from finance (Goldfarb and Iyengar, 2003) and energy systems (Zhao and Jiang, 2018) to communication networks (Li et al., 2014) and healthcare (Meng et al., 2015). We therefore believe that the distributionally robust CVRP serves as a complement to the existing modeling paradigms for the CVRP under uncertainty, such as the robust CVRP and the chance constrained CVRP. In particular, the most appropriate formulation for a specific application may depend on a variety of factors, such as runtime and scalability requirements, the acceptable degree of conservatism and the availability of historical records, and it ultimately needs to be decided upon by the domain expert.

Remark 1 (Joint Chance Constrained CVRP)**.** *Following the conventions of the vehicle routing literature, we consider* individual *chance constraints. Instead, one could consider a* joint *chance constraint, where the individual capacity requirements* $\mathbb{P}\left[\boldsymbol{R}_k \in \mathcal{R}(\tilde{\boldsymbol{q}})\right] \geq 1-\epsilon$, $k \in K$, *are replaced* *with a single joint capacity requirement* $\mathbb{P}\left[\boldsymbol{R}_k \in \mathcal{R}(\tilde{\boldsymbol{q}}) \;\; \forall k \in K\right] \geq 1 - \epsilon$. *The individual chance constraints provide a guarantee for each individual route (and hence, for every customer along that route), whereas the joint chance constraint offers a guarantee for the entire route plan. Since joint chance constrained optimization problems are typically much more challenging from a computational perspective (see,* e.g.*, Hanasusanto et al. 2017 and Xie and Ahmed 2017), we will focus on the individually chance constrained CVRP throughout this paper.*

3 Distributionally Robust Two-Index Vehicle Flow Formulation

The distributionally robust CVRP enforces chance constraints for each route $\boldsymbol{R}_k$ with respect to all probability distributions $\mathbb{P} \in \mathcal{P}$, of which there could be uncountably many. It is therefore not *a priori* clear how RVRP($\mathcal{P}$) can be solved numerically. In the following, we show that under certain conditions, RVRP($\mathcal{P}$) is equivalent to a two-index vehicle flow (2VF) formulation of the form

$$
\begin{array}{lll}
\text{minimize} & \displaystyle\sum_{(i,j) \in A} c(i,j)\, x_{ij} & \\
\text{subject to} & \displaystyle\sum_{\substack{j \in V: \\ (i,j) \in A}} x_{ij} = \sum_{\substack{j \in V: \\ (j,i) \in A}} x_{ji} = \delta_i & \forall i \in V \\
& \displaystyle\sum_{i \in V \setminus S} \sum_{j \in S} x_{ij} \geq d_{\mathcal{P}}(S) & \forall S \subseteq V_C, \; S \neq \emptyset \\
& x_{ij} \in \{0, 1\} & \forall (i,j) \in A,
\end{array}
\qquad (2\text{VF}(\mathcal{P}))
$$

where $\delta_i = 1$ for $i \in V_C$ and $\delta_0 = m$, and the demand estimator $d_{\mathcal{P}} : 2^{V_C} \mapsto \mathbb{R}_+$ maps subsets of the customer set V_C to the non-negative real line. In this formulation, we have $x_{ij} = 1$ if and only if one of the m vehicles traverses the arc $(i,j) \in A$. The objective function minimizes the overall transportation costs across all vehicles. The first constraint set ensures that each customer is visited by exactly one vehicle, and that m vehicles leave and return to the depot. The second constraint set is commonly referred to as rounded capacity inequalities (RCIs), and they ensure that the vehicles' capacity constraints are met and that every route contains the depot node.

For a fixed set S of customers, the left-hand side of the associated RCI represents an upper bound on the number of vehicles entering S (since some vehicles may enter S several times). Thus, the demand estimator $d_{\mathcal{P}}(S)$ on the right-hand side of the RCI has to provide a (sufficiently tight) lower bound on the number of vehicles required to serve the customers in S. Since there are exponentially many RCIs, they are typically introduced iteratively as part of a branch-and-cut scheme. 2VF($\mathcal{P}$) is one of the most well-studied formulations for the CVRP, and a large number of branch-and-cut schemes have been designed for its solution (see, *e.g.*, Lysgaard et al. 2004 and Semet et al. 2014). Thus, if we can show that RVRP($\mathcal{P}$) is equivalent to 2VF($\mathcal{P}$) for some demand estimator $d_{\mathcal{P}}$, then we can solve RVRP($\mathcal{P}$) as long as we can evaluate $d_{\mathcal{P}}$ quickly.

For the deterministic CVRP, a popular choice for the demand estimator is $\left\lceil \frac{1}{Q} \sum_{i \in S} q_i \right\rceil$, which represents the minimum number of vehicles required to serve S if the deliveries could be split continuously across vehicles, rounded up to the next integer number. It has been shown that this lower bound is sufficiently tight to ensure that the capacity constraint of each vehicle is met by any feasible solution to the corresponding 2VF formulation (Laporte et al., 1985). Moreover, this demand estimator eliminates short cycles that do not contain the depot node as long as the customer demands satisfy $\boldsymbol{q} > \boldsymbol{0}$ component-wise. Although tighter RCIs could in principle be obtained through the solution of bin packing problems, the increased strength of the cuts typically does not justify the additional computational effort required to evaluate the demand estimator.

To quantify the number of vehicles required to serve a customer set S in the distributionally robust CVRP, we define the value-at-risk of a random variable $\tilde{X}$ governed by the distribution $\mathbb{Q}$ as

$$\mathbb{Q}\text{-VaR}_{1-\epsilon}\big[\tilde{X}\big] = \inf\Big\{x \in \mathbb{R} \,:\, \mathbb{Q}\big[\tilde{X} \leq x\big] \geq 1 - \epsilon\Big\},$$

which denotes the $(1-\epsilon)$-quantile of $\tilde{X}$. Indeed, we have that

$$\mathbb{Q}\big[\tilde{X} \leq \tau\big] \geq 1 - \epsilon \quad \Longleftrightarrow \quad \mathbb{Q}\text{-VaR}_{1-\epsilon}\big[\tilde{X}\big] \leq \tau,$$

which in the case of the CVRP translates to

$$\mathbb{Q}\left[\boldsymbol{R}_k \in \mathcal{R}(\tilde{\boldsymbol{q}})\right] \geq 1 - \epsilon \quad \Longleftrightarrow \quad \mathbb{Q}\Big[\sum_{i \in \boldsymbol{R}_k} \tilde{q}_i \leq Q\Big] \geq 1 - \epsilon \quad \Longleftrightarrow \quad \mathbb{Q}\text{-VaR}_{1-\epsilon}\Big[\sum_{i \in \boldsymbol{R}_k} \tilde{q}_i\Big] \leq Q.$$

Instead of considering a single probability distribution $\mathbb{Q}$, however, RVRP($\mathcal{P}$) enforces chance constraints for *all* probability distributions $\mathbb{P} \in \mathcal{P}$. A similar reasoning as before shows that

$$\mathbb{P}\big[\tilde{X} \leq \tau\big] \geq 1 - \epsilon \;\; \forall \mathbb{P} \in \mathcal{P} \quad \Longleftrightarrow \quad \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{X}\big] \leq \tau \;\; \forall \mathbb{P} \in \mathcal{P} \quad \Longleftrightarrow \quad \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{X}\big] \leq \tau,$$

or, in the context of our distributionally robust CVRP,

$$\mathbb{P}\left[\boldsymbol{R}_k \in \mathcal{R}(\tilde{\boldsymbol{q}})\right] \geq 1-\epsilon \;\; \forall \mathbb{P} \in \mathcal{P} \quad \Longleftrightarrow \quad \mathbb{P}\Big[\sum_{i \in \boldsymbol{R}_k} \tilde{q}_i \leq Q\Big] \geq 1-\epsilon \;\; \forall \mathbb{P} \in \mathcal{P}$$
$$\Longleftrightarrow \quad \sup_{\mathbb{P} \in \mathcal{P}} \; \mathbb{P}\text{-VaR}_{1-\epsilon}\Big[\sum_{i \in \boldsymbol{R}_k} \tilde{q}_i\Big] \leq Q. \qquad (1)$$

In view of the above equivalences and inspired by the RCIs for the deterministic CVRP, we are led to the following demand estimator for the distributionally robust CVRP:

$$d_{\mathcal{P}}(S) = \max\left\{ \left\lceil \frac{1}{Q} \sup_{\mathbb{P} \in \mathcal{P}} \; \mathbb{P}\text{-VaR}_{1-\epsilon}\Big[\sum_{i \in S} \tilde{q}_i\Big] \right\rceil, \, 1 \right\} \qquad \forall S \neq \emptyset, \qquad (2)$$

as well as $d_{\mathcal{P}}(\emptyset) = 0$. In this expression, the supremum corresponds to the worst-case $(1-\epsilon)$-quantile of the cumulative customer demands in S (also called $(1-\epsilon)$-worst-case value-at-risk), and the division of this term by Q is supposed to provide a lower bound on the number of vehicles required to serve the customers in S. We take the maximum between this quantity (rounded up to the next integer) and 1 to ensure the elimination of short cycles. Indeed, contrary to the cumulative customer demands in the deterministic CVRP, the worst-case $(1-\epsilon)$-quantile could be zero even if no individual customer demand is deterministically zero. Similar to the deterministic RCIs, our demand estimator could in principle be tightened through the solution of a distributionally robust chance constrained bin packing problem. As in the deterministic case, however, this would usually not be attractive from a computational perspective.

One could expect RVRP($\mathcal{P}$) and 2VF($\mathcal{P}$) to be equivalent under any ambiguity set $\mathcal{P}$ as long as the demand estimator $d_{\mathcal{P}}$ is chosen as in (2). Unfortunately, this is *not* the case.

Example 1. *Consider a distributionally robust CVRP instance with $n = 2$ customers and $m = 2$ vehicles of capacity $Q = 1$. We define the ambiguity set for the customer demands as*

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^2) \; : \; \begin{bmatrix} \mathbb{P}(\tilde{q}_1 = 1) = 0.925, & \mathbb{P}(\tilde{q}_1 = 2) = 0.075 \\ \mathbb{P}(\tilde{q}_2 = 1) = 0.925, & \mathbb{P}(\tilde{q}_2 = 2) = 0.075 \end{bmatrix} \right\},$$

that is, each customer has a demand of 1 *(*2*) with probability* 0.925 *(*0.075*). Note that the ambiguity set does* not *specify that the customer demands are independent.*

For $\epsilon = 0.1$, the route set $\mathbf{R} = (\boldsymbol{R}_1, \boldsymbol{R}_2)$ with $\boldsymbol{R}_1 = (1)$ and $\boldsymbol{R}_2 = (2)$ is feasible in RVRP($\mathcal{P}$) since $\mathbb{P}\big[\tilde{q}_i \leq 1\big] = 0.925 \geq 1 - \epsilon = 0.9$ for $i = 1, 2$ and all $\mathbb{P} \in \mathcal{P}$. However, this route set $\mathbf{R}$ is

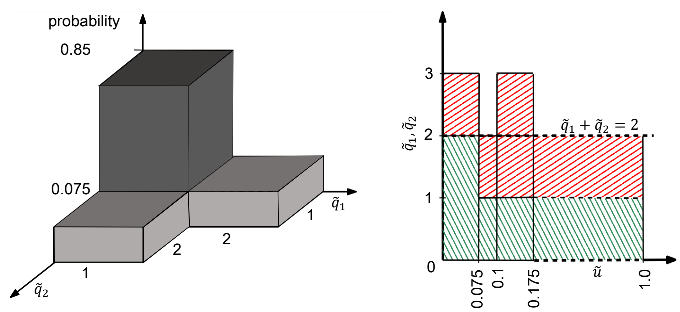

Figure 1. Probability distribution $\mathbb{P}^\star$ which illustrates that RVRP($\mathcal{P}$) and 2VF($\mathcal{P}$) are not equivalent. The left graph shows the probability distribution itself, whereas the right graph visualizes the customer demands $\tilde{q}_1$ and $\tilde{q}_2$ as a function of the underlying random variable $\tilde{u}$ used in the construction of $\mathbb{P}^\star$.

infeasible in 2VF($\mathcal{P}$) since it violates the RCI constraint for $S = \{1, 2\}$. Indeed, we have that

$$\begin{aligned} d_{\mathcal{P}}(\{1,2\}) &= \max\left\{ \left\lceil \frac{1}{Q} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\left[\tilde{q}_1 + \tilde{q}_2\right] \right\rceil, 1 \right\} \\ &\geq \mathbb{P}^\star\text{-VaR}_{1-\epsilon}\left[\tilde{q}_1 + \tilde{q}_2\right] = 3 \end{aligned}$$

since the probability distribution $\mathbb{P}^\star$ with the dependence structure

$$\tilde{q}_1 = \begin{cases} 2 & \text{if } \tilde{u} \in [0, 0.075], \\ 1 & \text{otherwise,} \end{cases} \qquad \tilde{q}_2 = \begin{cases} 2 & \text{if } \tilde{u} \in [0.1, 0.175], \\ 1 & \text{otherwise,} \end{cases}$$

where $\tilde{u}$ is a uniformly distributed random variable supported on $[0, 1]$, is contained in $\mathcal{P}$ (see Figure 1). We thus conclude that RVRP($\mathcal{P}$) and 2VF($\mathcal{P}$) are not *equivalent for this instance.*

Intuitively, the equivalence between RVRP($\mathcal{P}$) and 2VF($\mathcal{P}$) fails to hold in Example 1 due to the combination of two differences between the formulations. Firstly, RVRP($\mathcal{P}$) ignores the amount by which a capacity restriction is violated, whereas this amount is considered in the demand estimator (2) of 2VF($\mathcal{P}$). In particular, whenever the cumulative demands within a single vehicle exceed that vehicle's capacity in Example 1, then the cumulative demands are so large that they could not be served by both vehicles in 2VF($\mathcal{P}$) either, even if the demands could be split continuously. Secondly, since the vehicles' capacity restrictions in Example 1 are violated in non-overlapping scenarios, the probability of exceeding *some* vehicle's capacity is equal to the sum of probabilities of exceeding each individual vehicle's capacity. More generally, RVRP($\mathcal{P}$) only considers the probabilities of violating each individual vehicle's capacity, whereas the demand estimator (2) of 2VF($\mathcal{P}$) considers the joint violation probability (under the assumption that customer demands can be split continuously).

The aforementioned differences between RVRP($\mathcal{P}$) and 2VF($\mathcal{P}$) relate to the fact that the RCIs are agnostic to the assignment of customers to vehicles, and as such they consider the interplay between demands allocated to different vehicles even though such dependencies should be ignored. To avoid this problem, the demand estimator $d_{\mathcal{P}}$ should not assign 'excessively large' numbers of vehicles $d_{\mathcal{P}}(S)$ to large customer subsets S. It turns out that this intuition can be formalized.

(**S**) **Subadditivity.** For all customer subsets $S, T \subseteq V_C$, we have $d_{\mathcal{P}}(S \cup T) \leq d_{\mathcal{P}}(S) + d_{\mathcal{P}}(T)$.

Indeed, the demand estimator in Example 1 violates the subadditivity condition (**S**).

Example 1 (cont'd)**.** *For the distributionally robust CVRP instance from Example 1, we have*

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{0.9}(\tilde{q}_1) \;=\; \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{0.9}(\tilde{q}_2) \;=\; 1,$$

but at the same time we have

$$\begin{aligned} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{0.9}(\tilde{q}_1 + \tilde{q}_2) \;&\geq\; \mathbb{P}^{\star}\text{-VaR}_{0.9}(\tilde{q}_1 + \tilde{q}_2) \;=\; 3 \\ &>\; \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{0.9}(\tilde{q}_1) \;+\; \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{0.9}(\tilde{q}_2) \;=\; 2. \end{aligned}$$

*In other words, the demand estimator $d_{\mathcal{P}}$ violates the subadditivity condition (**S**) since*

$$d_{\mathcal{P}}(\{1\} \cup \{2\}) \;\not\leq\; d_{\mathcal{P}}(\{1\}) \;+\; d_{\mathcal{P}}(\{2\}).$$

We now show that the condition (**S**) is sufficient for RVRP($\mathcal{P}$) and 2VF($\mathcal{P}$) to be equivalent.

Theorem 1. *Assume that $\tilde{\boldsymbol{q}} \geq \mathbf{0}$ $\mathbb{P}$-a.s. for all $\mathbb{P} \in \mathcal{P}$ and that $d_{\mathcal{P}}$ satisfies the subadditivity condition (**S**). Then the problems* RVRP($\mathcal{P}$) *and* 2VF($\mathcal{P}$) *are equivalent in the following sense:*

(i) Any route set $\mathbf{R}$ that is feasible in RVRP($\mathcal{P}$) *induces a unique solution $\boldsymbol{x}$ that is feasible in* 2VF($\mathcal{P}$) *via*

$$x_{ij} = 1 \iff \exists k \in K,\ \exists l \in \{0, \ldots, n_k\} \,:\, (i,j) = (R_{k,l}, R_{k,l+1}), \tag{3}$$

and $\boldsymbol{x}$ and $\mathbf{R}$ attain the same transportation costs.

(ii) Any solution $\boldsymbol{x}$ that is feasible in 2VF($\mathcal{P}$) *induces a route set* $\mathbf{R}$ *that is feasible in* RVRP($\mathcal{P}$) *via* (3)*, and this route set is unique up to a reordering of the individual routes* $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m$*. Moreover,* $\boldsymbol{x}$ *and* $\mathbf{R}$ *attain the same transportation costs.*

The proof of Theorem 1, together with all other proofs, can be found in the electronic companion of the paper. According to the theorem, any ambiguity set $\mathcal{P}$ whose demand estimator $d_{\mathcal{P}}$ satisfies the subadditivity condition (**S**) allows us to use a branch-and-cut algorithm to solve 2VF($\mathcal{P}$) *in lieu* of RVRP($\mathcal{P}$). Example 1 has shown that the subadditivity condition may be violated if the ambiguity set $\mathcal{P}$ specifies the marginal distribution of each customer's demand. The example immediately implies that hypothesis test ambiguity sets (Bertsimas et al., 2018), which converge to ambiguity sets that exactly specify the marginal distribution of each customer's demand as the available data increases, also give rise to demand estimators that violate the subadditivity condition. Moreover, data-driven ambiguity sets, such as ϕ-divergence ambiguity sets (Ben-Tal et al., 2013), Wasserstein ambiguity sets (Esfahani and Kuhn, 2017) and hypothesis test ambiguity sets (Bertsimas et al., 2018), converge to singleton ambiguity sets as the available data increases, and the resulting demand estimators also violate the subadditivity condition since the involved worst-case values-at-risk converge to values-at-risk which are known to violate subadditivity.

In this paper, we study moment ambiguity sets of the form

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) \,:\, \mathbb{P}(\tilde{\boldsymbol{q}} \in \mathcal{Q}) = 1, \;\; \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{q}}] = \boldsymbol{\mu}, \;\; \mathbb{E}_{\mathbb{P}}[\boldsymbol{\varphi}(\tilde{\boldsymbol{q}})] \leq \boldsymbol{\sigma} \right\}. \tag{4}$$

The moment ambiguity set (4) specifies that the uncertain customer demands $\tilde{\boldsymbol{q}}$ are supported on a rectangular set $\mathcal{Q} = [\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}]$ with $\underline{\boldsymbol{q}} \geq \mathbf{0}$. It also stipulates that the expected customer demands $\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{q}}]$ are known to be $\boldsymbol{\mu}$, and that the upper bounds σ_i on the demand variations $\mathbb{E}_{\mathbb{P}}[\varphi_i(\tilde{\boldsymbol{q}})]$, $i = 1, \ldots, p$, of the customer demands are known. The demand variations are characterized by a dispersion measure $\boldsymbol{\varphi} : \mathbb{R}^n \mapsto \mathbb{R}^p$ which measures how 'stretched out' the joint probability distribution of the customer demands is. Possible choices of dispersion measures include the mean absolute deviations, $\varphi_i(\boldsymbol{q}) = |q_i - \mu_i|$, the variances $\varphi_i(\boldsymbol{q}) = (q_i - \mu_i)^2$, higher order moments $\varphi_i(\boldsymbol{q}) = |q_i - \mu_i|^q$, $q \geq 3$, or Huber loss functions of the customer demands $\tilde{\boldsymbol{q}}$. We will explore different dispersion measures in Sections 4 and 5. Throughout this paper, we make the standard regularity assumptions that $\boldsymbol{\mu} \in \operatorname{int} \mathcal{Q}$, that is, the expected demands are contained in the interior of the support $\mathcal{Q}$, that the dispersion measure $\boldsymbol{\varphi}$ is closed and component-wise convex, and that $\boldsymbol{\varphi}(\boldsymbol{\mu}) < \boldsymbol{\sigma}$. These assumptions will allow us to invoke strong convex duality, which is required for our results to hold. Moment ambiguity sets are amongst the most popular ambiguity sets studied in the distributionally robust optimization literature, see, *e.g.*, El Ghaoui et al. (2003), Delage and Ye (2010), Zymler et al. (2013) and Wiesemann et al. (2014).

We now show that in contrast to ambiguity sets constructed by marginal histograms, hypothesis tests or deviation measures such as the Wasserstein distance and ϕ-divergences, moment ambiguity sets lead to demand estimators $d_{\mathcal{P}}$ that satisfy the desired subadditivity property.

Theorem 2. *The demand estimator $d_{\mathcal{P}}$ for moment ambiguity sets of the form* (4) *is subadditive.*

In addition to satisfying the subadditivity condition (**S**), the distributions that minimize the probability of satisfying a vehicle's capacity requirement have a particularly simple structure if we restrict ourselves to moment ambiguity sets of the form (4).

Proposition 1. *Consider an instance of the moment ambiguity set* (4)*. Then for any customer subset $S \subseteq V_C$, there is a sequence of two-point distributions $\mathbb{P}^t = p_1^t \cdot \delta_{\boldsymbol{q}_1^t} + p_2^t \cdot \delta_{\boldsymbol{q}_2^t} \in \mathcal{P}$, $p_1^t, p_2^t \in \mathbb{R}_+$ and $\boldsymbol{q}_1^t, \boldsymbol{q}_2^t \in \mathcal{Q}$, such that $\mathbb{P}^t\text{-VaR}_{1-\epsilon}\big[\sum_{i \in S} \tilde{q}_i\big] \longrightarrow \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\sum_{i \in S} \tilde{q}_i\big]$ as $t \longrightarrow \infty$.*

Proposition 1 shows that for moment ambiguity sets of the form (4), the worst-case value-at-risk $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\sum_{i \in S} \tilde{q}_i\big]$ is asymptotically attained by a series of probability distributions that place all probability mass on two demand scenarios. We emphasize that the two-point nature of the worst-case distribution does *not* depend on the number of moment constraints contained in the ambiguity set (4). In that sense, Proposition 1 strengthens the findings of the Richter-Rogosinski theorem (Shapiro et al., 2014, Theorem 7.37), which applies to more general risk measures, to the special case of the worst-case value-at-risk.

Proposition 1 confirms our intuition that the distributionally robust CVRP constitutes a compromise between the deterministic CVRP, which optimizes in view of a single expected (or most likely) demand scenario, and the robust CVRP, which optimizes in view of the worst demand scenario contained in an uncertainty set. At the same time, the distributionally robust CVRP also offers a trade-off between the classical chance constrained CVRP, which is often challenging to solve as it optimizes in view of a distribution that may place positive probability mass on many demand scenarios, and the robust CVRP, which optimizes in view of a single worst-case scenario.

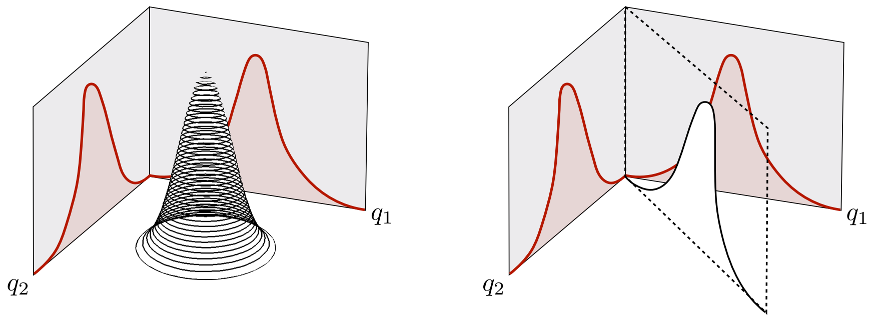

Figure 2. Examples of probability distributions contained in a marginalized ambiguity set. Since the conditions in the ambiguity set (5) only restrict the shapes of the marginal distributions, the ambiguity set contains distributions of varying dependence structure, ranging from independent (left graph) to perfectly correlated (right graph) ones.

4 Marginalized Moment Ambiguity Sets

In this section we study marginalized moment ambiguity sets of the form

$$\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) \, : \, \mathbb{P}(\tilde{\boldsymbol{q}} \in \mathcal{Q}) = 1, \;\; \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{q}}] = \boldsymbol{\mu}, \;\; \mathbb{E}_{\mathbb{P}}[\boldsymbol{\varphi}_i(\tilde{q}_i)] \leq \boldsymbol{\sigma}_i \;\; \forall i \in V_C\}\,, \tag{5}$$

where $\mathcal{Q} = [\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}]$ with $\underline{\boldsymbol{q}} \geq \mathbf{0}$, $\boldsymbol{\mu} \in \operatorname{int} \mathcal{Q}$, and $\boldsymbol{\varphi}_i : \mathbb{R} \mapsto \mathbb{R}^{p_i}$ is closed as well as component-wise convex and satisfies $\boldsymbol{\varphi}_i(\mu_i) < \boldsymbol{\sigma}_i$ with $\boldsymbol{\sigma}_i \in \mathbb{R}^{p_i}$, $i \in V_C$. In contrast to the generic moment ambiguity set (4), a marginalized moment ambiguity set only specifies the variability of individual customer demands and does not characterize the interactions between different customer demands. Nevertheless, Figure 2 shows that the customer demands may still exhibit dependencies under the probability distributions contained in a marginalized moment ambiguity set.

Marginalized moment ambiguity sets of the form (5) constitute a subclass of the generic moment ambiguity sets (4), and thus the demand estimator $d_{\mathcal{P}}$ over the ambiguity set (5) is subadditive due to Theorem 2. In fact, a much stronger additivity property holds for the ambiguity set (5).

Theorem 3. *Every marginalized moment ambiguity set of the form* (5) *satisfies*

$$\sup_{\mathbb{P} \in \mathcal{P}} \; \mathbb{P}\text{-VaR}_{1-\epsilon}\Big[\sum_{i \in S} \tilde{q}_i\Big] \;\; = \; \sum_{i \in S} \sup_{\mathbb{P} \in \mathcal{P}} \; \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big] \qquad \forall S \subseteq V_C, \; S \neq \emptyset.$$

Theorem 3 shows that the worst-case value-at-risk for a marginalized moment ambiguity set is additive. Nevertheless, we provide an example in Section EC.2 of the electronic companion which illustrates that for individual probability distributions within such an ambiguity set, the corresponding values-at-risk typically *fail* to be additive.

An immediate consequence of Theorem 3 is the following.

Corollary 1. *The distributionally robust CVRP over a marginalized moment ambiguity set* (5) *is equivalent to the deterministic CVRP with customer demands* $q_i = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big]$, $i \in V_C$.

If only marginal moment information is available, then Corollary 1 implies that a distributionally robust CVRP can be solved with existing solution schemes for deterministic CVRPs, such as branch-and-cut (Lysgaard et al., 2004; Semet et al., 2014) or branch-and-cut-and-price (Fukasawa et al., 2006; Pecin et al., 2017) algorithms. In other words, we can 'robustify' a deterministic CVRP instance without modifying the employed solution scheme by replacing the deterministic demands q_i with the (deterministic) worst-case value-at-risks $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big]$ for all customers $i \in V_C$.

While very attractive from a computational perspective, Corollary 1 also points at several weaknesses of the marginalized moment ambiguity sets. Firstly, marginalized moment ambiguity sets fail to capture any potentially known dependencies between customer demands. As a result, under the worst-case distribution all customer demands will attain their worst values jointly with probability ϵ. This contradicts the common wisdom that extreme demands are not typically attained simultaneously across all customers. Secondly, when using marginalized moment ambiguity sets we are unable to obtain a structurally different feasible region than for a suitably modified deterministic problem instance. As we will see in Section 5, this is in stark contrast to generic moment ambiguity sets that may not correspond to any deterministic problem instance. Finally, under the marginalized moment ambiguity sets the worst-case demand distribution does not depend on the selected route set, as we would usually expect to be the case under the distributionally robust optimization framework. In other words, under the marginalized moment ambiguity sets the decision maker cannot benefit from knowing the true probability distribution, as long as this distribution could be any of the distributions within the ambiguity set.

If we remove the expectation and the dispersion constraint in the marginalized moment ambiguity set (5), then the distributionally robust CVRP reduces to a deterministic CVRP with component-wise worst-case customer demands $\boldsymbol{q} = \overline{\boldsymbol{q}}$. If we remove the support and the dispersion constraint, on the other hand, then the distributionally robust CVRP becomes infeasible since the distribution $\kappa \cdot \delta_{\boldsymbol{q}_1} + (1-\kappa) \cdot \delta_{\boldsymbol{q}_2}$ with $\kappa \in (\epsilon, 1)$, $\boldsymbol{q}_1 = 2Q \cdot \mathbf{e}$ and $\boldsymbol{q}_2 = (\boldsymbol{\mu} - 2\kappa Q \cdot \mathbf{e})/(1-\kappa)$ is contained in the ambiguity set and places probability mass $\kappa > \epsilon$ on the demand scenario $2Q \cdot \mathbf{e}$, resulting in a worst-case value-at-risk of $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\sum_{i \in S} \tilde{q}_i\big] = 2Q|S|$. In the next three subsections, we develop closed-form solutions for the worst-case value-at-risk under support and expectation constraints combined with first-order, variance and semi-variance dispersion measures.

4.1 First-Order Ambiguity Sets

We begin with first-order marginalized moment ambiguity sets of the form

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) \,:\, \mathbb{P}\left(\tilde{\boldsymbol{q}} \in \mathcal{Q}\right) = 1,\;\; \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{q}}\right] = \boldsymbol{\mu},\;\; \mathbb{E}_{\mathbb{P}}\left[|\tilde{\boldsymbol{q}} - \boldsymbol{\mu}|\right] \leq \boldsymbol{\sigma} \right\}, \tag{6}$$

where the absolute value operator $|\cdot|$ is applied component-wise. As before, we assume that $\mathcal{Q} = [\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}]$ with $\underline{\boldsymbol{q}} \geq \mathbf{0}$ and $\boldsymbol{\mu} \in \operatorname{int} \mathcal{Q}$, and we additionally stipulate that $\boldsymbol{\sigma} > \mathbf{0}$. Note that (6) is a special case of the marginalized moment ambiguity set (5) where $\varphi_i(q_i) = |q_i - \mu_i|$, $i \in V_C$. The dispersion constraint imposes an upper bound of σ_i on the mean absolute deviation $\mathbb{E}_{\mathbb{P}}\left[|\tilde{q}_i - \mu_i|\right]$ of customer i's demand, $i \in V_C$.

Similar to the standard deviation, the mean absolute deviation measures the dispersion of a random variable around its expected value. Compared to the standard deviation, however, the mean absolute deviation is less affected by outliers and deviations from the standard modeling assumptions (such as normality). Due to these properties, the mean absolute deviation is preferred in the robust statistics literature, see, *e.g.*, Casella and Berger (2002).

We now show that the worst-case value-at-risk of a customer's demand $\tilde{q}_i$ under the marginalized first-order moment ambiguity set (6) admits a closed-form solution.

Proposition 2. *Every marginalized first-order ambiguity set of the form* (6) *satisfies*

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big] \;=\; \mu_i + \min\left\{ \overline{q}_i - \mu_i,\; \frac{1-\epsilon}{\epsilon}(\mu_i - \underline{q}_i),\; \frac{1}{2\epsilon}\sigma_i \right\} \qquad \forall i \in V_C. \tag{7}$$

Proposition 2 confirms our intuition that the worst-case value-at-risk of the demand of customer $i \in V_C$ increases with the range $\overline{q}_i - \underline{q}_i$, the safety threshold $1 - \epsilon$ as well as the upper bound on the mean absolute deviation σ_i. In particular, if customer i's demand is unbounded, then the worst-case value-at-risk simplifies to $\mu_i + \frac{1}{2\epsilon}\sigma_i$, and if the dispersion bound in (6) is disregarded, then the worst-case value-at-risk becomes $\min\left\{\overline{q}_i,\, \frac{1}{\epsilon}\mu_i - \frac{1-\epsilon}{\epsilon}\underline{q}_i\right\}$.

It is tempting to conclude from Proposition 2 that the worst-case value-at-risk (7) is attained by the Dirac distribution that places all probability mass on the single demand realization $\boldsymbol{\mu}$ +

$\min\left\{\overline{\boldsymbol{q}}-\boldsymbol{\mu},\ \frac{1-\epsilon}{\epsilon}(\boldsymbol{\mu}-\underline{\boldsymbol{q}}),\ \frac{1}{2\epsilon}\boldsymbol{\sigma}\right\}$, where the minimum is applied component-wise. This distribution, however, is *not* contained in the ambiguity set as it violates the expected value constraint in (6). Nevertheless, one can construct sequences of two-point distributions that are contained in the ambiguity set and that attain the worst-case value-at-risk (7) asymptotically. We characterize this sequence of distributions in Section EC.3.1 of the electronic companion.

4.2 Variance Ambiguity Sets

We next consider marginalized variance ambiguity sets of the form

$$\mathcal{P}=\left\{\mathbb{P}\in\mathcal{P}_0(\mathbb{R}^n)\,:\,\mathbb{P}\left[\tilde{\boldsymbol{q}}\in\mathcal{Q}\right]=1,\ \ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{q}}\right]=\boldsymbol{\mu},\ \ \mathbb{E}_{\mathbb{P}}\left[(\tilde{q}_i-\mu_i)^2\right]\le\sigma_i\ \ \forall i\in V_C\right\},\tag{8}$$

where $\mathcal{Q}=[\underline{\boldsymbol{q}},\overline{\boldsymbol{q}}]$ with $\underline{\boldsymbol{q}}\geq\mathbf{0}$ and $\boldsymbol{\mu}\in\operatorname{int}\mathcal{Q}$ as well as $\boldsymbol{\sigma}>\mathbf{0}$.

Similar to the mean absolute deviation, the worst-case value-at-risk of a customer's demand $\tilde{q}_i$ under the marginalized variance ambiguity set (8) admits a closed-form solution.

Proposition 3. *Every marginalized variance ambiguity set of the form* (8) *satisfies*

$$\sup_{\mathbb{P}\in\mathcal{P}}\ \mathbb{P}\text{-VaR}_{1-\epsilon}\left[\tilde{q}_i\right]\ =\ \mu_i+\min\left\{\overline{q}_i-\mu_i,\ \frac{1-\epsilon}{\epsilon}(\mu_i-\underline{q}_i),\ \sqrt{\frac{1-\epsilon}{\epsilon}\sigma_i}\right\}\qquad\forall i\in V_C.\tag{9}$$

The worst-case value-at-risk in Proposition 3 differs from the one in Proposition 2 only in the last term of the minimum operator, which corresponds to the variance bound in (8). Similar to the previous subsection, the expression (9) can be used to derive the worst-case value-at-risk if the support constraints or the variance constraint in (8) is disregarded. We characterize a sequence of two-point distributions attaining the worst-case value-at-risk in Section EC.3.2 of the electronic companion.

4.3 Semi-Variance Ambiguity Sets

We finally consider marginalized semi-variance ambiguity sets of the form

$$\mathcal{P}=\left\{\mathbb{P}\in\mathcal{P}_0(\mathbb{R}^n)\,:\,\left[\begin{array}{l}\mathbb{P}\left[\tilde{\boldsymbol{q}}\in\mathcal{Q}\right]=1,\ \ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{q}}\right]=\boldsymbol{\mu},\\ \mathbb{E}_{\mathbb{P}}\left[[\tilde{q}_i-\mu_i]_+^2\right]\le\sigma_i^+,\ \ \mathbb{E}_{\mathbb{P}}\left[[\mu_i-\tilde{q}_i]_+^2\right]\le\sigma_i^-\ \ \forall i\in V_C\end{array}\right]\right\},\tag{10}$$

where $\mathcal{Q}=[\underline{\boldsymbol{q}},\overline{\boldsymbol{q}}]$ with $\underline{\boldsymbol{q}}\geq\mathbf{0}$ and $\boldsymbol{\mu}\in\operatorname{int}\mathcal{Q}$ as well as $\boldsymbol{\sigma}^+,\boldsymbol{\sigma}^->\mathbf{0}$.

As in the preceding two subsections, the worst-case value-at-risk of a customer's demand $\tilde{q}_i$ under the marginalized semi-variance ambiguity set (10) admits a closed-form solution.

Proposition 4. *Every marginalized semi-variance ambiguity set of the form* (10) *satisfies*

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big] \;=\; \mu_i + \min\left\{\overline{q}_i - \mu_i, \frac{1-\epsilon}{\epsilon}(\mu_i - \underline{q}_i), \sqrt{\frac{\sigma_i^+}{\epsilon}}, \frac{\sqrt{(1-\epsilon)\sigma_i^-}}{\epsilon}\right\} \qquad \forall i \in V_C. \quad (11)$$

The worst-case value-at-risk in Proposition 4 differs from the previous ones in the last two terms of the minimum operator, which correspond to the semi-variance bounds in (10). Again, the expression (11) can be used to derive the worst-case value-at-risk if the support constraints or either or both of the two semi-variance constraint in (10) are disregarded. We characterize a sequence of two-point distributions attaining the worst-case value-at-risk in Section EC.3.3 of the electronic companion.

5 Generic Moment Ambiguity Sets

We now study generic moment ambiguity sets of the form (4), where the dispersion measure $\boldsymbol{\varphi}$ characterizes the joint variability of multiple demands. In particular, we consider ambiguity sets that stipulate bounds on the mean absolute deviations (Section 5.1) and the covariances (Section 5.2).

5.1 First-Order Ambiguity Sets

We begin with first-order generic moment ambiguity sets of the form

$$\mathcal{P} = \left\{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) \,:\, \mathbb{P}\left[\tilde{\boldsymbol{q}} \in \mathcal{Q}\right] = 1,\;\; \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{q}}\right] = \boldsymbol{\mu},\;\; \mathbb{E}_{\mathbb{P}}\big[\mathbf{1}_{S_i}^\top |\tilde{\boldsymbol{q}} - \boldsymbol{\mu}|\big] \leq \nu_i \;\; \forall i = 1,\ldots,p\right\}, \qquad (12)$$

where $\mathcal{Q} = [\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}]$ with $\underline{\boldsymbol{q}} \geq \mathbf{0}$, $\boldsymbol{\mu} \in \operatorname{int}\mathcal{Q}$ and $\boldsymbol{\nu} > \mathbf{0}$. As in Section 4.1, the absolute value operator $|\cdot|$ is applied component-wise. Note that (12) is a special case of the generic moment ambiguity set (4) where $\varphi_i(\boldsymbol{q}) = \sum_{j\in S_i} |\tilde{q}_j - \mu_j|$, $i = 1,\ldots,p$. In particular, the demand estimator $d_{\mathcal{P}}$ over the ambiguity set (12) is subadditive due to Theorem 2.

As we pointed out in Section 4.1, the mean absolute deviation in (12) is a popular dispersion measure in robust statistics. It is reminiscent of the standard deviation in classical statistics, but it is less affected by outliers and deviations from the classical model assumptions (*e.g.*, normality), which makes it more robust if the distribution is estimated from historical data. It can be shown that the sample mean absolute deviation outperforms the standard deviation in terms of asymptotic relative efficiency if the sample distribution has fat tails or if it is contaminated with another distribution

(Casella and Berger, 2002). For the use of the mean absolute deviation in (distributionally) robust optimization, see Bandi and Bertsimas (2012), Wiesemann et al. (2014) and Postek et al. (2018).

The first-order generic moment ambiguity set (12) generalizes the first-order marginalized moment ambiguity set (6). The possibility to impose upper bounds on the mean absolute deviations of sums of customer demands allows to reduce the ambiguity whenever customer demands are not perfectly correlated. While one could in principle impose upper dispersion bounds on the cumulative demands of any customer subset $S_i \subseteq V_C$, this approach would require large amounts of data to estimate the corresponding dispersion bounds ν_i, and it would be computationally demanding to determine the associated RCI cuts. Instead, one may impose upper dispersion bounds on some 'canonical' customer subsets that are dictated by the application area, for example, all customers within a specific municipality, county or state. For a given set of demand observations, the dispersion bounds ν_i for different subsets $S_i \subseteq V_C$ can be derived analytically using asymptotic arguments (Pham-Gia and Hung, 2001; Segers, 2014) or empirically via bootstrapping (Chernick, 2007).

Contrary to the marginalized ambiguity sets studied in Section 4, distributionally robust CVRPs with ambiguity sets of the form (12) typically *cannot* be reformulated as deterministic CVRPs.

Theorem 4. *For some instances of the distributionally robust CVRP with ambiguity set* (12) *there is no deterministic CVRP instance with the same set of feasible route sets.*

Intuitively speaking, the non-existence of a deterministic reformulation is owed to the fact that the deterministic CVRP cannot capture dependencies between customer demands. The proof of Theorem 4 constructs a distributionally robust CVRP instance with four customers where the demands of the customers 1 and 3 (as well as 2 and 4) cannot vary much jointly, whereas the demands of the customers 1 and 2 (as well as 3 and 4) can vary much jointly. As a result, the customer subsets $\{1, 3\}$ and $\{2, 4\}$ can each be served by a single vehicle in the distributionally robust CVRP instance. In the deterministic CVRP, on the other hand, the potential presence of joint variability of customer demands implies that at least some of the demands have to be sufficiently high, which in turn excludes the possibility to serve all customers by two vehicles.

Although we are unable to evaluate the worst-case value-at-risk $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\sum_{i \in S} \tilde{q}_i\big]$ in closed form for the ambiguity set (12), the quantity can be computed in polynomial time.

Theorem 5. *The worst-case value-at-risk* $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\sum_{i \in S} \tilde{q}_i\big]$ *over the generic first-order* *ambiguity set* (12) *is equal to the optimal objective value of the following optimization problem.*

$$\begin{array}{ll} \text{minimize} & \mathbf{1}_S{}^\top \boldsymbol{\mu} + \min\left\{(\overline{\boldsymbol{q}} - \boldsymbol{\mu}),\ \dfrac{1-\epsilon}{\epsilon}(\boldsymbol{\mu} - \underline{\boldsymbol{q}})\right\}^\top \left[\mathbf{1}_S - 2\sum_{i=1}^{p} \gamma_i \mathbf{1}_{S_i}\right]_+ + \dfrac{1}{\epsilon}\boldsymbol{\nu}^\top \boldsymbol{\gamma} \\ \text{subject to} & \boldsymbol{\gamma} \in \mathbb{R}^p_+ \end{array} \tag{13}$$

Problem (13) minimizes a non-smooth convex function over the non-negative orthant. It can be reformulated as a linear program and solved with a 'practical' complexity of $\mathcal{O}([|S|+p]^3)$, see Boyd and Vandenberghe (2004, §11). Faster solution times can be obtained through warm-starting. We characterize a sequence of two-point distributions attaining the worst-case value-at-risk in Section EC.3.4 of the electronic companion.

Although problem (13) can be solved in polynomial time, its solution may still be prohibitively expensive for large CVRP instances, where many RCIs have to be separated during the execution of a branch-and-cut scheme. It is therefore instructive to study special cases of the first-order generic moment ambiguity set (12) that allow for a faster computation of the worst-case value-at-risk (13).

Corollary 2. *If the ambiguity set* (12) *satisfies* $S_i \cap S_j = \emptyset$, $1 \leq i < j < p$, $\bigcup_{i=1}^{p-1} S_i = V_C$ *and* $S_p = V_C$, *then the worst-case value-at-risk* $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\left[\sum_{i\in S} \tilde{q}_i\right]$ *evaluates to*

$$\mathbf{1}_S{}^\top \boldsymbol{\mu} + \min\left\{\frac{\nu_p}{2\epsilon},\ \sum_{i=1}^{p-1} \min\left\{\mathbf{1}_{S\cap S_i}{}^\top \hat{\boldsymbol{q}},\ \frac{\nu_i}{2\epsilon}\right\}\right\}, \tag{14}$$

where $\hat{\boldsymbol{q}} = \min\left\{(\overline{\boldsymbol{q}} - \boldsymbol{\mu}),\ \frac{1-\epsilon}{\epsilon}(\boldsymbol{\mu} - \underline{\boldsymbol{q}})\right\}$.

The assumption that $\bigcup_{i=1}^{p-1} S_i = V_C$ comes without loss of generality as we can always add an auxiliary customer set S_i with a sufficiently large dispersion bound ν_i. The ambiguity set in Corollary 2 is a generalization of the first-order marginalized ambiguity set (6) that allows to impose upper dispersion bounds on the cumulative demand of arbitrary non-overlapping customer subsets as well as on the sum of all customer demands. The expression (14) can be evaluated in time $\mathcal{O}(|S|)$. Moreover, if a customer subset $S' \subseteq V_C$ differs from a customer subset $S \subseteq V_C$ through the inclusion or removal of a single customer, then the expression (14) associated with S' can be computed from the expression (14) associated with S in time $\mathcal{O}(p)$.

An important special case of Corollary 2 arises when $p = n + 1$ and $S_i = \{i\}$, $i = 1, \ldots, p$.

Corollary 3. *If the ambiguity set* (12) *satisfies* $p = n + 1$, $S_i = \{i\}$, $i = 1, \ldots, n$, *and* $S_{n+1} = V_C$, *then the worst-case value-at-risk* $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\left[\sum_{i \in S} \tilde{q}_i\right]$ *has the closed-form expression*

$$\mathbf{1}_S{}^\top \boldsymbol{\mu} + \min\left\{\frac{\nu_{n+1}}{2\epsilon},\ \sum_{i \in S} \min\left\{\hat{q}_i,\ \frac{\nu_i}{2\epsilon}\right\}\right\}, \tag{15}$$

where $\hat{\boldsymbol{q}} = \min\left\{(\overline{\boldsymbol{q}} - \boldsymbol{\mu}),\ \frac{1-\epsilon}{\epsilon}(\boldsymbol{\mu} - \underline{\boldsymbol{q}})\right\}$.

Compared to the first-order marginalized ambiguity set (6), the ambiguity set in Corollary 3 additionally imposes an upper bound on the sum of mean absolute deviations of all customer demands. The expression (15) can be evaluated in time $\mathcal{O}(|S|)$. Moreover, if a customer subset $S' \subseteq V_C$ differs from a customer subset $S \subseteq V_C$ through the inclusion or removal of a single customer, then the expression (15) associated with S' can be computed from the expression (15) associated with S in constant time $\mathcal{O}(1)$. If no support is present in Corollary 3, then the worst-case value-at-risk (15) reduces to $\mathbf{1}_S{}^\top \boldsymbol{\mu} + \min\{\nu_{n+1},\ \sum_{i \in S} \nu_i\}/(2\epsilon)$, which is reminiscent of a budget uncertainty set in classical robust optimization (Bertsimas and Sim, 2004).

5.2 Covariance Ambiguity Sets

We now consider second-order generic ambiguity sets (or *covariance ambiguity sets*) of the form

$$\mathcal{P} = \left\{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) \,:\, \mathbb{P}\left[\tilde{\boldsymbol{q}} \in \mathcal{Q}\right] = 1,\ \ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{q}}\right] = \boldsymbol{\mu},\ \ \mathbb{E}_{\mathbb{P}}\left[(\tilde{\boldsymbol{q}} - \boldsymbol{\mu})(\tilde{\boldsymbol{q}} - \boldsymbol{\mu})^\top\right] \preceq \boldsymbol{\Sigma}\right\}, \tag{16}$$

where $\mathcal{Q} = [\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}]$ with $\underline{\boldsymbol{q}} \geq \mathbf{0}$, $\boldsymbol{\mu} \in \operatorname{int} \mathcal{Q}$ and $\boldsymbol{\Sigma} \succ \mathbf{0}$. Note that (16) is a special case of the generic moment ambiguity set (4) where $\varphi(\boldsymbol{q}) = \max_{\boldsymbol{z} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \left\{\boldsymbol{z}^\top \left[(\boldsymbol{q} - \boldsymbol{\mu})(\boldsymbol{q} - \boldsymbol{\mu})^\top - \boldsymbol{\Sigma}\right] \boldsymbol{z}\right\}$ and $\sigma = 0$, and thus the demand estimator $d_{\mathcal{P}}$ over the ambiguity set (16) is subadditive due to Theorem 2.

The covariance ambiguity set (16) generalizes the marginalized variance ambiguity set (8). Similar to the mean absolute deviations on sums of customer demands in the previous subsection, the possibility to impose upper bounds on the covariances between pairs of customer demands allows to reduce the ambiguity whenever customer demands are not perfectly correlated. For a given set of demand observations, the upper covariance bound $\boldsymbol{\Sigma}$ can be derived analytically using McDiarmid's inequality (Delage and Ye, 2010) or empirically via bootstrapping (Chernick, 2007).

As in the previous section, distributionally robust CVRPs with ambiguity sets of the form (16) typically *cannot* be reformulated as deterministic CVRPs.

Theorem 6. *For some instances of the distributionally robust CVRP with ambiguity set* (16) *there is no deterministic CVRP instance with the same set of feasible route sets.*

Like the first-order generic ambiguity set from the previous section, the worst-case value-at-risk $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\sum_{i\in S}\tilde{q}_i\big]$ over the ambiguity set (16) can be computed in polynomial time.

Theorem 7. *The worst-case value-at-risk* $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}\big[\sum_{i\in S}\tilde{q}_i\big]$ *over the covariance ambiguity set* (16) *is equal to the optimal objective value of the optimization problem*

$$\begin{array}{ll} \text{maximize} & \mathbf{1}_S{}^\top \boldsymbol{\mu} + \mathbf{1}_S{}^\top \boldsymbol{q} \\ \text{subject to} & \boldsymbol{q}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{q} \leq \dfrac{1-\epsilon}{\epsilon} \\ & \boldsymbol{q} \in \big[\boldsymbol{q}^\ell,\, \boldsymbol{q}^u\big], \end{array} \tag{17}$$

where $\boldsymbol{q}^\ell = \max\big\{-\frac{1-\epsilon}{\epsilon}(\overline{\boldsymbol{q}} - \boldsymbol{\mu}),\, \underline{\boldsymbol{q}} - \boldsymbol{\mu}\big\}$ *and* $\boldsymbol{q}^u = \min\big\{\frac{1-\epsilon}{\epsilon}(\boldsymbol{\mu} - \underline{\boldsymbol{q}}),\, \overline{\boldsymbol{q}} - \boldsymbol{\mu}\big\}$.

Problem (17) is a convex quadratically constrained quadratic program that maximizes an affine function over the intersection of an ellipsoid and a hyperrectangle. The problem can be solved with a 'practical' complexity of $\mathcal{O}(n^3)$, see Boyd and Vandenberghe (2004, §11). We characterize a sequence of two-point distributions attaining the worst-case value-at-risk in Section EC.3.5 of the electronic companion.

Rather than solving problem (17) directly, we exploit strong convex duality, which applies since problem (17) affords a Slater point, to conclude that the dual second-order cone program

$$\begin{array}{ll} \text{minimize} & \mathbf{1}_S^\top \boldsymbol{\mu} + \sqrt{\dfrac{1-\epsilon}{\epsilon}} \big\|\boldsymbol{\Sigma}^{\frac{1}{2}}(\mathbf{e} - \boldsymbol{\lambda})\big\|_2 + \boldsymbol{q}^{\mathrm{u}\top}\boldsymbol{\lambda} \\ \text{subject to} & \boldsymbol{\lambda} \in \mathbb{R}^n_+ \end{array}$$

attains the same optimal objective value as problem (17). Due to its benign structure, the dual problem can be solved quickly using the Fast Iterative Shrinkage Thresholding Algorithm (Beck and Teboulle, 2009) with adaptive restarts (O'Donoghue and Candès, 2015) if we move the nonnegativity constraints to the objective function through indicator functions and apply a Moreau proximal smoothing (Beck and Teboulle, 2012) to the conic quadratic term in the objective function.

An important special case of Theorem 7 arises when the upper covariance bound in the ambiguity set (16) satisfies $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. This could be due to *a priori* structural knowledge about the customer demands, or by bounding a non-diagonal matrix $\boldsymbol{\Sigma}$ from above (with respect to the positive semidefinite cone) to obtain a conservative (outer) approximation of (16).

Corollary 4. *If the ambiguity set* (16) *satisfies* $\mathbf{\Sigma} = \mathrm{diag}\,(\sigma_1^2, \ldots, \sigma_n^2)$*, then the worst-case value-at-risk* $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}\big[\sum_{i\in S} \tilde{q}_i\big]$ *is equal to the optimal objective value of the optimization problem*

$$\begin{array}{ll} \text{maximize} & \mathbf{1}_S{}^\top \boldsymbol{\mu} + \displaystyle\sum_{i\in S(\theta)} q_i^{\mathrm{u}} + \sqrt{\left[\frac{1-\epsilon}{\epsilon} - \sum_{i\in S(\theta)} \left(\frac{q_i^{\mathrm{u}}}{\sigma_i}\right)^2\right]\left[\sum_{i\in S\setminus S(\theta)} \sigma_i^2\right]} \\ \text{subject to} & \theta \in \mathbb{R}_+, \end{array} \tag{18}$$

where $S(\theta) = \{i \in S \,:\, \sigma_i^2 > \theta\cdot q_i^u\}$ *and* $\boldsymbol{q}^{\mathrm{u}} = \min\left\{\frac{1-\epsilon}{\epsilon}(\boldsymbol{\mu} - \underline{\boldsymbol{q}}),\, \overline{\boldsymbol{q}} - \boldsymbol{\mu}\right\}$*, and where the feasible region is restricted to those values of* θ *for which the expression inside the square root is non-negative, that is, for which* $\sum_{i\in S(\theta)} \left(\frac{q_i^u}{\sigma_i}\right)^2 \leq \frac{1-\epsilon}{\epsilon}$.

Corollary 4 allows us to compute the worst-case value-at-risk $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}\big[\sum_{i\in S} \tilde{q}_i\big]$ over the covariance ambiguity set (16) with a diagonal upper bound $\mathbf{\Sigma}$ in time $\mathcal{O}(|S|)$, given that the ratios $q_i^{\mathrm{u}}/\sigma_i^2$ have been sorted upfront. Indeed, we can determine the set $S(\theta) \subseteq S$ that maximizes (18) through a linear search that adds a single customer to the candidate set $S(\theta)$ in each iteration.

6 Numerical Results

In this section, we compare the performance of our tailored RCI cut evaluation schemes for the generic ambiguity sets from Section 5 with a state-of-the-art commercial solver (Section 6.1), and we compare the runtimes of the resulting branch-and-cut algorithms for the distributionally robust chance constrained CVRP with a corresponding implementation for the deterministic CVRP (Section 6.2). Further numerical results can be found in Section EC.4 of the electronic companion, where we investigate how the parameter values of the marginal ambiguity sets from Section 4 impact the solution of the associated deterministic CVRP instances.

With the exception of Section 6.1 below, all numerical results are based on the CVRP benchmark problems compiled by Díaz (2006). The instances are named 'X-nY-kZ', where X denotes the literature source of the instance, Y is the number of nodes in the instance (including the depot) and Z is the number of vehicles. We only consider those problems for which two-dimensional coordinates for the nodes are available. Following the literature convention, we set the transportation costs c_{ij} to the Euclidean distance between i and j, rounded to the nearest integer.

Since the CVRP benchmark problems contain deterministic customer demands, we generate distributions for our stochastic demands according to the following procedure. The unscaled de-

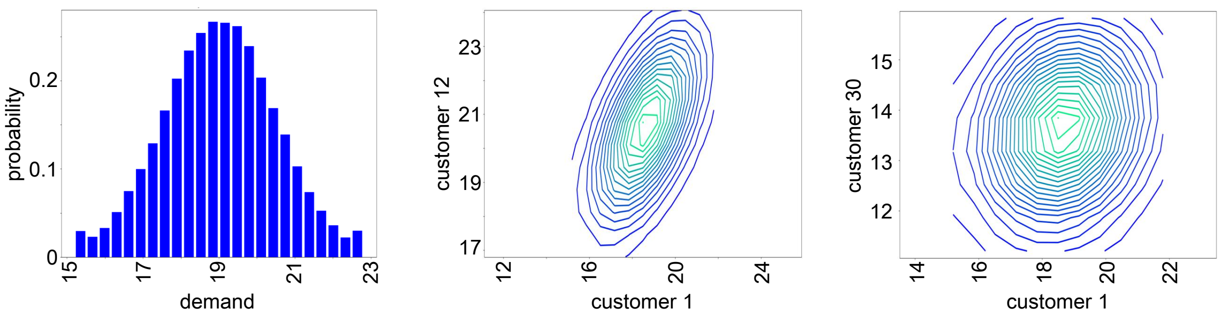

Figure 3. Demand distributions for the instance A-n32-k5. The left graph visualizes the histogram for customer 1, whereas the middle (right) graph illustrates the joint demand distribution of customers 1 and 12 (1 and 30), which are located nearby (far away).

mand of customer $i \in V_C$ is set to $\tilde{\chi}_i = \frac{1}{2}\tilde{\xi}_i + \frac{1}{2|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i}\tilde{\xi}_j$, where $\tilde{\boldsymbol{\xi}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ is an n-dimensional normally distributed random vector and $\mathcal{N}_i \subseteq V_C$ is the set of the $\lfloor 0.1n \rfloor$ customers closest to i in terms of Euclidean distance. We subsequently apply an affine transformation which ensures that the expected demand of customer i is μ_i, which we identify with customer i's nominal demand from the deterministic CVRP instance, and that 99% of customer i's demand falls into the interval $[\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}]$, where the bounds $(\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}})$ are set to $(\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}) = (0.8\boldsymbol{\mu}, 1.2\boldsymbol{\mu})$, unless specified otherwise. Finally, we clamp customer i's scaled demand distribution to the interval $[\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}]$. Our construction ensures that the customer demands exhibit a dependence structure that is informed by geographical proximity, see Figure 3. Since the unused vehicle capacities tend to be small already in the deterministic CVRP instances, we follow the approach in Gounaris et al. (2013) and increase the vehicle capacities Q in each benchmark instance by 20%. This ensures that all distributionally robust CVRP instances remain feasible. We set the risk threshold to $\epsilon = 0.2$.

We solve the deterministic and distributionally robust CVRP instances with a 'vanilla' branch-and-cut algorithm that only separates RCI cuts according to the Tabu Search procedure proposed by Augerat et al. (1998). Our branch-and-cut algorithm is implemented in C++ and uses the branch-and-bound capability of CPLEX 12.8.[1] The source code of the proposed branch-and-cut algorithm is available as part of the paper's online supplement. We solve all problems in single-core mode on an Intel Xeon 2.66GHz processor with 8GB memory and a runtime limit of 12 hours.

[1]CPLEX website: `https://www.ibm.com/analytics/cplex-optimizer`.

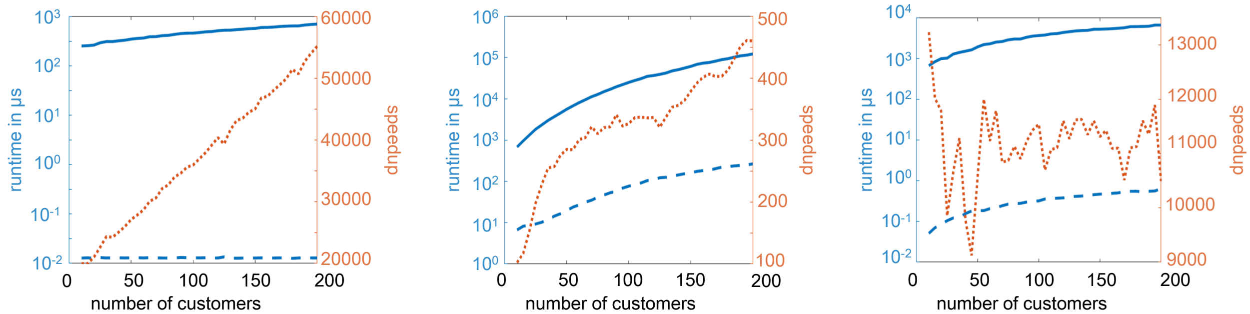

Figure 4. Runtimes for RCI cut evaluation. Shown are the average runtimes that CPLEX (solid lines) and our evaluation schemes (dashed lines) for first-order ambiguity sets (left graph), generic covariance ambiguity sets (middle graph) and covariance ambiguity sets with diagonal $\boldsymbol{\Sigma}$ (right graph) require to evaluate the right-hand side of a single RCI cut. The dotted lines represent the implied speedups.

6.1 Generic Ambiguity Sets: RCI Cut Evaluation

We first compare our tailored evaluation of the worst-case value-at-risk $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\sum_{i\in S} \tilde{q}_i\big]$ for first-order generic moment ambiguity sets (12) and covariance ambiguity sets (16) with their solution as linear and quadratically constrained quadratic programs via CPLEX, respectively. To this end, we generate random problem instances in which $n \in \{10, 15, \ldots, 200\}$ customers have nominal demands μ_i that are uniformly distributed on the set $\{1, \ldots, 10\}$ as well as random locations that are uniformly distributed on the square $[0, 10]^2$. We generate the demand distributions as described in the beginning of this section.

For the first-order ambiguity set (12), we partition the customers into four quadrants of equal size, and we select mean absolute deviations bounds for each quadrant as well as for the cumulative demands based on a sample from the joint demand distribution. Similarly, for the covariance ambiguity set (16), we select the covariance bound $\boldsymbol{\Sigma}$ based on a sample from the joint demand distribution. We also consider the special case of a diagonal covariance bound (see Corollary 4) where we set all non-diagonal elements of the previously described covariance bound $\boldsymbol{\Sigma}$ to zero.

Figure 4 compares the runtimes of our tailored evaluation schemes with those of CPLEX for evaluating the right-hand side of a single RCI cut, that is, an individual worst-case value-at-risk $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\sum_{i\in S} \tilde{q}_i\big]$, on 1,000 randomly generated problem instances for each instance size

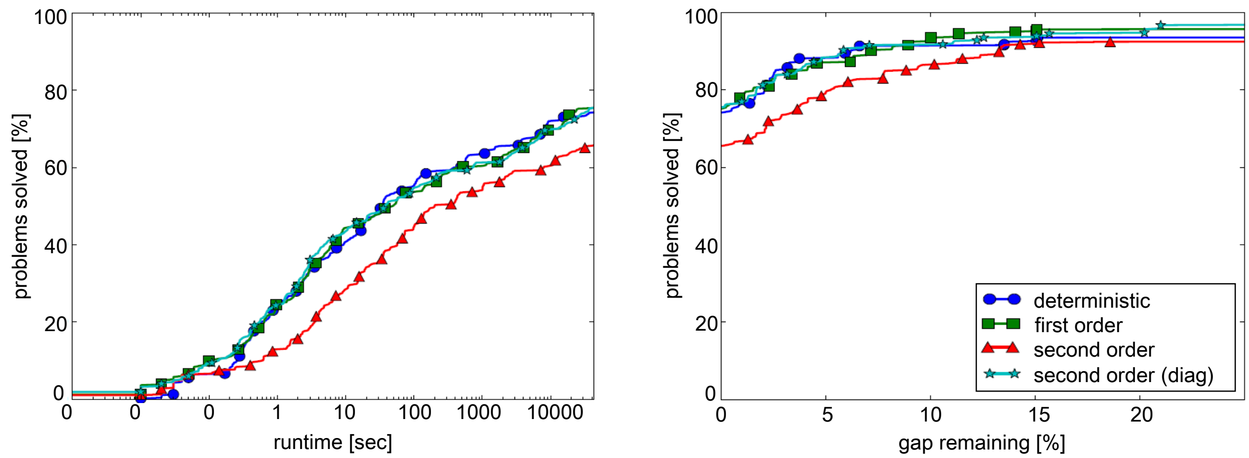

Figure 5. Runtimes and optimality gaps for our branch-and-cut schemes. Shown are the runtimes (left graph) and optimality gaps after 12 hours (right graph) for our deterministic branch-and-cut scheme with nominal demands $\boldsymbol{q} = \overline{\boldsymbol{q}}$ (blue, circles) as well as our distributionally robust branch-and-cut schemes over first order ambiguity sets (green, squares), second order ambiguity sets (red, triangles) and second order ambiguity sets with diagonal covariance bounds (cyan, stars).

n. While the achieved speedups are most significant for first-order ambiguity sets, they remain substantial for covariance ambiguity sets, especially if the covariance bound $\boldsymbol{\Sigma}$ is diagonal. The results are in line with our theoretical complexity estimates from Section 5, and they confirm our intuition that it is essential to study special classes of ambiguity sets $\mathcal{P}$ that give rise to easily computable demand estimators $d_{\mathcal{P}}$.

6.2 Generic Ambiguity Sets: Branch-and-Cut Scheme

We finally use our RCI cut evaluation schemes for the first-order and the two covariance ambiguity sets from the previous subsection to solve the CVRP benchmark instances of Díaz (2006). We compare the runtimes and optimality gaps of the resulting branch-and-cut procedures with those of a deterministic branch-and-cut algorithm applied to the deterministic CVRP with worst-case demands $\boldsymbol{q} = \overline{\boldsymbol{q}}$. The results are summarized in Figure 5 as well as in Section EC.5 of the electronic companion.

The results show that our branch-and-cut schemes for the first-order as well as the diagonal covariance ambiguity set perform very similar to the branch-and-cut scheme for the deterministic

CVRP, both in terms of the runtimes for successfully solved instances as well as the optimality gaps after 12 hours of runtime. In particular, all three algorithms can solve about 75% of the benchmark instances within the time limit, and the optimality gap is below 10% for roughly 90% of the instances. As expected from the previous subsection, our branch-and-cut scheme for covariance ambiguity sets with a non-diagonal bound $\boldsymbol{\Sigma}$ is slower; it solves about 65% of the benchmark instances within 12 hours, and the optimality gap is below 10% for about 80% of the instances.

To assess the conservatism of the obtained solutions, we consider 55 instances where all of the branch-and-cut schemes determined optimal solutions within the time limit and where the optimal solution of the deterministic CVRP with expected demands $\boldsymbol{q} = \boldsymbol{\mu}$ (henceforth the 'nominal solution') has strictly lower transportation costs than the optimal solution of the deterministic CVRP with component-wise worst-case demands $\boldsymbol{q} = \overline{\boldsymbol{q}}$ (henceforth the 'worst-case solution'). We then compute how much of the objective gap between the nominal solution and the robust solution is covered by each of the distributionally robust solutions. For our first-order ambiguity set as well as the covariance ambiguity set with a diagonal bound, every distributionally robust solution improves upon the worst-case solution, and the solutions close 75.5% of the objective gap on average. For our covariance ambiguity set with a non-diagonal bound, the distributionally robust solutions improve upon the worst-case solutions in 53 out of the 55 instances and close 52.1% of the objective gap on average. The results indicate that the distributionally robust CVRP can help to reduce the conservatism of naïve worst-case solutions.

7 Conclusions

Motivated by some of the shortcomings of the classical chance constrained CVRP, which assumes that the uncertain customer demands are governed by a precisely known distribution, we investigated the distributionally robust CVRP, in which this distribution is only partially characterized. In particular, we studied the computational tractability of the distributionally robust CVRP.

The solvability of the distributionally robust CVRP is largely determined by the choice of the ambiguity set. First and foremost, the ambiguity set should lead to a subadditive demand estimator so that standard branch-and-cut schemes can be used to solve the problem. This turns out to be the case for a large class of moment ambiguity sets. Secondly, the demand estimator should be easily computable. To this end, we have identified several classes of first-order and second-order moment ambiguity sets whose demand estimators can be computed by tailored algorithms that outperform an off-the-shelf commercial optimization package by orders of magnitude.

An interesting question that we have *not* touched upon is the comparison of different ambiguity sets in terms of their statistical properties, such as the degree of conservativeness of the obtained solutions as well as the amount of data required for a sufficiently accurate calibration of the ambiguity set. We believe that such a comparison would best be done empirically using a large set of benchmark instances, and we identify this as a fruitful avenue for future research. We also note that some of our results may have applications outside the domain of vehicle routing. For example, Theorem 3 can be readily generalized to show that linear worst-case chance constraints over marginalized moment ambiguity sets reduce to deterministic inequality constraints. It therefore appears instructive to further explore the consequences of highly structured ambiguity sets in distributionally robust optimization in general.

Acknowledgments

The authors gratefully acknowledge funding from the Imperial Business Analytics Centre and the EPSRC grants EP/M028240/1, EP/M027856/1 and EP/N020030/1. We are also indebted to Philipp Dufter for valuable help in the early stages of the manuscript, as well as the constructive suggestions of the anonymous referees.

Electronic Companion

This e-companion contains material that has been omitted from the main paper for the sake of brevity and readability. Section EC.8 discusses several shortcomings of the classical chance constrained CVRP formulation. Section EC.9 provides an example where the value-at-risk fails to be additive for individual probability distributions contained in a marginalized moment ambiguity set. Section EC.10 provides series of worst-case distributions that attain the worst-case values-at-risk for each of the five ambiguity sets considered in the main paper. Section EC.11 provides a numerical example that explores how the minimum number of vehicles as well as the transportation costs increase with the coefficient of variation for a marginalized variance ambiguity set. Section EC.12 provides omitted details of our numerical experiments from the main paper. Section EC.13, finally, contains the proofs of all results from the main text.

EC.8 Shortcomings of the Chance Constrained CVRP

In the following three examples, we discuss each of the challenges of modeling the random customer demands using a known distribution in the chance constrained CVRP.

Example 2 (Independence)**.** *Consider a chance constrained CVRP instance where a single vehicle of capacity $Q = 12$ serves the customers $V_C = \{1, \ldots, 20\}$. The marginal distribution of each*

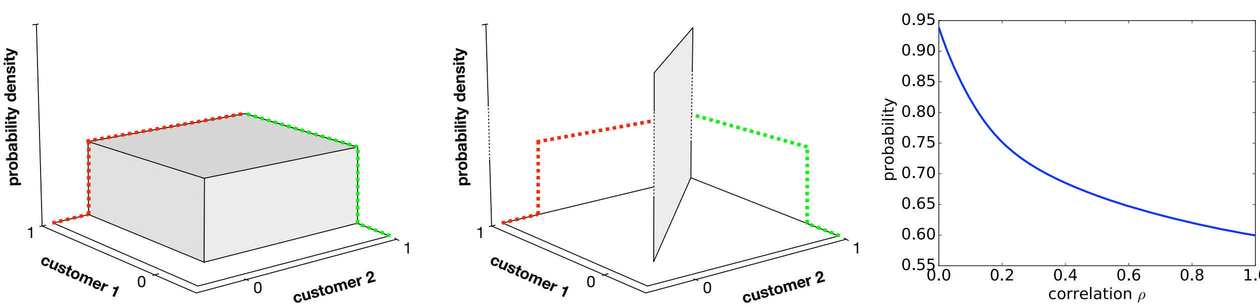

Figure 6. Chance constrained CVRP instance with uniformly distributed marginal customer demands (dotted lines) that are combined through a Gaussian copula. The left and middle graphs illustrate projections of the probability density functions corresponding to the correlations $\rho = 0$ and $\rho = 1$ onto two customers, respectively, and the right graph presents the probabilities of satisfying the vehicle's capacity for all 20 customers.

customer's demand is a uniform distribution over the interval $[0, 1]$. *We model the dependence between the customer demands via a Gaussian copula. The left and the middle graph in Figure 6 visualize the joint demand distribution for two customers when their demands are independent (correlation* $\rho = 0$*) and perfectly dependent (correlation* $\rho = 1$*), respectively. The right graph in Figure 6 visualizes the probability* $\mathbb{Q}\big[\mathbf{e}^\top \tilde{\boldsymbol{q}} \leq Q\big]$ *of satisfying the capacity restriction of the vehicle for varying levels of demand dependence. While* 20 *customers with independently distributed demands can be served with a high probability of approximately* 0.95*, this probability decreases to* 0.6 *for perfectly correlated (comonotone) demands. We thus conclude that it is crucial to model demand dependencies that may be present in the problem instance.*

For a branch-and-cut-and-price algorithm for the chance constrained CVRP that does not require independent customer demands, we refer to Dinh et al. (2016, 2017).

Example 3 (Complexity)**.** *Consider a chance constrained CVRP instance where the customer demands are uniformly distributed over a hyperrectangle* $[\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}]$ *with* $\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}} \in \mathbb{R}^n_+$. *In this case, evaluating the probability* $\mathbb{Q}\big[\mathbf{1}_{\boldsymbol{R}_k}{}^\top \tilde{\boldsymbol{q}} \leq Q\big]$ *of satisfying the capacity of vehicle* k *is tantamount to calculating the volume of the knapsack polytope, which is known to be #P-hard (Dyer and Stougie, 2006; Hanasusanto et al., 2016). This is problematic for exact solution schemes, which typically rely on the repeated evaluation of the feasibility of candidate routes to determine an optimal route set.*

We note that if the customer demands follow a multivariate normal distribution, then the cumulative demand along a candidate route is also normally distributed. In this case, the satisfaction of the corresponding vehicle's capacity reduces to evaluating the inverse cumulative distribution function of a standard normal distribution, which can be done efficiently. Moreover, by invoking a central limit theorem, a similar argument can be made for non-normally distributed customer demands as long as *(i)* the customer demands are (sufficiently) independent and *(ii)* each vehicle serves sufficiently many customers (with 30 being a common quote in the literature).

Example 4 (Estimation)**.** *Consider a chance constrained CVRP instance where three vehicles of capacity* $Q = 10$ *serve the customer set* $V_C = \{1, \ldots, 8\}$. *The expected customer demands are* $\boldsymbol{\mu} = (3, 5, 2, 5, 1, 6, 1, 1)^\top$*, and each customer demand* $\tilde{q}_i$ *follows an independent uniform distribution supported on* $[(2/3)\mu_i, (4/3)\mu_i]$. *The left part of Figure 7 shows the route set* $\mathbf{R} = (\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_3)$*,* $\boldsymbol{R}_1 = (1, 2)$*,* $\boldsymbol{R}_2 = (3, 4, 5)$ *and* $\boldsymbol{R}_3 = (6, 7, 8)$*, which is feasible at the tolerance* $\epsilon = 0.05$ *since*

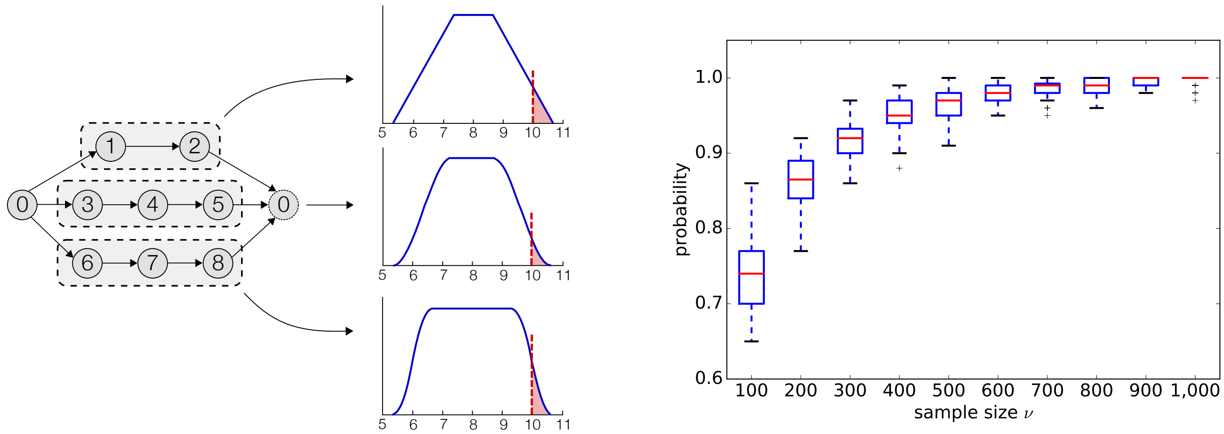

Figure 7. Chance constrained CVRP instance with independent and uniformly distributed customer demands. The three graphs in the middle visualize the true probability density functions of the cumulative customer demands, which correspond to generalized Irwin-Hall distributions, for the three routes on the left. The graph on the right shows the likelihood of the route set on the left being feasible if we replace the true distribution with an empirical distribution $\mathbb{Q}^\nu$ of varying sample size ν. The box-and-whisker plots report the ranges and the quartiles of 1,000 statistically independent sets of samples.

the vehicles' capacities are satisfied with probability 0.97*,* 0.98 *and* 0.97*, respectively. (For ease of illustration, we have duplicated the depot in the figure.) In practice, the true distribution $\mathbb{Q}$ of the customer demands is typically unknown. In this case, the literature often suggests to replace the unknown true distribution $\mathbb{Q}$ with the empirical distribution $\mathbb{Q}^\nu = \frac{1}{\nu}\sum_\ell \delta_{\boldsymbol{q}^\ell}$, where $\boldsymbol{q}^1, \ldots, \boldsymbol{q}^\nu$ denote historical observations of the customer demands under the distribution $\mathbb{Q}$; this approach is often referred to as 'sample average approximation' in the stochastic programming literature (Shapiro et al., 2014). The right part of Figure 7 shows the likelihood of the route set $\mathbf{R}$ being feasible (*i.e.*, satisfying the capacity constraints $\mathbb{Q}^\nu\left[\tilde{q}_1 + \tilde{q}_2 \leq 10\right] \geq 0.95$, $\mathbb{Q}^\nu\left[\tilde{q}_3 + \tilde{q}_4 + \tilde{q}_5 \leq 10\right] \geq 0.95$ and $\mathbb{Q}^\nu\left[\tilde{q}_6 + \tilde{q}_7 + \tilde{q}_8 \leq 10\right] \geq 0.95$) if we replace the unknown true distribution $\mathbb{Q}$ with the empirical distribution $\mathbb{Q}^\nu$ resulting from different sample sizes ν. We observe that despite the small number of customers and vehicles, $\nu \geq 800$ samples are required for the route set $\mathbf{R}$ to be feasible under the chance constraints corresponding to the empirical distribution $\mathbb{Q}^\nu$ with a confidence of* 0.99*.*

We note that smaller numbers of samples than those presented in Example 4 may suffice in practice if the risk tolerance ϵ of the value-at-risk is adjusted judiciously. The interested reader is referred to Luedtke and Ahmed (2008) for further details.

EC.9 Additivity of the Value-at-Risk over Marginalized Moment Ambiguity Sets

The following example illustrates that for individual probability distributions within a marginalized moment ambiguity set of the form (5), the corresponding values-at-risk typically *fail* to be additive.

Example 5. *Consider the following marginalized moment ambiguity set for two customers:*

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^2) \,:\, \mathbb{P}(\tilde{\boldsymbol{q}} \in [2,7] \times [5,15]) = 1,\;\; \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{q}}] = (3.2,\, 7.8)^\top,\;\; \mathbb{E}_{\mathbb{P}}[|\tilde{q}_i - \mu_i|] \leq 1.5,\, i = 1,2 \right\}$$

Here, $|\cdot|$ denotes the absolute value operator. We have $\mathbb{P}_1 \in \mathcal{P}$ for the distribution $\mathbb{P}_1 = \mathbb{P}_{11} \times \mathbb{P}_{12}$ under which the two customer demands are independent and governed by the marginal distributions $\mathbb{P}_{11} = 0.8 \cdot \delta_3 + 0.2 \cdot \delta_4 \in \mathcal{P}_0(\mathbb{R})$ and $\mathbb{P}_{12} = 0.8 \cdot \delta_7 + 0.2 \cdot \delta_{11} \in \mathcal{P}_0(\mathbb{R})$. One readily verifies that

$$14 \;=\; \mathbb{P}_1\text{-VaR}_{0.9}\left[\tilde{q}_1 + \tilde{q}_2\right] \;<\; \mathbb{P}_1\text{-VaR}_{0.9}\left[\tilde{q}_1\right] + \mathbb{P}_1\text{-VaR}_{0.9}\left[\tilde{q}_2\right] \;=\; 4 + 11,$$

that is, the 0.9-value-at-risk is subadditive *under $\mathbb{P}_1$. Likewise, we have $\mathbb{P}_2 \in \mathcal{P}$ for the distribution $\mathbb{P}_2 = \mathbb{P}_{21} \times \mathbb{P}_{22}$ with the marginals $\mathbb{P}_{21} = 0.6 \cdot \delta_2 + 0.4 \cdot \delta_5 \in \mathcal{P}_0(\mathbb{R})$ and $\mathbb{P}_{22} = 0.9 \cdot \delta_7 + 0.1 \cdot \delta_{15} \in \mathcal{P}_0(\mathbb{R})$. The distribution $\mathbb{P}_2$ satisfies*

$$12 \;=\; \mathbb{P}_2\text{-VaR}_{0.9}\left[\tilde{q}_1 + \tilde{q}_2\right] \;=\; \mathbb{P}_2\text{-VaR}_{0.9}\left[\tilde{q}_1\right] + \mathbb{P}_2\text{-VaR}_{0.9}\left[\tilde{q}_2\right] \;=\; 5 + 7,$$

that is, the value-at-risk is additive *under $\mathbb{P}_2$. Finally, we have $\mathbb{P}_3 \in \mathcal{P}$ for $\mathbb{P}_3 = \mathbb{P}_{31} \times \mathbb{P}_{32}$ with the marginals $\mathbb{P}_{31} = 0.9 \cdot \delta_3 + 0.1 \cdot \delta_5 \in \mathcal{P}_0(\mathbb{R})$ and $\mathbb{P}_{32} = 0.9 \cdot \delta_7 + 0.1 \cdot \delta_{15} \in \mathcal{P}_0(\mathbb{R})$. One verifies that*

$$12 \;=\; \mathbb{P}_3\text{-VaR}_{0.9}\left[\tilde{q}_1 + \tilde{q}_2\right] \;>\; \mathbb{P}_3\text{-VaR}_{0.9}\left[\tilde{q}_1\right] + \mathbb{P}_3\text{-VaR}_{0.9}\left[\tilde{q}_2\right] \;=\; 3 + 7,$$

that is, the value-at-risk is superadditive *under $\mathbb{P}_3$.*

EC.10 Worst Case Distributions

In this section, we consider each of the five ambiguity sets from the main paper and provide sequences of two-point distributions that are contained in the ambiguity set and that attain the worst-case values-at-risk asymptotically.

EC.10.1 First Order Marginalized Ambiguity Sets

Proposition EC.1. *A sequence of distributions $\mathbb{P}_i^t \in \mathcal{P}$, $t = 1, 2, \ldots$, that attain the worst-case value-at-risk* $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big]$ *in Proposition 2 asymptotically as $t \longrightarrow \infty$ can be defined as follows:*

(i) *If* (7) *is minimized by $\overline{q}_i - \mu_i$, then $\mathbb{P}_i^t = (1-\epsilon-1/t)\cdot\delta_{\boldsymbol{q}_1} + (\epsilon+1/t)\cdot\delta_{\boldsymbol{q}_2}$ with $\boldsymbol{q}_1 = \boldsymbol{\mu} - \frac{\epsilon}{1-\epsilon}(\overline{q}_i - \mu_i)\mathbf{e}_i$ and $\boldsymbol{q}_2 = \boldsymbol{\mu} + \frac{1-\epsilon-1/t}{\epsilon+1/t}\frac{\epsilon}{1-\epsilon}(\overline{q}_i - \mu_i)\mathbf{e}_i$.*

(ii) *If* (7) *is minimized by $\frac{1-\epsilon}{\epsilon}(\mu_i - \underline{q}_i)$, then $\mathbb{P}_i^t = (1-\epsilon-1/t)\cdot\delta_{\boldsymbol{q}_1} + (\epsilon+1/t)\cdot\delta_{\boldsymbol{q}_2}$ with $\boldsymbol{q}_1 = \boldsymbol{\mu} - (\mu_i - \underline{q}_i)\mathbf{e}_i$ and $\boldsymbol{q}_2 = \boldsymbol{\mu} + \frac{1-\epsilon-1/t}{\epsilon+1/t}(\mu_i - \underline{q}_i)\mathbf{e}_i$.*

(iii) *If* (7) *is minimized by $\frac{1}{2\epsilon}\sigma_i$, then $\mathbb{P}_i^t = (1-\epsilon-1/t)\cdot\delta_{\boldsymbol{q}_1} + (\epsilon+1/t)\cdot\delta_{\boldsymbol{q}_2}$ with $\boldsymbol{q}_1 = \boldsymbol{\mu} - \frac{\sigma_i}{2(1-\epsilon)}\mathbf{e}_i$ and $\boldsymbol{q}_2 = \boldsymbol{\mu} + \frac{1-\epsilon-1/t}{\epsilon+1/t}\frac{\sigma_i}{2(1-\epsilon)}\mathbf{e}_i$.*

Proof of Proposition EC.1. We have to show for each of the three cases that $\mathbb{P}_i^t \in \mathcal{P}$, that is, that *(a)* $\mathbb{P}_i^t$ is supported on $\mathcal{Q}$, *(b)* $\mathbb{E}_{\mathbb{P}_i^t}\big[\tilde{\boldsymbol{q}}\big] = \boldsymbol{\mu}$ and *(c)* $\mathbb{E}_{\mathbb{P}_i^t}\big[|\tilde{\boldsymbol{q}} - \boldsymbol{\mu}|\big] \leq \boldsymbol{\sigma}$ hold. The claim of the proposition then follows since in each of the three cases, the distribution places a probability mass of $\epsilon + 1/t$ on $\boldsymbol{q}_2$, and q_{2i} converges to $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big]$ as $t \longrightarrow \infty$. The proof that $\mathbb{P}_i^t \in \mathcal{P}$ follows along similar lines as the proof of Proposition 1 and is thus omitted for the sake of brevity. □

Proposition EC.1 presents a sequence of worst-case distributions $\mathbb{P}_i^t$ for the demand $\tilde{q}_i$ of each *individual* customer $i \in V_C$. Using similar arguments as in the proof of Theorem 3 from the main text, we can construct a sequence of worst-case distributions $\mathbb{P}^t$ for the demand of *all* customers $i \in V_C$ via the Fréchet-Hoeffding upper bound copula

$$\mathbb{P}^t(\tilde{\boldsymbol{q}} \leq \boldsymbol{q}) = \min_{i\in V_C} \mathbb{P}_i^t(\tilde{q}_i \leq q_i).$$

For the ambiguity set (6), this sequence of worst-case distributions $\mathbb{P}^t$ has a simple description: it satisfies $\mathbb{P}^t \longrightarrow (1-\epsilon)\cdot\delta_{\boldsymbol{q}_1} + \epsilon\cdot\delta_{\boldsymbol{q}_2}$ for the two-point distribution characterized by

$$(q_{1i}, q_{2i}) = \begin{cases} \left(\frac{1}{1-\epsilon}\mu_i - \frac{\epsilon}{1-\epsilon}\overline{q}_i,\ \overline{q}_i\right) & \text{if (7) is minimized by } \overline{q}_i - \mu_i, \\ \left(\underline{q}_i,\ \frac{1}{\epsilon}\mu_i - \frac{1-\epsilon}{\epsilon}\underline{q}_i\right) & \text{if (7) is minimized by } \frac{1-\epsilon}{\epsilon}(\mu_i - \underline{q}_i), \\ \left(\mu_i - \frac{\sigma_i}{2(1-\epsilon)},\ \mu_i + \frac{\sigma_i}{2\epsilon}\right) & \text{if (7) is minimized by } \frac{1}{2\epsilon}\sigma_i \end{cases} \qquad \forall i \in V_C.$$

This joint worst-case distribution is illustrated in Figure 8 for an example with two customers.

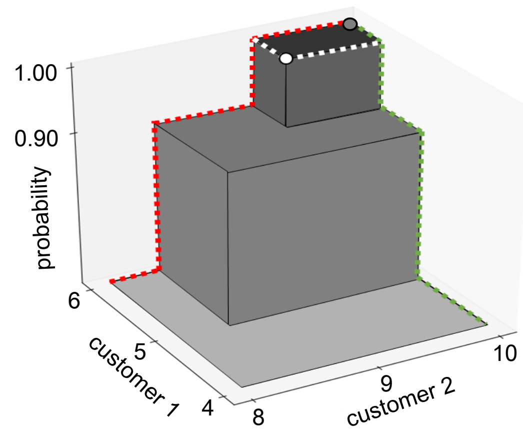

Figure 8. Fréchet-Hoeffding upper bound copula $\mathbb{P}^{\star}$ for a distributionally robust CVRP instance with two customers and a marginalized first-order ambiguity set (6) with support $\mathcal{Q} = [4,6] \times [8,10]$, mean $\boldsymbol{\mu}^{\top} = (4.5, 9)$, mean absolute deviation bounds $\boldsymbol{\sigma}^{\top} = (0.1, 0.15)$ and risk threshold $\epsilon = 0.1$. The marginal distributions are highlighted via dotted green and red lines, and the worst-case value-at-risk $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_1 + \tilde{q}_2\big]$ (the worst-case demand realization) is indicated by a white (dark gray) circle.

Example 5 (cont'd)**.** *Proposition EC.2 shows that the* 0.9*-worst-case values-at-risk for the two customer demands* $\tilde{q}_1$ *and* $\tilde{q}_2$ *from our previous example are* 5 *and* 15*, respectively. Moreover, Proposition EC.1 implies that each individual worst-case value-at-risk is attained asymptotically by a sequence of distributions that converges to the asymptotic distribution* $0.9 \cdot \delta_{\boldsymbol{q}_1} + 0.1 \cdot \delta_{\boldsymbol{q}_2}$ *with* $\boldsymbol{q}_1 = (3,7)^{\top}$ *and* $\boldsymbol{q}_2 = (5,15)^{\top}$*. Finally, since each element of the sequence is a two-point distribution that places a probability mass greater than* 0.1 *on each scenario, the value-at-risk is indeed additive for each member of the sequence. Note, however, that although the worst-case values-at-risk of* $\tilde{q}_1 + \tilde{q}_2$ *converge to* 20 *for the distributions in the sequence, the worst-case value-at-risk under the asymptotic distribution is* 10*.*

EC.10.2 Variance Marginalized Ambiguity Sets

Proposition EC.2. *A sequence of distributions* $\mathbb{P}_i^t \in \mathcal{P}$*,* $t = 1, 2, \ldots,$ *that attain the worst-case value-at-risk* $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big]$ *in Proposition 3 asymptotically as* $t \longrightarrow \infty$ *can be defined as follows:*

(i) *If* (9) *is minimized by* $\overline{q}_i - \mu_i$, *then* $\mathbb{P}_i^t = (1-\epsilon-1/t)\cdot\delta_{\boldsymbol{q}_1} + (\epsilon+1/t)\cdot\delta_{\boldsymbol{q}_2}$ *with* $\boldsymbol{q}_1 = \boldsymbol{\mu} - \frac{\epsilon}{1-\epsilon}(\overline{q}_i - \mu_i)\mathbf{e}_i$ *and* $\boldsymbol{q}_2 = \boldsymbol{\mu} + \frac{1-\epsilon-1/t}{\epsilon+1/t}\frac{\epsilon}{1-\epsilon}(\overline{q}_i - \mu_i)\mathbf{e}_i$.

(ii) *If* (9) *is minimized by* $\frac{1-\epsilon}{\epsilon}(\mu_i - \underline{q}_i)$, *then* $\mathbb{P}_i^t = (1-\epsilon-1/t)\cdot\delta_{\boldsymbol{q}_1} + (\epsilon+1/t)\cdot\delta_{\boldsymbol{q}_2}$ *with* $\boldsymbol{q}_1 = \boldsymbol{\mu} - (\mu_i - \underline{q}_i)\mathbf{e}_i$ *and* $\boldsymbol{q}_2 = \boldsymbol{\mu} + \frac{1-\epsilon-1/t}{\epsilon+1/t}(\mu_i - \underline{q}_i)\mathbf{e}_i$.

(iii) *If* (9) *is minimized by* $\sqrt{\frac{1-\epsilon}{\epsilon}}\sigma_i$, *then* $\mathbb{P}_i^t = (1-\epsilon-1/t)\cdot\delta_{\boldsymbol{q}_1} + (\epsilon+1/t)\cdot\delta_{\boldsymbol{q}_2}$ *with* $\boldsymbol{q}_1 = \boldsymbol{\mu} - \sqrt{\frac{\epsilon}{1-\epsilon}}\sigma_i\mathbf{e}_i$ *and* $\boldsymbol{q}_2 = \boldsymbol{\mu} + \frac{1-\epsilon-1/t}{\epsilon+1/t}\sqrt{\frac{\epsilon}{1-\epsilon}}\sigma_i\mathbf{e}_i$.

Proof of Proposition EC.2. This proposition can be proved in the same way as Proposition EC.1. □

EC.10.3 Semi-Variance Marginalized Ambiguity Sets

Proposition EC.3. *A sequence of distributions* $\mathbb{P}_i^t \in \mathcal{P}$, $t = 1, 2, \ldots$, *that attain the worst-case value-at-risk* $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big]$ *in Proposition 4 asymptotically as* $t \longrightarrow \infty$ *can be defined as follows:*

(i) *If* (11) *is minimized by* $\overline{q}_i - \mu_i$, *then* $\mathbb{P}_i^t = (1-\epsilon-1/t)\cdot\delta_{\boldsymbol{q}_1} + (\epsilon+1/t)\cdot\delta_{\boldsymbol{q}_2}$ *with* $\boldsymbol{q}_1 = \boldsymbol{\mu} - \frac{\epsilon}{1-\epsilon}(\overline{q}_i - \mu_i)\mathbf{e}_i$ *and* $\boldsymbol{q}_2 = \boldsymbol{\mu} + \frac{1-\epsilon-1/t}{\epsilon+1/t}\frac{\epsilon}{1-\epsilon}(\overline{q}_i - \mu_i)\mathbf{e}_i$.

(ii) *If* (11) *is minimized by* $\frac{1-\epsilon}{\epsilon}(\mu_i - \underline{q}_i)$, *then* $\mathbb{P}_i^t = (1-\epsilon-1/t)\cdot\delta_{\boldsymbol{q}_1} + (\epsilon+1/t)\cdot\delta_{\boldsymbol{q}_2}$ *with* $\boldsymbol{q}_1 = \boldsymbol{\mu} - (\mu_i - \underline{q}_i)\mathbf{e}_i$ *and* $\boldsymbol{q}_2 = \boldsymbol{\mu} + \frac{1-\epsilon-1/t}{\epsilon+1/t}(\mu_i - \underline{q}_i)\mathbf{e}_i$.

(iii) *If* (11) *is minimized by* $\sqrt{\frac{\sigma_i^+}{\epsilon}}$, *then* $\mathbb{P}_i^t = (1-\epsilon-1/t)\cdot\delta_{\boldsymbol{q}_1} + (\epsilon+1/t)\cdot\delta_{\boldsymbol{q}_2}$ *with* $\boldsymbol{q}_1 = \boldsymbol{\mu} - \frac{\epsilon}{1-\epsilon}\sqrt{\frac{\sigma_i^+}{\epsilon}}\mathbf{e}_i$ *and* $\boldsymbol{q}_2 = \boldsymbol{\mu} + \frac{1-\epsilon-1/t}{\epsilon+1/t}\frac{\epsilon}{1-\epsilon}\sqrt{\frac{\sigma_i^+}{\epsilon}}\mathbf{e}_i$.

(iv) *If* (11) *is minimized by* $\frac{\sqrt{(1-\epsilon)\sigma_i^-}}{\epsilon}$, *then* $\mathbb{P}_i^t = (1-\epsilon-1/t)\cdot\delta_{\boldsymbol{q}_1} + (\epsilon+1/t)\cdot\delta_{\boldsymbol{q}_2}$ *with* $\boldsymbol{q}_1 = \boldsymbol{\mu} - \sqrt{\frac{\sigma_i^-}{1-\epsilon}}\mathbf{e}_i$ *and* $\boldsymbol{q}_2 = \boldsymbol{\mu} + \frac{1-\epsilon-1/t}{\epsilon+1/t}\sqrt{\frac{\sigma_i^-}{1-\epsilon}}\mathbf{e}_i$.

Proof of Proposition EC.3. This proposition can be proved in the same way as Proposition EC.1. □

EC.10.4 First Order Generic Ambiguity Sets

Proposition EC.4. *A sequence of distributions $\mathbb{P}^t \in \mathcal{P}$, $t = 1, 2, \ldots$, that attain the worst-case value-at-risk* $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\sum_{i\in S} \tilde{q}_i\big]$ *in Theorem 5 asymptotically as $t \longrightarrow \infty$ can be defined as $\mathbb{P}^t = (\xi_1 - 1/t) \cdot \delta_{\boldsymbol{q}_1} + (\xi_2 + 1/t) \cdot \delta_{\boldsymbol{q}_2}$, where $\boldsymbol{q}_1 = \frac{\boldsymbol{\zeta}_1}{\xi_1}$ and $\boldsymbol{q}_2 = \frac{\boldsymbol{\zeta}_2}{\xi_2} - \frac{1}{t}\left(\frac{\boldsymbol{\zeta}_2}{\xi_2} - \frac{\boldsymbol{\zeta}_1}{\xi_1}\right)$, and $(\xi_i, \boldsymbol{\zeta}_i)$ is an optimal solution to the linear program*

$$
\begin{array}{lll}
\text{minimize} & \xi_1 & \\
\text{subject to} & \xi_1 + \xi_2 = 1, & \boldsymbol{\zeta}_1 + \boldsymbol{\zeta}_2 = \boldsymbol{\mu} \\
& \underline{\boldsymbol{q}}\xi_i \le \boldsymbol{\zeta}_i \le \overline{\boldsymbol{q}}\xi_i, & -\boldsymbol{\rho}_i \le \boldsymbol{\zeta}_i - \boldsymbol{\mu}\xi_i \le \boldsymbol{\rho}_i \\
& \xi_2 \tau \le \mathbf{1}_S^\top \boldsymbol{\zeta}_2, & \mathbf{S}(\boldsymbol{\rho}_1 + \boldsymbol{\rho}_2) \le \boldsymbol{\nu} \\
& \xi_i \in \mathbb{R}_+,\ \boldsymbol{\zeta}_i \in \mathbb{R}^n,\ \boldsymbol{\rho}_i \in \mathbb{R}^n_+,\ i = 1, 2, &
\end{array}
$$

where $\mathbf{S}$ is a $p \times n$ matrix with rows $\mathbf{1}_{S_i}^\top$, $i = 1, \ldots, p$.

Proof of Proposition EC.4. The proof follows along similar lines as the proof of Proposition EC.1. □

EC.10.5 Covariance Generic Ambiguity Sets

Proposition EC.5. *A sequence of distributions $\mathbb{P}^t \in \mathcal{P}$, $t = 1, 2, \ldots$, that attain the worst-case value-at-risk* $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\sum_{i\in S} \tilde{q}_i\big]$ *in Theorem 7 asymptotically as $t \longrightarrow \infty$ can be defined as $\mathbb{P}^t = (\xi_1 - 1/t) \cdot \delta_{\boldsymbol{q}_1} + (\xi_2 + 1/t) \cdot \delta_{\boldsymbol{q}_2}$, where $\boldsymbol{q}_1 = \frac{\boldsymbol{\zeta}_1}{\xi_1}$ and $\boldsymbol{q}_2 = \frac{\boldsymbol{\zeta}_2}{\xi_2} + \frac{1}{t(\xi_2 + 1/t)}(\frac{\boldsymbol{\zeta}_1}{\xi_1} - \frac{\boldsymbol{\zeta}_2}{\xi_2})$, and $(\xi_i, \boldsymbol{\zeta}_i)$is an optimal solution to the convex optimization problem*

$$
\begin{array}{lll}
\text{minimize} & \xi_1 & \\
\text{subject to} & \xi_1 + \xi_2 = 1, & \boldsymbol{\zeta}_1 + \boldsymbol{\zeta}_2 = \boldsymbol{\mu} \\
& \underline{\boldsymbol{q}}\xi_i \le \boldsymbol{\zeta}_i \le \overline{\boldsymbol{q}}\xi_i, & \mathbf{1}_S^\top \boldsymbol{\zeta}_2 \ge \xi_2 \tau \\
& \multicolumn{2}{l}{\dfrac{1}{\xi_1}\left(\boldsymbol{\zeta}_1 - \xi_1\boldsymbol{\mu}\right)\left(\boldsymbol{\zeta}_1 - \xi_1\boldsymbol{\mu}\right)^\top + \dfrac{1}{\xi_2}\left(\boldsymbol{\zeta}_2 - \xi_2\boldsymbol{\mu}\right)\left(\boldsymbol{\zeta}_2 - \xi_2\boldsymbol{\mu}\right)^\top \preceq \boldsymbol{\Sigma}} \\
& \xi_i \in \mathbb{R}_+,\ \boldsymbol{\zeta}_i \in \mathbb{R}^n,\ i = 1, 2. &
\end{array}
$$

Proof of Proposition EC.5. The proof follows along similar lines as the proof of Proposition EC.1. □

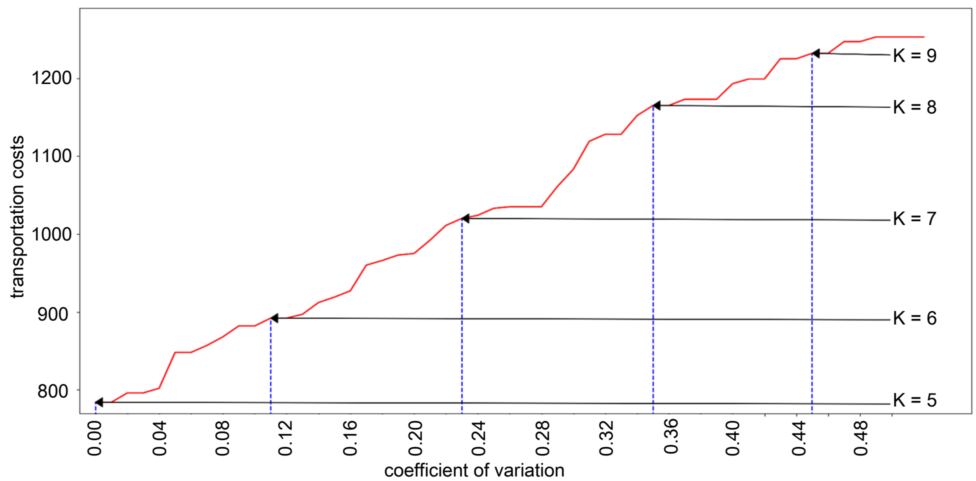

Figure 9. Minimum number of vehicles and optimal transportation costs for the instance A-n32-k5 with a marginalized variance ambiguity set and different variation coefficients ρ.

EC.11 Numerical Results for Marginalized Ambiguity Sets

We solve a distributionally robust version of the benchmark instance A-n32-k5 with a marginalized variance ambiguity set of the form (8). To this end, we identify the expected customer demands $\boldsymbol{\mu}$ with the nominal customer demands from the benchmark instance and set $(\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}) = (0.5\boldsymbol{\mu}, 2\boldsymbol{\mu})$. Moreover, we select $\sigma_i = (\rho\mu_i)^2$, $i \in V_C$, where ρ represents the coefficient of variation, which is assumed to be common for all customer demands. Contrary to the other experiments, we use the same vehicle capacities as in the benchmark instances.

Figure 9 illustrates the minimum number of vehicles required to serve all customers' demands, as well as the resulting transportation costs, as a function of the coefficient of variation ρ. Moreover, Figure 10 shows the optimal route sets corresponding to three different values of ρ. We observe that a higher coefficient of variation ρ in the ambiguity set (8) hedges against larger sets of demand distributions in the distributionally robust CVRP instance, which in turn leads to higher nominal customer demands in the corresponding deterministic CVRP instance. As a result, both the number of vehicles and the transportation costs tend to increase with larger values of ρ.

EC.12 Detailed Numerical Results

Table 1 summarizes the best determined solution as well as the runtime of our branch-and-cut

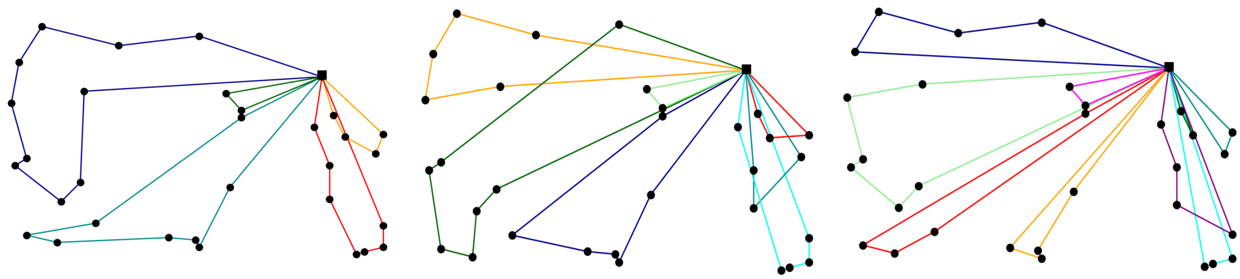

Figure 10. Optimal route sets for the instance A-n32-k5 with a marginalized variance ambiguity set and variation coefficients $\rho = 0$ (left; 5 vehicles), $\rho = 0.23$ (middle; 7 vehicles) and $\rho = 0.45$ (right; 9 vehicles).

scheme for the deterministic CVRP (‘Deterministic’) as well as the distributionally robust CVRP over a first order ambiguity set (‘First Order’), a second order ambiguity set (‘Second Order’) as well as a second order ambiguity set with diagonal covariance bounds (‘Diagonal’). Instances that have been solved to certified optimality (within the runtime limit of 12 hours) are marked with an asterisk: in this case, the ‘Opt’ column denotes the optimal objective value, and the ‘t (sec)’ column provides the runtime. For all other instances, the ‘Opt’ column denotes the objective value of the best route set found, and the ‘[LB]’ column presents the lower bound at termination.

Problem	Deterministic		First Order		Second Order		Diagonal	
	Opt	t (sec) [LB]	Opt	t (sec) [LB]	Opt	t (sec) [LB]	Opt	t (sec) [LB]
A-n32-k5	784.0⋆	0.17	753.0⋆	0.18	756.0⋆	0.24	753.0⋆	0.19
A-n33-k5	661.0⋆	0.4	639.0⋆	0.31	652.0⋆	3.61	639.0⋆	0.62
A-n33-k6	742.0⋆	0.68	716.0⋆	0.91	731.0⋆	34.02	716.0⋆	1.28
A-n34-k5	778.0⋆	0.78	702.0⋆	0.27	724.0⋆	34.21	702.0⋆	0.26
A-n36-k5	799.0⋆	1.82	762.0⋆	5.39	770.0⋆	14.79	762.0⋆	10.57
A-n37-k5	669.0⋆	0.2	656.0⋆	0.62	663.0⋆	1.5	656.0⋆	0.28
A-n37-k6	949.0⋆	35.67	884.0⋆	6.98	902.0⋆	95.44	884.0⋆	12.33
A-n38-k5	730.0⋆	1.67	684.0⋆	2.83	704.0⋆	10.76	684.0⋆	1.22
A-n39-k5	822.0⋆	10.09	767.0⋆	13.11	792.0⋆	409.91	767.0⋆	8.49
A-n39-k6	831.0⋆	2.47	791.0⋆	3.02	800.0⋆	56.15	791.0⋆	2.17
A-n44-k6	937.0⋆	43.99	903.0⋆	241.45	917.0⋆	2088.51	903.0⋆	160.27
A-n45-k6	944.0⋆	25.07	873.0⋆	1.94	903.0⋆	70.57	873.0⋆	2.36
A-n45-k7	1146.0⋆	484.95	1077.0⋆	4510.15	1119.0	[1095.0]	1077.0⋆	2614.56
A-n46-k7	917.0⋆	2.67	887.0⋆	7.28	892.0⋆	11.99	887.0⋆	20.46
A-n48-k7	1073.0⋆	28.28	1027.0⋆	6439.49	1036.0⋆	11257.9	1027.0⋆	27366.7
A-n53-k7	1010.0⋆	27.12	968.0⋆	30.43	980.0⋆	192.44	968.0⋆	19.55
A-n54-k7	1167.0⋆	528.51	1087.0⋆	4709.69	1138.0	[1090.83]	1087.0⋆	12903.8
A-n55-k9	1073.0⋆	18.31	1023.0⋆	57.89	1047.0⋆	1396.58	1023.0⋆	88.19
A-n60-k9	1371.0	[1342.33]	1320.0	[1225.95]	1362.0	[1234.71]	1265.0	[1233.0]
A-n61-k9	1044.0	[1021.4]	958.0⋆	1829.01	998.0	[969.727]	958.0⋆	5888.66
A-n62-k8	1288.0⋆	8480.34	1237.0	[1160.15]	1245.0	[1185.42]	1228.0	[1162.67]
A-n63-k9	1682.0	[1588.17]	1605.0	[1444.61]	1588.0	[1465.27]	1551.0	[1442.24]
A-n63-k10	1326.0	[1293.44]	1231.0	[1212.3]	1254.0	[1219.17]	1233.0	[1203.94]
A-n64-k9	1457.0	[1360.81]	1380.0	[1267.45]	no-feas	—	1421.0	[1271.71]
A-n65-k9	1174.0⋆	1080.06	1098.0⋆	994.57	1193.0	[1100.48]	1098.0⋆	2119.66
A-n69-k9	1159.0	[1143.28]	1094.0	[1091.54]	1143.0	[1085.97]	1094.0⋆	16385.9
A-n80-k10	no-feas	—	1744.0	[1588.17]	1778.0	[1596.94]	1863.0	[1582.24]
B-n31-k5	672.0⋆	0.25	651.0⋆	0.63	652.0⋆	1.35	651.0⋆	0.26
B-n34-k5	788.0⋆	1.32	764.0	[757.8]	769.0⋆	29.69	768.0	[755.5]
B-n35-k5	955.0⋆	0.28	867.0⋆	0.05	888.0⋆	5.74	867.0⋆	0.07
B-n38-k6	805.0⋆	0.3	732.0⋆	0.29	732.0⋆	0.62	732.0⋆	0.25
B-n39-k5	549.0⋆	0.27	521.0⋆	0.19	532.0⋆	0.38	521.0⋆	0.15
B-n41-k6	829.0⋆	1.02	791.0⋆	1.6	797.0⋆	8.05	791.0⋆	0.68
B-n43-k6	742.0⋆	14.58	680.0⋆	4.69	683.0⋆	4.49	680.0⋆	2.88
B-n44-k7	909.0⋆	1.74	841.0⋆	1.78	847.0⋆	21.92	841.0⋆	2.02
B-n45-k5	751.0⋆	3.78	677.0⋆	2.36	702.0⋆	3.33	677.0⋆	2.47
B-n45-k6	678.0⋆	13.82	626.0⋆	2.36	660.0⋆	59.67	626.0⋆	4.19

Table 1. Optimally solved instances are highlighted with an asterisk, and the adjacent entries report the solution times t. For all other instances, we present the best solution found within 12 hours and the lower bound at termination.

We now provide detailed numerical results for each branch-and-cut scheme in turn. To this end, Tables 2–5 provide the percentage gap at the root node (measured relative to the best route set found at termination), the time to process the root node, the number of RCI cuts introduced throughout the execution of our branch-and-cut scheme, the amount of time spent on identifying

Problem	Deterministic		First Order		Second Order		Diagonal	
	Opt	t (sec) [LB]	Opt	t (sec) [LB]	Opt	t (sec) [LB]	Opt	t (sec) [LB]
B-n50-k7	741.0*	0.52	679.0*	0.1	724.0*	2.99	679.0*	0.13
B-n50-k8	1312.0*	1487.59	1233.0	[1209.0]	1231.0*	12567.6	1230.0	[1214.84]
B-n51-k7	1032.0*	6.63	929.0*	3.63	964.0*	17285.8	929.0*	1.66
B-n52-k7	747.0*	0.44	676.0*	0.96	679.0*	2.33	676.0*	0.45
B-n56-k7	707.0*	0.73	623.0*	7.43	624.0*	60.66	623.0*	2.02
B-n57-k9	1598.0*	94.52	1541.0*	15505.3	1580.0	[1547.24]	1541.0*	1788.11
B-n63-k10	1496.0*	3488.23	1434.0	[1368.87]	1587.0	[1379.17]	1419.0	[1362.11]
B-n64-k9	861.0*	141.12	803.0*	6.67	803.0*	3.21	803.0*	2.87
B-n66-k9	1316.0*	6490.27	1212.0*	14206.8	1361.0	[1208.0]	1212.0*	4563.34
B-n67-k10	1032.0*	65.55	980.0*	340.58	1014.0*	2416.87	980.0*	690.97
B-n68-k9	1275.0	[1266.71]	1315.0	[1154.56]	1354.0	[1176.14]	1207.0	[1153.67]
B-n78-k10	1221.0*	3359.2	1132.0	[1105.92]	1148.0	[1129.2]	1123.0	[1104.28]
E-n101-k8	820.0	[806.437]	793.0	[784.607]	806.0	[785.587]	799.0	[783.205]
E-n101-k14	1206.0	[1024.4]	1034.0	[991.635]	no-feas	—	1240.0	[989.23]
E-n22-k4	375.0*	0.03	373.0*	0.02	373.0*	0.02	373.0*	0.02
E-n23-k3	569.0*	0.01	564.0*	0.0	569.0*	0.04	564.0*	0.0
E-n30-k3	534.0*	2.25	492.0*	0.05	495.0*	0.11	492.0*	0.04
E-n33-k4	835.0*	0.41	814.0*	0.45	814.0*	2.17	814.0*	0.53
E-n51-k5	521.0*	0.88	516.0*	36.0	516.0*	15.45	516.0*	34.72
E-n76-k7	682.0*	13066.8	661.0*	1909.01	667.0	[665.143]	661.0*	6614.39
E-n76-k8	737.0	[725.244]	706.0	[700.631]	716.0	[700.12]	706.0	[699.462]
E-n76-k10	851.0	[801.247]	792.0	[765.87]	799.0	[767.762]	790.0	[765.091]
E-n76-k14	no-feas	—	973.0	[912.894]	1004.0	[915.242]	968.0	[911.564]
F-n135-k7	1162.0*	106.59	1086.0*	13961.3	1209.0	[1094.58]	1086.0*	26643.3
F-n45-k4	724.0*	0.18	715.0*	0.59	720.0*	0.76	715.0*	0.5
F-n72-k4	237.0*	11.59	232.0*	0.43	232.0*	1.53	232.0*	0.33
M-n101-k10	820.0*	5.58	804.0*	52.45	809.0*	102.53	804.0*	36.29
M-n121-k7	1041.0	[1017.04]	1065.0	[945.488]	no-feas	—	997.0	[949.634]
M-n151-k12	1120.0	[968.683]	no-feas	—	no-feas	—	1182.0	[935.127]
P-n19-k2	212.0*	0.03	195.0*	0.0	195.0*	0.0	195.0*	0.0
P-n20-k2	216.0*	0.05	208.0*	0.01	209.0*	0.02	208.0*	0.01
P-n21-k2	211.0*	0.03	208.0*	0.01	211.0*	0.02	208.0*	0.01
P-n22-k2	216.0*	0.03	213.0*	0.03	215.0*	0.03	213.0*	0.03
P-n22-k8	604.0*	0.3	559.0*	0.04	593.0*	0.55	559.0*	0.07
P-n23-k8	529.0*	5.62	504.0*	0.92	524.0*	43.7	504.0*	1.25
P-n40-k5	458.0*	0.39	449.0*	0.26	454.0*	4.02	449.0*	0.39
P-n45-k5	510.0*	2.17	500.0*	2.13	501.0*	16.21	500.0*	2.37
P-n50-k7	554.0*	32.13	543.0*	24.85	545.0*	114.94	543.0*	31.51
P-n50-k8	644.0	[620.417]	588.0*	130.92	592.0*	390.6	588.0*	77.8
P-n50-k10	696.0*	7635.86	662.0*	208.73	670.0*	1013.43	662.0*	208.7
P-n51-k10	741.0*	7841.63	695.0*	134.74	714.0*	7168.28	695.0*	84.04

Table 1. (Continued from previous page.)

Problem	Deterministic		First Order		Second Order		Diagonal	
	Opt	t (sec) [LB]	Opt	t (sec) [LB]	Opt	t (sec) [LB]	Opt	t (sec) [LB]
P-n55-k10	696.0	[694.0]	665.0*	237.22	680.0*	23861.2	665.0*	171.93
P-n55-k7	568.0*	339.01	551.0*	50.72	554.0*	152.11	551.0*	55.96
P-n55-k8	594.0*	97.82	575.0*	21.8	584.0*	386.29	575.0*	5.53
P-n55-k15	no-feas	—	886.0*	4922.87	933.0	[885.795]	886.0*	3433.36
P-n60-k10	744.0*	30258.2	712.0*	2678.99	716.0*	9448.37	712.0*	5553.23
P-n60-k15	973.0	[948.558]	926.0*	17824.9	949.0	[920.5]	926.0*	17177.0
P-n65-k10	799.0	[782.652]	761.0*	23003.3	770.0	[761.0]	761.0*	25022.8
P-n70-k10	830.0	[803.776]	789.0	[768.618]	801.0	[768.222]	796.0	[768.104]
P-n76-k4	593.0*	16.13	588.0*	4.89	589.0*	103.87	588.0*	4.14
P-n76-k5	627.0*	570.43	614.0*	1457.71	615.0*	873.37	614.0*	595.71
P-n101-k4	681.0*	9.27	673.0*	3.34	673.0*	8.93	673.0*	3.98
att-n48-k4	40002.0*	3.25	38637.0*	1.73	38966.0*	6.58	38637.0*	1.21

Table 1. (Continued from previous page.)

RCI cuts as well as the number of branch-and-bound nodes created.

Problem	Root gap	Root time (sec)	# of Cuts	Cuts time (sec)	# of B&B nodes
A-n32-k5	8.80%	0.03	223	0.04	45
A-n33-k5	2.53%	0.15	274	0.12	59
A-n33-k6	1.61%	0.42	307	0.16	59
A-n34-k5	2.83%	0.35	358	0.21	74
A-n36-k5	2.07%	0.32	522	0.79	330
A-n37-k5	0.00%	0.18	103	0.02	0
A-n37-k6	4.97%	0.75	2,758	10.79	2,940
A-n38-k5	12.88%	0.05	697	0.62	255
A-n39-k5	17.86%	0.04	1,604	2.41	774
A-n39-k6	4.75%	0.29	665	1.04	351
A-n44-k6	4.12%	0.75	2,898	12.30	2,689
A-n45-k6	5.15%	0.65	2,017	6.91	1,602
A-n45-k7	4.78%	2.40	6,351	63.29	13,686
A-n46-k7	1.06%	0.60	854	0.93	136
A-n48-k7	3.77%	1.57	2,303	8.25	1,245
A-n53-k7	10.80%	0.12	1,822	8.50	1,485
A-n54-k7	5.01%	3.41	5,213	68.83	12,059
A-n55-k9	2.70%	1.49	1,121	8.24	1,774
A-n60-k9	13.68%	1.74	24,243	2,253.15	296,214
A-n61-k9	5.79%	3.24	50,713	647.81	106,151
A-n62-k8	13.52%	1.53	11,191	483.35	76,017
A-n63-k9	17.42%	4.65	37,127	1,391.02	199,160
A-n63-k10	22.40%	0.08	27,591	1,794.41	227,390
A-n64-k9	15.95%	2.74	45,515	972.17	131,772
A-n65-k9	6.08%	3.07	4,982	178.73	18,669
A-n69-k9	6.34%	4.14	23,146	1,712.53	170,185
A-n80-k10	1,611.78%	4.97	55,387	813.87	59,793
B-n31-k5	4.61%	0.10	276	0.04	43
B-n34-k5	6.21%	0.20	513	0.46	295
B-n35-k5	7.51%	0.13	318	0.03	67
B-n38-k6	0.24%	0.23	186	0.05	14
B-n39-k5	3.28%	0.10	202	0.08	32
B-n41-k6	4.34%	0.41	407	0.29	144
B-n43-k6	20.65%	0.09	2,081	4.38	1,167
B-n44-k7	9.35%	0.73	885	0.22	22
B-n45-k5	0.90%	0.63	912	1.27	473
B-n45-k6	3.05%	0.52	1,586	5.29	1,505

Table 2. Detailed numerical results for our deterministic branch-and-cut scheme. Shown is the gap at the root node, the time required to process the root node, the number of RCI cuts introduced throughout the search, the time spent on identifying RCI cuts and the size of the overall branch-and-bound tree (in order of appearance).

Problem	Root gap	Root time (sec)	# of Cuts	Cuts time (sec)	# of B&B nodes
B-n50-k7	4.12%	0.09	423	0.14	73
B-n50-k8	11.90%	2.62	10,186	137.43	20,986
B-n51-k7	6.36%	0.33	1,068	2.41	798
B-n52-k7	4.21%	0.15	287	0.09	33
B-n56-k7	3.49%	0.08	387	0.23	43
B-n57-k9	9.89%	0.10	3,446	29.59	3,903
B-n63-k10	11.43%	5.37	23,103	161.14	16,698
B-n64-k9	6.11%	0.32	5,842	27.82	3,169
B-n66-k9	17.28%	5.51	22,026	268.89	30,531
B-n67-k10	5.97%	1.71	2,692	17.61	2,455
B-n68-k9	11.66%	2.82	25,156	1,488.26	255,735
B-n78-k10	16.79%	2.84	16,680	178.24	15,615
E-n101-k8	4.50%	3.60	30,712	1,108.90	109,437
E-n101-k14	20.17%	0.72	58,691	1,389.14	65,811
E-n22-k4	0.80%	0.03	41	0.00	3
E-n23-k3	0.00%	0.01	12	0.00	0
E-n30-k3	9.59%	0.04	663	0.72	1,319
E-n33-k4	1.22%	0.33	221	0.05	10
E-n51-k5	2.00%	0.38	373	0.25	23
E-n76-k7	4.01%	3.05	13,073	609.01	111,864
E-n76-k8	5.61%	5.77	23,396	1,304.07	169,627
E-n76-k10	9.97%	7.59	46,341	899.08	68,335
E-n76-k14	1,004.04%	2.61	73,328	1,157.97	58,012
F-n135-k7	7.20%	4.28	2,875	9.45	559
F-n45-k4	2.07%	0.05	173	0.04	27
F-n72-k4	1.69%	0.77	1,431	2.89	1,533
M-n101-k10	0.77%	4.90	609	0.51	13
M-n121-k7	9.19%	31.06	15,294	1,171.71	94,427
M-n151-k12	17.09%	20.39	38,922	1,129.98	43,820
P-n19-k2	0.00%	0.03	25	0.00	0
P-n20-k2	1.70%	0.03	64	0.01	33
P-n21-k2	0.24%	0.03	30	0.00	2
P-n22-k2	0.00%	0.03	30	0.00	0
P-n22-k8	4.33%	0.02	355	0.14	67
P-n23-k8	10.42%	0.03	1,747	2.48	471
P-n40-k5	1.52%	0.18	203	0.13	28
P-n45-k5	4.31%	0.04	789	0.93	257
P-n50-k7	3.17%	0.66	2,346	11.15	1,841
P-n50-k8	10.07%	0.71	45,881	771.23	140,653
P-n50-k10	6.24%	1.11	15,289	625.22	76,105
P-n51-k10	6.59%	0.92	20,397	376.89	44,558

Table 2. (Continued from previous page.)

Problem	Root gap	Root time (sec)	# of Cuts	Cuts time (sec)	# of B&B nodes
P-n55-k10	5.74%	1.35	18,663	2,761.96	439,225
P-n55-k7	4.24%	0.53	4,957	67.23	13,381
P-n55-k8	3.99%	0.77	3,306	27.15	4,988
P-n55-k15	986.58%	3.89	85,026	779.67	60,092
P-n60-k10	6.26%	2.89	18,509	1,817.38	220,933
P-n60-k15	11.18%	0.16	29,304	2,736.74	217,385
P-n65-k10	6.44%	3.13	26,086	1,869.01	220,693
P-n70-k10	7.81%	4.15	38,508	1,277.46	100,265
P-n76-k4	1.32%	1.48	1,637	3.30	694
P-n76-k5	3.55%	1.94	5,427	54.18	11,008
P-n101-k4	0.77%	1.87	961	2.65	404
att-n48-k4	2.62%	0.51	547	1.19	951

Table 2. (Continued from previous page.)

Problem	Root gap	Root time (sec)	# of Cuts	Cuts time (sec)	# of B&B nodes
A-n32-k5	0.64%	0.16	87	0.02	6
A-n33-k5	7.08%	0.04	224	0.12	81
A-n33-k6	2.65%	0.17	334	0.36	205
A-n34-k5	2.14%	0.12	185	0.06	44
A-n36-k5	8.19%	0.07	749	1.90	1,613
A-n37-k5	3.66%	0.13	259	0.21	125
A-n37-k6	9.07%	0.17	1,044	2.76	1,288
A-n38-k5	4.24%	0.22	636	1.03	493
A-n39-k5	6.26%	0.64	1,412	3.19	1,460
A-n39-k6	4.63%	0.19	537	1.30	747
A-n44-k6	5.83%	0.64	4,040	35.06	13,726
A-n45-k6	3.79%	0.48	334	0.71	219
A-n45-k7	14.90%	0.20	13,254	304.42	91,182
A-n46-k7	5.88%	0.13	1,026	3.08	940
A-n48-k7	8.81%	0.72	9,168	350.46	133,682
A-n53-k7	8.32%	0.25	1,284	11.10	2,504
A-n54-k7	7.28%	3.11	6,864	395.41	114,838
A-n55-k9	7.54%	0.44	2,129	24.05	4,667
A-n60-k9	19.03%	0.69	41,073	1,200.80	235,201
A-n61-k9	10.10%	0.67	8,835	207.29	29,502
A-n62-k8	18.19%	0.90	33,033	921.39	140,083
A-n63-k9	21.80%	0.66	45,214	1,189.42	230,934
A-n63-k10	13.15%	0.47	19,973	2,248.64	388,767
A-n64-k9	12.52%	10.18	41,546	944.52	133,645
A-n65-k9	8.01%	0.70	5,914	140.58	28,429
A-n69-k9	7.52%	1.13	14,681	2,457.60	407,054
A-n80-k10	19.59%	0.59	38,716	1,037.21	102,557
B-n31-k5	1.96%	0.11	421	0.18	154
B-n34-k5	3.80%	0.15	22,506	830.25	1,006,939
B-n35-k5	0.17%	0.05	36	0.00	5
B-n38-k6	4.17%	0.03	314	0.09	78
B-n39-k5	2.98%	0.06	203	0.03	68
B-n41-k6	1.69%	0.10	634	0.64	309
B-n43-k6	7.79%	0.09	1,007	1.74	595
B-n44-k7	18.61%	0.08	612	0.62	100
B-n45-k5	2.73%	0.27	526	0.97	365
B-n45-k6	2.39%	0.32	660	0.87	383

Table 3. Detailed numerical results for our distributionally robust branch-and-cut scheme over first order ambiguity sets. The columns are the same as in Table 2.

Problem	Root gap	Root time (sec)	# of Cuts	Cuts time (sec)	# of B&B nodes
B-n50-k7	1.18%	0.08	87	0.01	19
B-n50-k8	11.92%	0.81	31,056	1,375.71	393,742
B-n51-k7	1.27%	0.12	695	2.08	667
B-n52-k7	4.73%	0.16	400	0.32	122
B-n56-k7	3.37%	0.41	960	3.14	1,034
B-n57-k9	6.77%	0.33	19,596	1,422.81	242,756
B-n63-k10	17.19%	0.23	36,888	1,991.49	372,514
B-n64-k9	3.13%	0.34	734	3.80	437
B-n66-k9	14.29%	0.74	24,773	515.11	75,099
B-n67-k10	7.92%	0.18	3,435	102.70	14,767
B-n68-k9	22.59%	1.10	37,064	1,503.80	221,916
B-n78-k10	18.37%	1.00	22,453	2,512.12	332,470
E-n101-k8	3.70%	1.10	23,580	1,699.66	120,921
E-n101-k14	8.13%	0.92	34,050	2,525.48	112,334
E-n22-k4	0.00%	0.02	17	0.00	0
E-n23-k3	0.00%	0.00	6	0.00	0
E-n30-k3	0.24%	0.04	39	0.01	2
E-n33-k4	2.18%	0.14	372	0.08	102
E-n51-k5	3.69%	0.42	1,709	9.45	3,565
E-n76-k7	4.39%	0.67	7,776	187.81	28,781
E-n76-k8	5.69%	1.80	14,279	2,435.24	304,396
E-n76-k10	9.18%	0.44	29,876	1,787.28	221,637
E-n76-k14	10.95%	1.42	37,663	2,619.79	195,994
F-n135-k7	9.28%	4.90	11,290	743.93	79,677
F-n45-k4	3.87%	0.05	380	0.10	105
F-n72-k4	0.00%	0.41	69	0.02	0
M-n101-k10	4.73%	2.07	1,319	24.62	2,003
M-n121-k7	21.29%	2.17	25,715	889.68	100,398
M-n151-k12	1,007.57%	17.79	33,566	1,780.50	52,380
P-n19-k2	0.00%	0.00	2	0.00	0
P-n20-k2	0.00%	0.01	8	0.00	0
P-n21-k2	0.00%	0.01	10	0.00	0
P-n22-k2	0.35%	0.03	19	0.00	4
P-n22-k8	0.00%	0.04	34	0.00	0
P-n23-k8	5.65%	0.02	429	0.50	282
P-n40-k5	0.98%	0.15	117	0.07	18
P-n45-k5	3.20%	0.11	493	1.06	397
P-n50-k7	4.71%	0.33	1,427	10.26	3,145
P-n50-k8	4.96%	0.64	2,405	38.69	10,523
P-n50-k10	5.72%	0.44	3,556	64.73	12,038

Table 3. (Continued from previous page.)

Problem	Root gap	Root time (sec)	# of Cuts	Cuts time (sec)	# of B&B nodes
P-n51-k10	5.07%	0.34	2,574	48.55	8,253
P-n55-k10	5.00%	0.63	3,233	75.73	13,805
P-n55-k7	3.31%	0.41	1,936	17.92	4,612
P-n55-k8	3.22%	0.28	1,372	10.28	1,821
P-n55-k15	5.76%	0.59	11,686	820.36	84,568
P-n60-k10	5.80%	1.25	5,997	478.57	77,248
P-n60-k15	5.89%	0.66	9,642	3,647.42	397,186
P-n65-k10	6.92%	0.36	11,478	2,064.94	290,779
P-n70-k10	8.20%	0.78	19,677	3,024.37	400,755
P-n76-k4	2.64%	0.41	669	2.07	422
P-n76-k5	3.89%	0.62	7,598	108.29	25,154
P-n101-k4	0.78%	0.62	312	1.10	123
att-n48-k4	1.54%	0.36	379	0.54	394

Table 3. (Continued from previous page.)

Problem	Root gap	Root time (sec)	# of Cuts	Cuts time (sec)	# of B&B nodes
A-n32-k5	0.88%	0.17	100	0.07	3
A-n33-k5	9.27%	0.05	366	2.96	472
A-n33-k6	5.24%	0.22	968	26.66	4,342
A-n34-k5	4.76%	0.23	1,232	25.71	4,234
A-n36-k5	6.30%	0.21	659	11.97	1,472
A-n37-k5	4.90%	0.10	243	1.15	70
A-n37-k6	9.13%	0.25	1,963	67.19	7,016
A-n38-k5	4.81%	0.28	657	8.75	933
A-n39-k5	10.41%	0.31	3,884	167.48	22,474
A-n39-k6	5.32%	0.24	1,725	43.72	3,427
A-n44-k6	7.94%	0.47	6,821	907.55	58,357
A-n45-k6	9.52%	0.11	1,465	55.35	3,617
A-n45-k7	14.54%	0.29	30,388	7,769.22	454,813
A-n46-k7	4.61%	0.41	565	10.32	403
A-n48-k7	11.31%	0.50	9,292	3,841.96	148,705
A-n53-k7	7.41%	0.45	1,530	160.80	5,100
A-n54-k7	15.64%	0.50	32,890	8,621.02	273,935
A-n55-k9	7.56%	1.30	5,775	890.82	30,153
A-n60-k9	24.01%	0.14	31,016	11,954.02	228,278
A-n61-k9	12.84%	0.66	19,059	13,276.38	304,638
A-n62-k8	19.74%	1.05	30,499	6,682.17	196,778
A-n63-k9	21.39%	0.92	36,943	9,421.22	226,004
A-n63-k10	20.45%	0.07	30,239	10,365.47	216,149
A-n64-k9	1,236.57%	0.34	51,599	5,620.44	130,836
A-n65-k9	16.33%	0.91	32,861	11,697.69	316,333
A-n69-k9	11.88%	2.03	26,141	12,435.64	244,488
A-n80-k10	19.83%	1.71	42,743	10,253.69	134,555
B-n31-k5	1.76%	0.13	437	0.88	250
B-n34-k5	8.19%	0.13	1,200	23.28	3,333
B-n35-k5	1.34%	0.09	870	4.16	783
B-n38-k6	4.17%	0.02	317	0.41	80
B-n39-k5	3.99%	0.07	275	0.14	84
B-n41-k6	2.22%	0.19	662	6.36	1,024
B-n43-k6	8.20%	0.06	806	3.12	376
B-n44-k7	19.19%	0.07	1,149	17.78	1,121
B-n45-k5	7.95%	0.09	308	2.74	247
B-n45-k6	2.70%	0.13	1,273	47.87	4,590

Table 4. Detailed numerical results for our distributionally robust branch-and-cut scheme over second order ambiguity sets. The columns are the same as in Table 2.

Problem	Root gap	Root time (sec)	# of Cuts	Cuts time (sec)	# of B&B nodes
B-n50-k7	1.75%	0.09	593	2.16	241
B-n50-k8	13.48%	0.13	16,809	3,840.93	116,174
B-n51-k7	3.30%	0.32	14,994	4,899.57	404,450
B-n52-k7	2.36%	0.24	400	1.69	128
B-n56-k7	3.89%	0.42	1,629	51.74	1,720
B-n57-k9	7.31%	0.40	27,203	13,061.37	459,242
B-n63-k10	23.09%	0.60	43,961	10,941.47	272,627
B-n64-k9	3.55%	0.52	266	2.48	24
B-n66-k9	24.98%	0.63	56,503	6,127.27	158,376
B-n67-k10	10.89%	0.18	7,345	1,655.82	31,546
B-n68-k9	22.17%	0.69	37,478	14,580.46	293,702
B-n78-k10	17.13%	2.99	30,486	14,825.17	139,882
E-n101-k8	5.06%	1.54	29,045	14,248.27	121,826
E-n101-k14	1,054.00%	0.91	36,531	19,669.39	73,112
E-n22-k4	0.00%	0.02	18	0.00	0
E-n23-k3	0.00%	0.02	10	0.02	0
E-n30-k3	0.13%	0.08	47	0.03	3
E-n33-k4	2.58%	0.16	534	1.43	235
E-n51-k5	3.91%	0.29	814	12.59	942
E-n76-k7	5.10%	0.57	14,438	15,433.94	284,828
E-n76-k8	8.00%	0.34	16,598	19,485.62	271,932
E-n76-k10	9.48%	1.28	30,125	14,273.25	136,172
E-n76-k14	15.29%	0.44	47,456	15,796.83	118,497
F-n135-k7	16.10%	9.14	28,690	14,498.58	55,654
F-n45-k4	4.54%	0.05	204	0.51	69
F-n72-k4	0.86%	0.39	140	1.11	24
M-n101-k10	4.46%	1.57	958	94.88	478
M-n121-k7	950.33%	3.42	20,421	14,041.34	69,004
M-n151-k12	998.08%	3.98	19,697	26,267.34	40,896
P-n19-k2	0.00%	0.00	2	0.00	0
P-n20-k2	0.00%	0.02	10	0.00	0
P-n21-k2	0.95%	0.01	21	0.01	5
P-n22-k2	1.55%	0.02	31	0.01	9
P-n22-k8	2.56%	0.03	277	0.43	95
P-n23-k8	9.26%	0.02	1,956	33.25	5,106
P-n40-k5	1.55%	0.24	355	3.40	297
P-n45-k5	2.59%	0.04	660	14.21	1,292
P-n50-k7	5.31%	0.45	1,853	93.86	3,666
P-n50-k8	7.70%	0.19	2,732	306.98	9,854
P-n50-k10	7.91%	0.28	3,907	785.45	18,990

Table 4. (Continued from previous page.)

Problem	Root gap	Root time (sec)	# of Cuts	Cuts time (sec)	# of B&B nodes
P-n51-k10	7.63%	0.36	6,623	4,141.96	132,791
P-n55-k10	7.76%	0.40	12,911	10,772.11	261,138
P-n55-k7	3.92%	0.44	1,937	124.32	4,649
P-n55-k8	5.06%	0.32	2,768	297.57	9,981
P-n55-k15	10.68%	0.40	36,951	18,004.99	232,056
P-n60-k10	7.61%	0.44	5,653	6,506.57	123,579
P-n60-k15	7.73%	1.07	16,838	28,090.32	214,042
P-n65-k10	7.86%	0.67	10,710	24,639.16	294,987
P-n70-k10	9.64%	1.35	20,694	18,413.52	245,107
P-n76-k4	2.80%	0.36	1,549	80.69	2,438
P-n76-k5	3.60%	0.68	4,342	567.82	12,544
P-n101-k4	0.79%	0.80	206	7.50	107
att-n48-k4	1.65%	0.45	476	4.72	560

Table 4. (Continued from previous page.)

Problem	Root gap	Root time (sec)	# of Cuts	Cuts time (sec)	# of B&B nodes
A-n32-k5	0.59%	0.18	59	0.01	4
A-n33-k5	7.08%	0.05	298	0.29	220
A-n33-k6	3.14%	0.19	403	0.51	391
A-n34-k5	3.15%	0.18	130	0.04	13
A-n36-k5	8.14%	0.06	1,189	3.45	2,421
A-n37-k5	3.09%	0.14	94	0.08	38
A-n37-k6	8.19%	0.16	1,201	4.77	2,657
A-n38-k5	2.34%	0.35	388	0.39	204
A-n39-k5	6.52%	0.33	1,057	2.28	1,218
A-n39-k6	4.96%	0.14	478	0.99	736
A-n44-k6	5.98%	0.55	3,590	27.83	11,279
A-n45-k6	4.19%	0.33	588	0.89	259
A-n45-k7	12.33%	0.40	8,480	253.71	74,868
A-n46-k7	5.88%	0.16	1,968	6.61	1,723
A-n48-k7	11.19%	0.25	13,612	1,005.50	412,658
A-n53-k7	7.15%	0.45	1,070	8.35	2,040
A-n54-k7	11.98%	0.46	10,251	831.14	233,414
A-n55-k9	7.93%	0.35	2,052	40.24	7,419
A-n60-k9	15.69%	0.55	24,807	2,156.01	308,711
A-n61-k9	10.65%	0.34	11,494	651.52	100,321
A-n62-k8	17.26%	0.35	35,582	1,151.97	282,771
A-n63-k9	16.00%	4.32	45,954	1,204.07	196,922
A-n63-k10	12.26%	0.65	21,539	2,570.20	368,590
A-n64-k9	19.04%	0.42	47,264	1,287.94	218,242
A-n65-k9	8.25%	1.61	9,073	229.99	41,152
A-n69-k9	7.80%	0.78	12,260	1,445.01	189,889
A-n80-k10	22.56%	4.63	45,429	1,287.34	147,232
B-n31-k5	1.91%	0.10	227	0.05	68
B-n34-k5	8.79%	0.08	18,094	1,625.71	1,711,050
B-n35-k5	0.00%	0.07	37	0.00	2
B-n38-k6	4.17%	0.04	195	0.09	55
B-n39-k5	2.98%	0.07	139	0.02	31
B-n41-k6	0.54%	0.40	181	0.17	68
B-n43-k6	7.79%	0.08	576	1.10	504
B-n44-k7	18.61%	0.08	644	0.67	154
B-n45-k5	4.13%	0.23	468	1.14	495
B-n45-k6	2.65%	0.34	965	1.49	634

Table 5. Detailed numerical results for our distributionally robust branch-and-cut scheme over second order ambiguity sets with diagonal covariance bounds. The columns have the same interpretation as in Table 2.

Problem	Root gap	Root time (sec)	# of Cuts	Cuts time (sec)	# of B&B nodes
B-n50-k7	1.18%	0.08	102	0.02	24
B-n50-k8	11.13%	0.81	25,668	1,574.21	368,764
B-n51-k7	1.27%	0.12	430	0.91	263
B-n52-k7	4.73%	0.13	307	0.09	78
B-n56-k7	3.49%	0.46	465	0.73	260
B-n57-k9	6.55%	0.71	9,297	288.68	39,564
B-n63-k10	16.31%	0.24	32,541	3,134.08	518,803
B-n64-k9	2.88%	0.45	730	0.97	180
B-n66-k9	14.60%	0.63	12,257	330.91	60,197
B-n67-k10	7.92%	0.21	4,569	164.72	24,308
B-n68-k9	15.63%	0.67	39,837	2,391.10	407,652
B-n78-k10	17.12%	1.12	17,842	3,132.44	323,583
E-n101-k8	4.06%	2.07	22,403	2,245.80	168,540
E-n101-k14	23.35%	1.27	33,049	3,777.12	194,335
E-n22-k4	0.00%	0.02	18	0.00	0
E-n23-k3	0.00%	0.00	6	0.00	0
E-n30-k3	1.07%	0.04	44	0.00	2
E-n33-k4	2.21%	0.13	294	0.12	127
E-n51-k5	3.49%	0.46	1,749	9.24	4,370
E-n76-k7	4.53%	0.63	11,179	580.14	78,018
E-n76-k8	5.92%	1.37	11,679	3,486.02	243,887
E-n76-k10	8.92%	0.49	23,723	2,574.44	252,486
E-n76-k14	11.90%	0.58	31,041	3,238.67	199,176
F-n135-k7	9.53%	4.29	16,108	1,917.77	123,567
F-n45-k4	3.87%	0.04	313	0.12	134
F-n72-k4	0.81%	0.31	92	0.02	2
M-n101-k10	5.76%	0.73	1,257	18.00	1,501
M-n121-k7	15.67%	2.86	26,325	993.00	67,071
M-n151-k12	23.35%	20.86	30,041	3,050.43	88,252
P-n19-k2	0.00%	0.00	2	0.00	0
P-n20-k2	0.00%	0.01	8	0.00	0
P-n21-k2	0.00%	0.01	11	0.00	0
P-n22-k2	0.63%	0.03	26	0.00	7
P-n22-k8	0.00%	0.07	36	0.00	0
P-n23-k8	5.65%	0.02	657	0.63	458
P-n40-k5	1.11%	0.14	162	0.15	66
P-n45-k5	3.20%	0.11	471	1.18	582
P-n50-k7	4.10%	0.55	1,380	13.13	4,200
P-n50-k8	4.74%	0.56	2,438	27.12	5,894
P-n50-k10	6.06%	0.28	2,890	68.16	14,987

Table 5. (Continued from previous page.)

Problem	Root gap	Root time (sec)	# of Cuts	Cuts time (sec)	# of B&B nodes
P-n51-k10	5.47%	0.21	2,745	33.21	6,673
P-n55-k10	5.40%	0.43	3,688	55.22	11,508
P-n55-k7	4.13%	0.24	1,732	21.91	5,742
P-n55-k8	3.31%	0.28	743	3.08	781
P-n55-k15	5.59%	0.37	8,309	955.02	95,393
P-n60-k10	6.48%	0.63	7,364	828.66	136,069
P-n60-k15	6.74%	0.33	9,798	4,132.93	462,037
P-n65-k10	5.80%	0.61	11,542	2,407.29	360,166
P-n70-k10	9.17%	0.44	22,465	3,135.99	345,192
P-n76-k4	2.64%	0.40	618	1.62	341
P-n76-k5	3.21%	0.65	4,857	74.25	23,370
P-n101-k4	0.82%	0.96	368	1.36	180
att-n48-k4	1.83%	0.33	320	0.33	234

Table 5. (Continued from previous page.)

EC.13 Proofs

Proof of Theorem 1. For the first statement, assume that the route set $\mathbf{R}$ is feasible in RVRP($\mathcal{P}$). We need to show that $\boldsymbol{x}$ defined through (3) satisfies the constraints of 2VF($\mathcal{P}$) and attains the same transportation costs. One readily verifies that $\boldsymbol{x}$ satisfies the binarity and the degree constraints of 2VF($\mathcal{P}$). In view of the RCI constraints, we note that for any $S \subseteq V_C$, $S \neq \emptyset$, we have

$$\begin{aligned} d_{\mathcal{P}}(S) \; &= \; d_{\mathcal{P}}\Big(\bigcup_{k \in K} [\boldsymbol{R}_k \cap S]\Big) \; \leq \; \sum_{\substack{k \in K: \\ \boldsymbol{R}_k \cap S \neq \emptyset}} d_{\mathcal{P}}(\boldsymbol{R}_k \cap S) \; \leq \; \sum_{\substack{k \in K: \\ \boldsymbol{R}_k \cap S \neq \emptyset}} d_{\mathcal{P}}(\boldsymbol{R}_k) \\ &= \; |k \in K \, : \, \boldsymbol{R}_k \cap S \neq \emptyset| \; \leq \; \sum_{i \in V \setminus S} \sum_{j \in S} x_{ij}(\boldsymbol{R}), \end{aligned}$$

where the first identity follows from the fact that $\mathbf{R} \in \mathfrak{P}(V_C, m)$ and thus $\bigcup_k \boldsymbol{R}_k = V_C$, the first inequality holds because $d_{\mathcal{P}}$ is subadditive, and the second inequality is due to the fact that $\boldsymbol{R}_k \cap S \subseteq \boldsymbol{R}_k$ and $\tilde{\boldsymbol{q}} \geq \mathbf{0}$ $\mathbb{P}$-a.s. for all $\mathbb{P} \in \mathcal{P}$, which in turn implies that $d_{\mathcal{P}}(S) \leq d_{\mathcal{P}}(T)$ for all $S \subseteq T \subseteq V_C$. The second equality holds since $\mathbb{P}\left[\boldsymbol{R}_k \in \mathcal{R}(\tilde{\boldsymbol{q}})\right] \geq 1 - \epsilon$ for all $\mathbb{P} \in \mathcal{P}$ implies that $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\sum_{i \in \boldsymbol{R}_k} \tilde{q}_i\big] \leq Q$ and hence $d_{\mathcal{P}}(\boldsymbol{R}_k) = 1$. In view of the last inequality, let $j_k \in \boldsymbol{R}_k \cap S$ be the first customer on the route $\boldsymbol{R}_k$ that is contained in S, where $k \in K$ satisfies $\boldsymbol{R}_k \cap S \neq \emptyset$. By the feasibility of $\mathbf{R}$ and the definition of j_k, we have $\sum_{i \in V \setminus S} x_{ij_k}(\boldsymbol{R}) = 1$. The inequality now follows from the fact that there are $|k \in K \, : \, \boldsymbol{R}_k \cap S \neq \emptyset|$ different customer nodes j_k with this property.[2] We thus conclude that $\boldsymbol{x}$ also satisfies the RCI constraints of 2VF($\mathcal{P}$). Moreover, equation (3) implies that the transportation costs of $\boldsymbol{x}$ and $\mathbf{R}$ coincide.

For the second statement, we fix a feasible solution $\boldsymbol{x} \in$ 2VF($\mathcal{P}$) and construct a route set $\mathbf{R}$ satisfying (3) as follows. Since $\sum_{j \in V_C} x_{0j} = m$, there are $j_1, \ldots, j_m \in V_C$, $j_1 < \ldots < j_m$, such that $x_{0,j_1} = \ldots = x_{0,j_m} = 1$. For each route $\boldsymbol{R}_k$, $k \in K$, we set $R_{k,1} \leftarrow j_k$ and $n_k \leftarrow 1$. Since $\sum_{j \in V} x_{R_{k,n_k},j} = 1$, we either have $x_{R_{k,n_k},j} = 1$ for some $j \in V_C$ or $x_{R_{k,n_k},0} = 1$. In the former case, we extend route k by the customer $R_{k,n_k+1} \leftarrow j$, we set $n_k \leftarrow n_k + 1$ and we continue the procedure with customer j. In the latter case, we have completed the route $\boldsymbol{R}_k$. By construction, the resulting route set $\mathbf{R}$ satisfies (3). We now show that $\mathbf{R}$ is feasible in RVRP($\mathcal{P}$).

To see that $\mathbf{R} \in \mathfrak{P}(V_C, m)$, we first observe that $\boldsymbol{R}_k \neq \emptyset$ due to the existence of the customers $j_1, \ldots, j_m$. Moreover, the degree constraints in 2VF($\mathcal{P}$) ensure that $\boldsymbol{R}_k \cap \boldsymbol{R}_l = \emptyset$ for all $k \neq l$. It

[2] Note that the same vehicle may enter and leave the customer set S several times, which implies that we cannot strengthen the inequality to an equality in general.

remains to be shown that $\bigcup_k \boldsymbol{R}_k = V_C$. Imagine, to the contrary, that there is a customer $j \in V_C$ such that $j \notin \bigcup_k \boldsymbol{R}_k$. By construction of the above algorithm, j must lie on a short cycle $S \subset V_C$ that is not connected to the depot node 0. Since $d_{\mathcal{P}}(S) \geq 1$ but $\sum_{i \in V \setminus S} \sum_{j \in S} x_{ij} = 0$, the RCI constraint corresponding to the customer set S is violated. We thus conclude that $\boldsymbol{x}$ cannot be feasible in 2VF($\mathcal{P}$), which is a contradiction.

We now show that $\mathbb{P}\left[\boldsymbol{R}_k \in \mathcal{R}(\tilde{\boldsymbol{q}})\right] \geq 1 - \epsilon$ for all $\mathbb{P} \in \mathcal{P}$ and $k \in K$. By construction of the route set $\mathbf{R}$ and the feasibility of $\boldsymbol{x}$ in 2VF($\mathcal{P}$), we have $\sum_{i \in V \setminus \boldsymbol{R}_k} \sum_{j \in \boldsymbol{R}_k} x_{ij} = 1 \geq d_{\mathcal{P}}(\boldsymbol{R}_k)$ for all $k \in K$, and the definition of $d_{\mathcal{P}}$ then implies that $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\sum_{i \in \boldsymbol{R}_k} \tilde{q}_i\big] \leq Q$ and thus $\mathbb{P}\left[\boldsymbol{R}_k \in \mathcal{R}(\tilde{\boldsymbol{q}})\right] \geq 1 - \epsilon$ for all $\mathbb{P} \in \mathcal{P}$.

Finally, imagine that two route sets $\mathbf{R}$ and $\mathbf{R}'$ satisfy (3), and that there is no reordering of the routes in $\mathbf{R}'$ that yields $\mathbf{R}$. Then there must be a customer pair $(i, j) \in V_C \times V_C$ such that (i, j) is visited by the same vehicle in immediate succession in $\mathbf{R}$ but not in $\mathbf{R}'$. This, however, violates the assumption that both $\mathbf{R}$ and $\mathbf{R}'$ satisfy (3), as x_{ij} would have to be both 0 and 1 in that case. We thus conclude that the route set $\mathbf{R}$ satisfying (3) is indeed unique up to a reordering of the individual routes $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m$. □

The proof of Theorem 2 relies on the following auxiliary result, which we prove first.

Lemma 1 (Strong Duality)**.** *Let $\mathcal{Q} = [\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}]$, $f : \mathbb{R}^n \mapsto \mathbb{R}$ be an arbitrary function and $\boldsymbol{\varphi} : \mathbb{R}^n \mapsto \mathbb{R}^p$ be continuous. Assume that $\boldsymbol{\mu} \in \operatorname{int} \mathcal{Q}$ and that $\boldsymbol{\varphi}(\boldsymbol{\mu}) < \boldsymbol{\sigma}$. Then, strong duality holds between the primal moment problem*

$$\begin{array}{ll} \text{minimize} & \displaystyle\int_{\mathcal{Q}} f(\boldsymbol{q})\, \mathbb{P}(\mathrm{d}\boldsymbol{q}) \\ \text{subject to} & \displaystyle\int_{\mathcal{Q}} \mathbb{P}(\mathrm{d}\boldsymbol{q}) = 1 \\ & \displaystyle\int_{\mathcal{Q}} \boldsymbol{q}\, \mathbb{P}(\mathrm{d}\boldsymbol{q}) = \boldsymbol{\mu} \\ & \displaystyle\int_{\mathcal{Q}} \boldsymbol{\varphi}(\boldsymbol{q})\, \mathbb{P}(\mathrm{d}\boldsymbol{q}) \leq \boldsymbol{\sigma} \\ & \mathbb{P} \in \mathcal{M}_+(\mathcal{Q}) \end{array}$$

and its semi-infinite dual problem

$$\begin{array}{lll} \text{maximize} & \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\sigma}^\top \boldsymbol{\gamma} & \\ \text{subject to} & \alpha + \boldsymbol{q}^\top \boldsymbol{\beta} - \boldsymbol{\varphi}(\boldsymbol{q})^\top \boldsymbol{\gamma} \leq f(\boldsymbol{q}) & \forall \boldsymbol{q} \in \mathcal{Q} \\ & \alpha \in \mathbb{R}, \;\; \boldsymbol{\beta} \in \mathbb{R}^n, \;\; \boldsymbol{\gamma} \in \mathbb{R}^p_+. & \end{array}$$

Proof of Lemma 1. The result follows from Proposition 3.4 in Shapiro (2001) if we can show that the point $(1, \boldsymbol{\mu}, \boldsymbol{\sigma})$ resides in the interior of the convex cone

$$\mathcal{V} = \left\{ (a, \boldsymbol{b}, \boldsymbol{c}) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^p \,:\, \exists \mu \in \mathcal{M}_+(\mathcal{Q}) \text{ such that } \begin{array}{l} \int \mu(\mathrm{d}\boldsymbol{\xi}) = a, \\ \int \boldsymbol{q}\, \mu(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{b}, \\ \int \boldsymbol{\varphi}(\boldsymbol{q})\, \mu(\mathrm{d}\boldsymbol{\xi}) \leq \boldsymbol{c} \end{array} \right\}.$$

In the following, we denote by $\mathbb{B}_\rho(\boldsymbol{x})$ the closed Euclidean ball of radius $\rho > 0$ that is centered at $\boldsymbol{x}$. We prove the statement by showing that any point $(s, \boldsymbol{m}, \boldsymbol{s}) \in \mathbb{B}_\kappa(1) \times \mathbb{B}_\kappa(\boldsymbol{\mu}) \times \mathbb{B}_\kappa(\boldsymbol{\sigma})$, where $\kappa > 0$ is sufficiently small, is contained in $\mathcal{V}$. Indeed, assume that κ is small enough so that $\boldsymbol{m}/s \in \mathcal{Q}$ and $\boldsymbol{\varphi}(\boldsymbol{m}/s) \leq \boldsymbol{s}/s$. This is possible since $\boldsymbol{\mu} \in \operatorname{int} \mathcal{Q}$, $\boldsymbol{\varphi}(\boldsymbol{\mu}) < \boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$ is continuous. We then have that the scaled Dirac measure $s \cdot \delta_{\boldsymbol{m}/s}$ satisfies $s \cdot \delta_{\boldsymbol{m}/s} \in \mathcal{M}_+(\mathcal{Q})$, $\int s \cdot \delta_{\boldsymbol{m}/s} = s$, $\int \boldsymbol{q}\, s \cdot \delta_{\boldsymbol{m}/s} = \boldsymbol{m}$ as well as $\int \boldsymbol{\varphi}(\boldsymbol{q})\, s \cdot \delta_{\boldsymbol{m}/s} = s \cdot \boldsymbol{\varphi}(\boldsymbol{m}/s) \leq \boldsymbol{s}$. We thus conclude that $(s, \boldsymbol{m}, \boldsymbol{s}) \in \mathcal{V}$ as desired. $\square$

Proof of Theorem 2. We claim that the epigraph of the worst-case value-at-risk,

$$M \;=\; \left\{ (\boldsymbol{\lambda}, \tau) \in \mathbb{R}^n \times \mathbb{R} \,:\, \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\boldsymbol{\lambda}^\top \tilde{\boldsymbol{q}}\big] \leq \tau \right\}, \tag{19}$$

is convex for moment ambiguity sets of the form (4). We then have

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\Big[\sum_{i \in S} \tilde{q}_i\Big] \;=\; n \cdot \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\Big[\frac{1}{n} \sum_{i \in S} \tilde{q}_i\Big] \;\leq\; \sum_{i \in S} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big],$$

where the identity follows from the positive homogeneity of the value-at-risk (which carries over to the worst-case value-at-risk), and the inequality follows from the stated convexity of the epigraph of the worst-case value-at-risk (Rockafellar, 1970, Theorem 4.2).

We now show that the epigraph (19) is indeed convex for moment ambiguity sets. To this end, we note that $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\boldsymbol{\lambda}^\top \tilde{\boldsymbol{q}}\big] \leq \tau$ if and only if the optimal value of the moment problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\int_{\mathcal{Q}} \mathbb{I}_{[\boldsymbol{\lambda}^\top \boldsymbol{q} \leq \tau]}\, \mathbb{P}(\mathrm{d}\boldsymbol{q}) \\ \text{subject to} & \displaystyle\int_{\mathcal{Q}} \mathbb{P}(\mathrm{d}\boldsymbol{q}) = 1 \\ & \displaystyle\int_{\mathcal{Q}} \boldsymbol{q}\, \mathbb{P}(\mathrm{d}\boldsymbol{q}) = \boldsymbol{\mu} \\ & \displaystyle\int_{\mathcal{Q}} \boldsymbol{\varphi}(\boldsymbol{q})\, \mathbb{P}(\mathrm{d}\boldsymbol{q}) \leq \boldsymbol{\sigma} \\ & \mathbb{P} \in \mathcal{M}_+(\mathcal{Q}) \end{array}$$

is greater than or equal to $1-\epsilon$. By Lemma 1, this is the case if and only if the optimal objective value of the semi-infinite dual problem

$$\begin{array}{ll} \text{maximize} & \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\sigma}^\top \boldsymbol{\gamma} \\ \text{subject to} & \alpha + \boldsymbol{q}^\top \boldsymbol{\beta} - \boldsymbol{\varphi}(\boldsymbol{q})^\top \boldsymbol{\gamma} \leq \mathbb{I}_{[\boldsymbol{\lambda}^\top \boldsymbol{q} \leq \tau]} \qquad \forall \boldsymbol{q} \in \mathcal{Q} \\ & \alpha \in \mathbb{R}, \;\; \boldsymbol{\beta} \in \mathbb{R}^n, \;\; \boldsymbol{\gamma} \in \mathbb{R}^p_+ \end{array}$$

is greater than or equal to $1-\epsilon$. By splitting up the semi-infinite constraint, we obtain

$$\begin{array}{lll} \text{maximize} & \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\sigma}^\top \boldsymbol{\gamma} & \\ \text{subject to} & \alpha + \boldsymbol{q}^\top \boldsymbol{\beta} - \boldsymbol{\varphi}(\boldsymbol{q})^\top \boldsymbol{\gamma} \leq 1 & \forall \boldsymbol{q} \in \mathcal{Q} \\ & \alpha + \boldsymbol{q}^\top \boldsymbol{\beta} - \boldsymbol{\varphi}(\boldsymbol{q})^\top \boldsymbol{\gamma} \leq 0 & \forall \boldsymbol{q} \in \mathcal{Q} \,:\, \boldsymbol{\lambda}^\top \boldsymbol{q} > \tau \\ & \alpha \in \mathbb{R}, \;\; \boldsymbol{\beta} \in \mathbb{R}^n, \;\; \boldsymbol{\gamma} \in \mathbb{R}^p_+. & \end{array} \tag{20}$$

We first assume that $\tau \neq [\boldsymbol{\lambda}]_+^\top \overline{\boldsymbol{q}} - [-\boldsymbol{\lambda}]_+^\top \underline{\boldsymbol{q}}$. In that case, we have $\{\boldsymbol{q} \in \mathcal{Q} \,:\, \boldsymbol{\lambda}^\top \boldsymbol{q} > \tau\} \neq \emptyset$ if and only if $\{\boldsymbol{q} \in \mathcal{Q} \,:\, \boldsymbol{\lambda}^\top \boldsymbol{q} \geq \tau\} \neq \emptyset$, and we can replace the strict inequality in the parameterization of the second constraint with a weak one due to the convexity (and, *a fortiori*, continuity) of $\boldsymbol{\varphi}$.

The first constraint in (20) is satisfied if and only if

$$\left[\begin{array}{ll} \text{maximize} & \alpha + \boldsymbol{q}^\top \boldsymbol{\beta} - \boldsymbol{\varphi}(\boldsymbol{q})^\top \boldsymbol{\gamma} \\ \text{subject to} & \boldsymbol{q} \in \mathcal{Q} \end{array}\right] \leq 1 \quad \Longleftrightarrow \quad \left[\begin{array}{ll} \text{minimize} & -\boldsymbol{q}^\top \boldsymbol{\beta} + \boldsymbol{\varphi}(\boldsymbol{q})^\top \boldsymbol{\gamma} \\ \text{subject to} & \boldsymbol{q} \in [\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}] \end{array}\right] \geq \alpha - 1.$$

Strong convex duality, which holds since the support $\mathcal{Q}$ has a nonempty interior, implies that this is the case if and only if the optimal value of the dual problem,

$$\begin{array}{ll} \text{maximize} & \underline{\boldsymbol{q}}^\top \underline{\boldsymbol{\nu}}_1 - \overline{\boldsymbol{q}}^\top \overline{\boldsymbol{\nu}}_1 - \displaystyle\sum_{i=1}^p \gamma_i \varphi_i^\star(\boldsymbol{\phi}_i / \gamma_i) \\ \text{subject to} & \displaystyle\sum_{i=1}^p \boldsymbol{\phi}_{1i} = \boldsymbol{\beta} + \underline{\boldsymbol{\nu}}_1 - \overline{\boldsymbol{\nu}}_1 \\ & \underline{\boldsymbol{\nu}}_1, \overline{\boldsymbol{\nu}}_1 \in \mathbb{R}^n_+, \;\; \boldsymbol{\phi}_{1i} \in \mathbb{R}^n, \, i = 1, \ldots, p, \end{array}$$

is greater than or equal to $\alpha - 1$. Here, $\varphi_i^\star$ is the conjugate function of φ_i.

The second constraint in (20) is satisfied if and only if

$$\left[\begin{array}{ll} \text{maximize} & \alpha + \boldsymbol{q}^\top \boldsymbol{\beta} - \boldsymbol{\varphi}(\boldsymbol{q})^\top \boldsymbol{\gamma} \\ \text{subject to} & \boldsymbol{q} \in \mathcal{Q} \\ & \boldsymbol{\lambda}^\top \boldsymbol{q} \geq \tau \end{array}\right] \leq 0 \quad \Longleftrightarrow \quad \left[\begin{array}{ll} \text{minimize} & -\boldsymbol{q}^\top \boldsymbol{\beta} + \boldsymbol{\varphi}(\boldsymbol{q})^\top \boldsymbol{\gamma} \\ \text{subject to} & \boldsymbol{q} \in [\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}] \\ & \boldsymbol{\lambda}^\top \boldsymbol{q} \geq \tau \end{array}\right] \geq \alpha. \tag{21}$$

In the following, we distinguish three mutually exclusive and collectively exhaustive cases: *(i)* there is a Slater point $\boldsymbol{q} \in \operatorname{int} \mathcal{Q}$ satisfying $\boldsymbol{\lambda}^\top \boldsymbol{q} > \tau$; *(ii)* there is no $\boldsymbol{q} \in \mathcal{Q}$ that satisfies $\boldsymbol{\lambda}^\top \boldsymbol{q} \geq \tau$; and *(iii)* there are $\boldsymbol{q} \in \mathcal{Q}$ that satisfy $\boldsymbol{\lambda}^\top \boldsymbol{q} \geq \tau$, but none of them satisfies $\boldsymbol{q} \in \operatorname{int} \mathcal{Q}$ and $\boldsymbol{\lambda}^\top \boldsymbol{q} > \tau$. In the first case, strong convex duality holds, and (21) is satisfied if and only if the optimal value of the dual problem,

$$\begin{array}{ll} \text{maximize} & \underline{\boldsymbol{q}}^\top \underline{\boldsymbol{\nu}}_0 - \overline{\boldsymbol{q}}^\top \overline{\boldsymbol{\nu}}_0 + \tau\eta - \displaystyle\sum_{i=1}^p \gamma_i \varphi_i^\star(\boldsymbol{\phi}_{0i}/\gamma_i) \\ \text{subject to} & \displaystyle\sum_{i=1}^p \boldsymbol{\phi}_{0i} = \boldsymbol{\beta} + \underline{\boldsymbol{\nu}}_0 - \overline{\boldsymbol{\nu}}_0 + \eta\boldsymbol{\lambda} \\ & \underline{\boldsymbol{\nu}}_0, \overline{\boldsymbol{\nu}}_0 \in \mathbb{R}^n_+, \;\; \eta \in \mathbb{R}_+, \;\; \boldsymbol{\phi}_{0i} \in \mathbb{R}^n, \; i = 1, \ldots, p, \end{array} \tag{22}$$

is greater than or equal to α. In the second case, fix $\eta \in \mathbb{R}_+$ and set $\underline{\boldsymbol{\nu}}_0 = [-\boldsymbol{\beta} - \eta\boldsymbol{\lambda}]_+$, $\overline{\boldsymbol{\nu}}_0 = [\boldsymbol{\beta} + \eta\boldsymbol{\lambda}]_+$ and $\boldsymbol{\phi}_{0i} = \boldsymbol{0}$, $i = 1, \ldots, p$. This choice is feasible in (22) and attains the objective value

$$\underline{\boldsymbol{q}}^\top [-\boldsymbol{\beta} - \eta\boldsymbol{\lambda}]_+ - \overline{\boldsymbol{q}}^\top [\boldsymbol{\beta} + \eta\boldsymbol{\lambda}]_+ + \tau\eta - c \;\; = \;\; \eta\Big(\underline{\boldsymbol{q}}^\top [-\boldsymbol{\beta}/\eta - \boldsymbol{\lambda}]_+ - \overline{\boldsymbol{q}}^\top [\boldsymbol{\beta}/\eta + \boldsymbol{\lambda}]_+ + \tau\Big) - c \;\; \longrightarrow \;\; +\infty$$

as $\eta \longrightarrow +\infty$ since $\underline{\boldsymbol{q}}^\top [-\boldsymbol{\beta}/\eta - \boldsymbol{\lambda}]_+ - \overline{\boldsymbol{q}}^\top [\boldsymbol{\beta}/\eta + \boldsymbol{\lambda}]_+ \longrightarrow -\hat{\boldsymbol{q}}^\top \boldsymbol{\lambda}$ for $\hat{\boldsymbol{q}} \in \mathcal{Q}$ defined via $\hat{q}_i = \underline{q}_i$ if $\lambda_i < 0$ and $\hat{q}_i = \overline{q}_i$ otherwise, $i = 1, \ldots, n$, and $\boldsymbol{\lambda}^\top \boldsymbol{q} < \tau$ for all $\boldsymbol{q} \in \mathcal{Q}$. In this argument, the term $c = \sum_{i=1}^p \gamma_i \varphi_i^\star(\boldsymbol{\phi}_{0i}/\gamma_i) = \boldsymbol{\gamma}^\top \boldsymbol{\varphi}^\star(\boldsymbol{0})$ is constant. As for the third case, denote by (21_ϵ) and (22_ϵ) the variants of problems (21) and (22) where we replace the parameters $(\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}})$ with $(\underline{\boldsymbol{q}} - \epsilon\mathbf{e}, \overline{\boldsymbol{q}} + \epsilon\mathbf{e})$, respectively. Ignoring the trivial case where $\boldsymbol{\lambda} = \boldsymbol{0}$ and $\tau = 0$, we observe that strong duality holds between (21_ϵ) and (22_ϵ) for every $\epsilon > 0$. Moreover, one readily verifies that the mapping $\epsilon \mapsto (21_\epsilon)$ is right continuous at $\epsilon = 0$, that the problems (21) $\equiv$ (21_0) and (22) $\equiv$ (22_0) are both feasible, and that the optimal value of (21) is greater than or equal to the optimal value of (22) by weak duality. Since the optimal value of ($22_{\epsilon'}$) is greater than or equal to the optimal value of (22_ϵ) for all $0 \leq \epsilon' \leq \epsilon$, we thus conclude that the optimal values of the problems (21) and (22) also coincide.

The previous two paragraphs imply that for all $\tau \in \mathbb{R}$, $\tau \neq [\boldsymbol{\lambda}]_+^\top \overline{\boldsymbol{q}} - [-\boldsymbol{\lambda}]_+^\top \underline{\boldsymbol{q}}$, we have that

$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\boldsymbol{\lambda}^\top \tilde{\boldsymbol{q}}\big] \leq \tau$ if and only if

$$\begin{array}{c} \exists \alpha \in \mathbb{R},\;\; \boldsymbol{\beta} \in \mathbb{R}^n,\;\; \boldsymbol{\gamma} \in \mathbb{R}^p_+,\;\; \underline{\boldsymbol{\nu}}_1, \overline{\boldsymbol{\nu}}_1, \underline{\boldsymbol{\nu}}_0, \overline{\boldsymbol{\nu}}_0 \in \mathbb{R}^n_+, \\ \eta \in \mathbb{R}_+,\;\; \boldsymbol{\phi}_{1i}, \boldsymbol{\phi}_{0i} \in \mathbb{R}^n,\; i = 1, \ldots, p \;: \end{array} \quad \begin{cases} \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\sigma}^\top \boldsymbol{\gamma} \geq 1 - \epsilon \\ \underline{\boldsymbol{q}}^\top \underline{\boldsymbol{\nu}}_1 - \overline{\boldsymbol{q}}^\top \overline{\boldsymbol{\nu}}_1 - \sum_{i=1}^p \gamma_i \varphi_i^\star(\boldsymbol{\phi}_{1i}/\gamma_i) \geq \alpha - 1 \\ \underline{\boldsymbol{q}}^\top \underline{\boldsymbol{\nu}}_0 - \overline{\boldsymbol{q}}^\top \overline{\boldsymbol{\nu}}_0 + \tau\eta - \sum_{i=1}^p \gamma_i \varphi_i^\star(\boldsymbol{\phi}_{0i}/\gamma_i) \geq \alpha \\ \sum_{i=1}^p \boldsymbol{\phi}_{1i} = \boldsymbol{\beta} + \underline{\boldsymbol{\nu}}_1 - \overline{\boldsymbol{\nu}}_1,\;\; \sum_{i=1}^p \boldsymbol{\phi}_{0i} = \boldsymbol{\beta} + \underline{\boldsymbol{\nu}}_0 - \overline{\boldsymbol{\nu}}_0 + \eta\boldsymbol{\lambda}. \end{cases}$$

We claim that any feasible solution to this system of equations satisfies $\eta > 0$. Assume to the contrary that there was a feasible solution with $\eta = 0$. In that case, the constraint system would be independent of τ, and we would have $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\boldsymbol{\lambda}^\top \tilde{\boldsymbol{q}}\big] \leq \tau$ either for all $\tau \in \mathbb{R}$ or for no $\tau \in \mathbb{R}$. However, this cannot be the case since $\tilde{\boldsymbol{q}} \in \mathcal{Q}$ $\mathbb{P}$-a.s. for all $\mathbb{P} \in \mathcal{P}$ and the support $\mathcal{Q}$ is bounded. We thus conclude that $\eta > 0$, which allows us to replace all decision variables by their division through η and replace η with $1/\eta$. We have thus established that for $\tau \neq [\boldsymbol{\lambda}]_+^\top \overline{\boldsymbol{q}} - [-\boldsymbol{\lambda}]_+^\top \underline{\boldsymbol{q}}$, we have $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\boldsymbol{\lambda}^\top \tilde{\boldsymbol{q}}\big] \leq \tau$ if and only if

$$\begin{array}{c} \exists \alpha \in \mathbb{R},\;\; \boldsymbol{\beta} \in \mathbb{R}^n,\;\; \boldsymbol{\gamma} \in \mathbb{R}^p_+,\;\; \underline{\boldsymbol{\nu}}_1, \overline{\boldsymbol{\nu}}_1, \underline{\boldsymbol{\nu}}_0, \overline{\boldsymbol{\nu}}_0 \in \mathbb{R}^n_+, \\ \eta \in \mathbb{R}_+,\;\; \boldsymbol{\phi}_{1i}, \boldsymbol{\phi}_{0i} \in \mathbb{R}^n,\; i = 1, \ldots, p \;: \end{array} \quad \begin{cases} \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\sigma}^\top \boldsymbol{\gamma} \geq (1 - \epsilon)\eta \\ \underline{\boldsymbol{q}}^\top \underline{\boldsymbol{\nu}}_1 - \overline{\boldsymbol{q}}^\top \overline{\boldsymbol{\nu}}_1 - \sum_{i=1}^p \gamma_i \varphi_i^\star(\boldsymbol{\phi}_{1i}/\gamma_i) \geq \alpha - \eta \\ \underline{\boldsymbol{q}}^\top \underline{\boldsymbol{\nu}}_0 - \overline{\boldsymbol{q}}^\top \overline{\boldsymbol{\nu}}_0 + \tau - \sum_{i=1}^p \gamma_i \varphi_i^\star(\boldsymbol{\phi}_{0i}/\gamma_i) \geq \alpha \\ \sum_{i=1}^p \boldsymbol{\phi}_{1i} = \boldsymbol{\beta} + \underline{\boldsymbol{\nu}}_1 - \overline{\boldsymbol{\nu}}_1,\;\; \sum_{i=1}^p \boldsymbol{\phi}_{0i} = \boldsymbol{\beta} + \underline{\boldsymbol{\nu}}_0 - \overline{\boldsymbol{\nu}}_0 + \boldsymbol{\lambda}. \end{cases} \tag{23}$$

Assume now that $\tau = [\boldsymbol{\lambda}]_+^\top \overline{\boldsymbol{q}} - [-\boldsymbol{\lambda}]_+^\top \underline{\boldsymbol{q}}$. In that case, the application of our continuity argument to equation (20) could imply that $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\boldsymbol{\lambda}^\top \tilde{\boldsymbol{q}}\big] \leq \tau$ but the equation system (23) is not satisfiable. To prove that this is not possible, we show that the set-valued mapping $\tau \mapsto S(\tau)$, where $S(\tau)$ is the set of all $(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \underline{\boldsymbol{\nu}}_i, \overline{\boldsymbol{\nu}}_i, \eta, \boldsymbol{\phi}_{ij})$ satisfying (23), is outer semicontinuous (Rockafellar and Wets, 1997, Definition 5.4). This is the case if and only if the graph of S—that is, the set of all $\boldsymbol{\lambda} \in \mathbb{R}^n$, $\tau \in \mathbb{R}$ and $\alpha \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^n$, $\boldsymbol{\gamma} \in \mathbb{R}^p_+$, $\underline{\boldsymbol{\nu}}_1, \overline{\boldsymbol{\nu}}_1, \underline{\boldsymbol{\nu}}_0, \overline{\boldsymbol{\nu}}_0 \in \mathbb{R}^n_+$, $\eta \in \mathbb{R}_+$ as well as $\boldsymbol{\phi}_{1i}, \boldsymbol{\phi}_{0i} \in \mathbb{R}^n$, $i = 1, \ldots, p$, satisfying (23)—is closed (Rockafellar and Wets, 1997, Theorem 5.7). Indeed, the conjugate functions $\varphi_i^\star$ are convex by construction, and their convexity is preserved by the perspective functions (Boyd and Vandenberghe, 2004, §3.2.6). The result now follows since convex functions are continuous (Rockafellar and Wets, 1997, Theorem 2.35) and the lower level sets of continuous functions are closed (Rockafellar and Wets, 1997, Theorem 1.6). We thus conclude that for *all* $\tau \in \mathbb{R}$, we have $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\boldsymbol{\lambda}^\top \tilde{\boldsymbol{q}}\big] \leq \tau$ if and only if (23) is satisfiable.

By construction, the set of all $\boldsymbol{\lambda} \in \mathbb{R}^n$, $\tau \in \mathbb{R}$ and $\alpha \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^n$, $\boldsymbol{\gamma} \in \mathbb{R}^p_+$, $\underline{\boldsymbol{\nu}}_1, \overline{\boldsymbol{\nu}}_1, \underline{\boldsymbol{\nu}}_0, \overline{\boldsymbol{\nu}}_0 \in \mathbb{R}^n_+$, $\eta \in \mathbb{R}_+$ as well as $\boldsymbol{\phi}_{1i}, \boldsymbol{\phi}_{0i} \in \mathbb{R}^n$, $i = 1, \ldots, p$, satisfying the equation system (23) is convex. The set M is a projection of this set onto $\boldsymbol{\lambda}$ and τ and is thus convex as well. □

Proof of Proposition 1. The proof of Theorem 2 implies that for every $\tau \geq \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\boldsymbol{\lambda}^\top \tilde{\boldsymbol{q}}\big]$, the optimal value of the optimization problem

$$
\begin{array}{ll}
\text{maximize} & \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\sigma}^\top \boldsymbol{\gamma} \\
\text{subject to} & \underline{\boldsymbol{q}}^\top \underline{\boldsymbol{\nu}}_1 - \overline{\boldsymbol{q}}^\top \overline{\boldsymbol{\nu}}_1 - \displaystyle\sum_{i=1}^p \gamma_i \varphi_i^\star(\boldsymbol{\phi}_{1i}/\gamma_i) \geq \alpha - 1 \\
& \underline{\boldsymbol{q}}^\top \underline{\boldsymbol{\nu}}_0 - \overline{\boldsymbol{q}}^\top \overline{\boldsymbol{\nu}}_0 + \tau\eta - \displaystyle\sum_{i=1}^p \gamma_i \varphi_i^\star(\boldsymbol{\phi}_{0i}/\gamma_i) \geq \alpha \\
& \displaystyle\sum_{i=1}^p \boldsymbol{\phi}_{1i} = \underline{\boldsymbol{\nu}}_1 - \overline{\boldsymbol{\nu}}_1, \;\; \displaystyle\sum_{i=1}^p \boldsymbol{\phi}_{0i} = \boldsymbol{\beta} + \underline{\boldsymbol{\nu}}_0 - \overline{\boldsymbol{\nu}}_0 + \eta\boldsymbol{\lambda} \\
& \alpha \in \mathbb{R}, \;\; \boldsymbol{\beta} \in \mathbb{R}^n, \;\; \boldsymbol{\gamma} \in \mathbb{R}^p_+, \;\; \underline{\boldsymbol{\nu}}_1, \overline{\boldsymbol{\nu}}_1, \underline{\boldsymbol{\nu}}_0, \overline{\boldsymbol{\nu}}_0 \in \mathbb{R}^n_+ \\
& \eta \in \mathbb{R}_+, \;\; \boldsymbol{\phi}_{1i}, \boldsymbol{\phi}_{0i} \in \mathbb{R}^n, \, i = 1, \ldots, p
\end{array}
\tag{24}
$$

is greater than or equal to $1-\epsilon$. We claim that for $\tau = \tau^\star$, where $\tau^\star = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\boldsymbol{\lambda}^\top \tilde{\boldsymbol{q}}\big]$, the optimal value of problem (24) is in fact *equal to* $1-\epsilon$. Indeed, assume to the contrary that for $\tau = \tau^\star$, the optimal solution $(\alpha^\star, \boldsymbol{\beta}^\star, \boldsymbol{\gamma}^\star, \underline{\boldsymbol{\nu}}_i^\star, \overline{\boldsymbol{\nu}}_i^\star, \eta^\star, \boldsymbol{\phi}_{ij}^\star)$ to problem (24) satisfied $\alpha^\star + \boldsymbol{\mu}^\top \boldsymbol{\beta}^\star - \boldsymbol{\sigma}^\top \boldsymbol{\gamma}^\star > 1-\epsilon$. In that case, we could replace $\alpha^\star$ with $\hat{\alpha} < \alpha^\star$ such that $(\hat{\alpha}, \boldsymbol{\beta}^\star, \boldsymbol{\gamma}^\star, \underline{\boldsymbol{\nu}}_i^\star, \overline{\boldsymbol{\nu}}_i^\star, \eta^\star, \boldsymbol{\phi}_{ij}^\star)$ remains feasible for $\hat{\tau} < \tau^\star$ and still satisfies $\hat{\alpha} + \boldsymbol{\mu}^\top \boldsymbol{\beta}^\star - \boldsymbol{\sigma}^\top \boldsymbol{\gamma}^\star \geq 1-\epsilon$. This, however, would contradict the definition of $\tau^\star$ as the smallest value of τ for which there is $(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \underline{\boldsymbol{\nu}}_i, \overline{\boldsymbol{\nu}}_i, \eta, \boldsymbol{\phi}_{ij})$ feasible in problem (24) with an objective value greater than or equal to $1-\epsilon$. We thus conclude that the optimal value of the problem (24) for $\tau = \tau^\star$ is *exactly* $1-\epsilon$. Strong convex duality, which holds since problem (24) admits a Slater point, then implies that the optimal value of the dual problem,

$$
\begin{array}{ll}
\text{minimize} & \xi_1 \\
\text{subject to} & \xi_0 + \xi_1 = 1, \;\; \boldsymbol{\zeta}_0 + \boldsymbol{\zeta}_1 = \boldsymbol{\mu} \\
& \xi_i \underline{\boldsymbol{q}} \leq \boldsymbol{\zeta}_i \leq \xi_i \overline{\boldsymbol{q}} \qquad\qquad\qquad\qquad \forall i \in \{0,1\} \\
& \boldsymbol{\lambda}^\top \boldsymbol{\zeta}_0 \geq \xi_0 \tau^\star \\
& \xi_0 \cdot \boldsymbol{\varphi}(\boldsymbol{\zeta}_0/\xi_0) + \xi_1 \cdot \boldsymbol{\varphi}(\boldsymbol{\zeta}_1/\xi_1) \leq \boldsymbol{\sigma} \\
& \xi_i \in \mathbb{R}_+, \;\; \boldsymbol{\zeta}_i \in \mathbb{R}^n, \; i = 0,1,
\end{array}
\tag{25}
$$

is also equal to $1-\epsilon$. Let $(\xi_i^\star, \boldsymbol{\zeta}_i^\star)$ be an optimal solution to this problem.

We claim that the sequence of two-point distribution $\mathbb{P}^t$ defined by

$$
\mathbb{P}^t = (\xi_1^\star - 1/t) \cdot \delta_{\frac{\boldsymbol{\zeta}_1^\star}{\xi_1^\star}} + (\xi_0^\star + 1/t) \cdot \delta_{\frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star} + \frac{1}{t(\xi_0^\star + 1/t)}\left(\frac{\boldsymbol{\zeta}_1^\star}{\xi_1^\star} - \frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star}\right)}, \quad t = 1, 2, \ldots,
$$

satisfies *(i)* $\mathbb{P}^t \in \mathcal{P}$ for sufficiently large t as well as *(ii)* $\mathbb{P}^t\text{-VaR}_{1-\epsilon}\big[\boldsymbol{\lambda}^\top \tilde{\boldsymbol{q}}\big] \longrightarrow \tau^\star$ as $t \longrightarrow \infty$.

In view of statement *(i)*, we note that $\mathbb{P}^t$ is a probability distribution for t sufficiently large since $(\xi_0^\star, \xi_1^\star) = (\epsilon, 1-\epsilon)$ due to the first constraint set in (25). Also, $\mathbb{P}^t$ is supported on $\mathcal{Q}$ for sufficiently large t since $\boldsymbol{\zeta}_i^\star/\xi_i^\star \in [\underline{\boldsymbol{q}}, \overline{\boldsymbol{q}}]$, $i = 0,1$, due to the second constraint in (25) and for t sufficiently large, $\frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star} + \frac{1}{t(\xi_0^\star + 1/t)}\left(\frac{\boldsymbol{\zeta}_1^\star}{\xi_1^\star} - \frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star}\right)$ is a convex combination of $\boldsymbol{\zeta}_0^\star/\xi_0^\star$ and $\boldsymbol{\zeta}_1^\star/\xi_1^\star$. Likewise, we have

$$
\mathbb{E}_{\mathbb{P}^t}\big[\tilde{\boldsymbol{q}}\big] \;=\; (\xi_1^\star - 1/t) \cdot \frac{\boldsymbol{\zeta}_1^\star}{\xi_1^\star} + (\xi_0^\star + 1/t) \cdot \frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star} + \frac{1}{t}\left(\frac{\boldsymbol{\zeta}_1^\star}{\xi_1^\star} - \frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star}\right) \;=\; \boldsymbol{\mu}
$$

due to the first constraint set in (25) as well as, for t sufficiently large,

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}^t}\big[\boldsymbol{\varphi}(\tilde{\boldsymbol{q}})\big] \;&=\; (\xi_1^\star - 1/t) \cdot \boldsymbol{\varphi}\left(\frac{\boldsymbol{\zeta}_1^\star}{\xi_1^\star}\right) + (\xi_0^\star + 1/t) \cdot \boldsymbol{\varphi}\left(\frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star} + \frac{1}{t(\xi_0^\star + 1/t)}\left[\frac{\boldsymbol{\zeta}_1^\star}{\xi_1^\star} - \frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star}\right]\right) \\
&\leq\; (\xi_1^\star - 1/t) \cdot \boldsymbol{\varphi}\left(\frac{\boldsymbol{\zeta}_1^\star}{\xi_1^\star}\right) + \frac{1}{t} \cdot \boldsymbol{\varphi}\left(\frac{\boldsymbol{\zeta}_1^\star}{\xi_1^\star}\right) + \xi_0^\star \cdot \boldsymbol{\varphi}\left(\frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star}\right) \;\leq\; \boldsymbol{\sigma},
\end{aligned}
$$

where the inequalities follow from the convexity of $\boldsymbol{\varphi}$ and the fourth constraint in (25), respectively.

To show statement *(ii)*, we note that $\mathbb{P}^t$ places a probability mass of $\xi_0^\star + 1/t = \epsilon + 1/t$ on the scenario $\frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star} + \frac{1}{t(\xi_0^\star + 1/t)}\left(\frac{\boldsymbol{\zeta}_1^\star}{\xi_1^\star} - \frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star}\right)$, which satisfies

$$
\boldsymbol{\lambda}^\top \left[\frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star} + \frac{1}{t(\xi_0^\star + 1/t)}\left(\frac{\boldsymbol{\zeta}_1^\star}{\xi_1^\star} - \frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star}\right)\right] \;\geq\; \tau^\star \;+\; \frac{1}{t(\xi_0^\star + 1/t)} \boldsymbol{\lambda}^\top \left(\frac{\boldsymbol{\zeta}_1^\star}{\xi_1^\star} - \frac{\boldsymbol{\zeta}_0^\star}{\xi_0^\star}\right) \;\underset{t \longrightarrow \infty}{\longrightarrow}\; \tau^\star.
$$

Here, the inequality follows from the fact that $\boldsymbol{\lambda}^\top \boldsymbol{\zeta}_0^\star/\xi_0^\star \geq \tau^\star$ due to the third constraint in (25). The convergence of the middle expression to $\tau^\star$ holds since $\boldsymbol{\lambda}^\top\left(\boldsymbol{\zeta}_1^\star/\xi_1^\star - \boldsymbol{\zeta}_0^\star/\xi_0^\star\right)$ is finite while $t(\xi_0^\star +$

$1/t) \longrightarrow \infty$ as $\xi_0^\star = \epsilon > 0$. We have thus established that $\mathbb{P}^t\text{-VaR}_{1-\epsilon}\big[\boldsymbol{\lambda}^\top \tilde{\boldsymbol{q}}\big] \longrightarrow \tau'$ with $\tau' \geq \tau^\star$ as $t \longrightarrow \infty$. Since $\mathbb{P}^t \in \mathcal{P}$ for sufficiently large t, on the other hand, the definition of $\tau^\star$ implies that $\tau' \leq \tau^\star$ as well, which concludes the proof. □

Proof of Theorem 3. Since the assumptions of Theorem 2 are satisfied, we conclude that

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\Big[\sum_{i\in S} \tilde{q}_i\Big] \;\leq\; \sum_{i\in S} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big] \qquad \forall S \subseteq V_C,\, S \neq \emptyset.$$

To show the reverse inequality, we note that for any $\kappa > 0$, we have

$$\begin{aligned} \sum_{i\in S} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big] \;&\leq\; \sum_{i\in S} \big(\mathbb{P}_i^\star\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big] + \kappa\big) \\ &=\; \sum_{i\in S} \big(\mathbb{P}^\star\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big] + \kappa\big) \;=\; \mathbb{P}^\star\text{-VaR}_{1-\epsilon}\Big[\sum_{i\in S} \tilde{q}_i\Big] + |S|\kappa. \end{aligned} \tag{26}$$

Here, $\mathbb{P}_i^\star \in \mathcal{P}$ is a distribution that satisfies $\mathbb{P}_i^\star\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big] \geq \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big] - \kappa$, which implies the first inequality. In the second row, we define the probability measure $\mathbb{P}^\star$ via

$$\mathbb{P}^\star\big(\tilde{\boldsymbol{q}} \leq \boldsymbol{q}\big) \;=\; \min_{i\in S} \mathbb{P}_i^\star\big(\tilde{q}_i \leq q_i\big) \qquad \forall \boldsymbol{q} \in \mathbb{R}^n.$$

By construction, $\mathbb{P}^\star$ has the same marginal distributions as $\mathbb{P}_i^\star$, $i \in V_C$, that is, $\mathbb{P}^\star\big(\tilde{q}_i \in A\big) = \mathbb{P}_i^\star\big(\tilde{q}_i \in A\big)$ for all $i \in V_C$ and for every measurable set A (Dhaene et al., 2002, Theorem 2). From the definition of the marginalized moment ambiguity sets, we thus conclude that $\mathbb{P}^\star \in \mathcal{P}$. Since $\mathbb{P}^\star$ is comonotonic (Dhaene et al., 2002, Definition 4 and Theorem 2), the last equality in (26) follows from the comonotone additivity of the value-at-risk (Pflug, 2000, Proposition 3). As κ was chosen arbitrarily in (26) and since $\mathbb{P}^\star \in \mathcal{P}$, we thus conclude that

$$\sum_{i\in S} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big] \;\leq\; \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\Big[\sum_{i\in S} \tilde{q}_i\Big]$$

as desired. This completes the proof. □

Proof of Corollary 1. Since the ambiguity set $\mathcal{P}$ is subadditive, Theorem 1 implies that RVRP($\mathcal{P}$) is equivalent to 2VF($\mathcal{P}$). Moreover, Theorem 3 allows us to interpret 2VF($\mathcal{P}$) as the two-index vehicle flow formulation of a deterministic CVRP with customer demands $q_i = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big]$, $i \in V_C$. The statement now follows from the well-known equivalence of the two-index vehicle flow formulation and the deterministic CVRP. □

Proof of Proposition 2. We apply Theorem 5 on page 21 of the main paper to conclude that the worst-case value-at-risk $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big]$ is equal to the optimal objective value of the following problem.

$$\begin{array}{ll} \text{minimize} & \mu_i + \min\left\{ \left(\overline{q}_i - \mu_i\right),\, \dfrac{1-\epsilon}{\epsilon}\left(\mu_i - \underline{q}_i\right) \right\} \cdot [1-2\gamma]_+ + \dfrac{1}{\epsilon}\nu\gamma \\ \text{subject to} & \gamma \in \mathbb{R}_+. \end{array}$$

The first and second terms in the objective function are constant and non-decreasing in γ, respectively, for $\gamma \geq 1/2$. Without loss of generality, we can therefore assume that $\gamma \leq 1/2$ at optimality. We thus obtain a linearized version of the problem as follows.

$$\begin{array}{ll} \text{minimize} & \mu_i + \min\left\{ \left(\overline{q}_i - \mu_i\right),\, \dfrac{1-\epsilon}{\epsilon}\left(\mu_i - \underline{q}_i\right) \right\} \cdot [1-2\gamma] + \dfrac{1}{\epsilon}\nu\gamma \\ \text{subject to} & \gamma \in [0,\, 1/2] \end{array}$$

Since the objective function is linear in γ, the problem is optimized by either $\gamma = 0$ or $\gamma = 1/2$. The result now follows from a case distinction. □

Proof of Proposition 3. We apply Theorem 7 on page 24 of the main paper to conclude that the worst-case value-at-risk $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big]$ is equal to the optimal objective value of the problem

$$\begin{array}{ll} \text{maximize} & \mu_i + q_i \\ \text{subject to} & \boldsymbol{q}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{q} \leq \dfrac{1-\epsilon}{\epsilon} \\ & \boldsymbol{q} \in \left[\boldsymbol{q}^\ell,\, \boldsymbol{q}^{\mathrm{u}}\right], \end{array} \tag{27}$$

where $\boldsymbol{q}^\ell = \max\left\{-\frac{1-\epsilon}{\epsilon}(\overline{\boldsymbol{q}} - \boldsymbol{\mu}),\, \underline{\boldsymbol{q}} - \boldsymbol{\mu}\right\}$, $\boldsymbol{q}^{\mathrm{u}} = \min\left\{\frac{1-\epsilon}{\epsilon}(\boldsymbol{\mu} - \underline{\boldsymbol{q}}),\, \overline{\boldsymbol{q}} - \boldsymbol{\mu}\right\}$ and the covariance matrix satisfies $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$. If $\boldsymbol{q}$ is feasible for the above problem, then so is $\boldsymbol{q}'$ with $q_i' = q_i$ and $q_j' = 0$ for all $j \neq i$. Indeed, we have $\boldsymbol{q}^\ell \leq \boldsymbol{0}$ since $-\frac{1-\epsilon}{\epsilon}(\overline{\boldsymbol{q}} - \boldsymbol{\mu}) \leq \boldsymbol{0}$ and $\underline{\boldsymbol{q}} - \boldsymbol{\mu} \leq \boldsymbol{0}$, as well as $\boldsymbol{q}^{\mathrm{u}} \geq \boldsymbol{0}$ since $\frac{1-\epsilon}{\epsilon}(\boldsymbol{\mu} - \underline{\boldsymbol{q}}) \geq \boldsymbol{0}$ and $\overline{\boldsymbol{q}} - \boldsymbol{\mu} \geq \boldsymbol{0}$. Moreover, we have $\boldsymbol{q}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{q} = \sum_{j=1}^n q_j^2/\sigma_j \geq q_i^2/\sigma_i = \boldsymbol{q}'^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{q}'$. Since $\boldsymbol{q}$ and $\boldsymbol{q}'$ attain the same objective value in (27), we thus conclude that problem (27) attains the same optimal value as the univariate optimization problem

$$\begin{array}{ll} \text{maximize} & \mu_i + q_i \\ \text{subject to} & q_i^2/\sigma_i \leq \dfrac{1-\epsilon}{\epsilon} \\ & q_i \in \left[q_i^\ell,\, q_i^{\mathrm{u}}\right]. \end{array} \tag{28}$$

At optimality we have $q_i^2/\sigma_i = \frac{1-\epsilon}{\epsilon}$ or $q_i = q_i^{\mathrm{u}}$. The result now follows from a case distinction. $\square$

Proof of Proposition 4. The rectangularity of the ambiguity set $\mathcal{P}$ allows us to conclude that

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big] = \sup_{\mathbb{P}\in\mathcal{P}_i} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big]$$

with

$$\mathcal{P}_i = \left\{ \mathbb{P}\in\mathcal{P}_0(\mathbb{R}) \;:\; \left[\begin{array}{l} \mathbb{P}\left[\tilde{q}_i \in [\underline{q}_i,\, \overline{q}_i]\right] = 1, \;\; \mathbb{E}_{\mathbb{P}}\left[\tilde{q}_i\right] = \mu_i, \\ \mathbb{E}_{\mathbb{P}}\left[[\tilde{q}_i - \mu_i]_+^2\right] \le \sigma_i^+, \;\; \mathbb{E}_{\mathbb{P}}\left[[\mu_i - \tilde{q}_i]_+^2\right] \le \sigma_i^- \;\; \forall i \in V_C \end{array}\right]\right\}.$$

For a fixed scalar $\tau\in\mathbb{R}$, we then have $\sup_{\mathbb{P}\in\mathcal{P}_i} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\tilde{q}_i\big] \le \tau$ if and only if the optimal objective value of the moment problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\int_{[\underline{q}_i,\overline{q}_i]} \mathbb{I}_{[q_i\le\tau]}\mathbb{P}(\mathrm{d}q_i) \\ \text{subject to} & \displaystyle\int_{[\underline{q}_i,\overline{q}_i]} \mathbb{P}(\mathrm{d}q_i) = 1 \\ & \displaystyle\int_{[\underline{q}_i,\overline{q}_i]} q_i\,\mathbb{P}(\mathrm{d}q_i) = \mu_i \\ & \displaystyle\int_{[\underline{q}_i,\overline{q}_i]} ([q_i-\mu_i]_+)^2\,\mathbb{P}(\mathrm{d}q_i) \le \sigma_i^+ \\ & \displaystyle\int_{[\underline{q}_i,\overline{q}_i]} ([\mu_i-q_i]_+)^2\,\mathbb{P}(\mathrm{d}q_i) \le \sigma_i^- \\ & \mathbb{P}\in\mathcal{M}_+(\mathbb{R}) \end{array}$$

is greater than or equal to $1-\epsilon$. A similar reasoning as in the proof of Theorem 2 shows that this is the case if and only if the optimal objective value of the problem

$$\begin{array}{ll} \text{maximize} & \alpha + \mu_i\beta - \sigma_i^+\gamma^+ - \sigma_i^-\gamma^- \\ \text{subject to} & \alpha - \mu_i(\chi_1^q - \pi_1^q) + \dfrac{1}{4\gamma^+}(\pi_1^q + \pi_1^0)^2 + \dfrac{1}{4\gamma^-}(\chi_1^q + \chi_1^0)^2 - \underline{q}_i\underline{\phi}_1 + \overline{q}_i\overline{\phi}_1 \le 1 \\ & \alpha - \mu_i(\chi_0^q - \pi_0^q) + \dfrac{1}{4\gamma^+}(\pi_0^q + \pi_0^0)^2 + \dfrac{1}{4\gamma^-}(\chi_0^q + \chi_0^0)^2 - \underline{q}_i\underline{\phi}_0 + \overline{q}_i\overline{\phi}_0 - \tau\omega \le 0 \\ & \pi_1^q - \chi_1^q + \overline{\phi}_1 - \underline{\phi}_1 = \beta, \;\; \pi_0^q - \chi_0^q + \overline{\phi}_0 - \underline{\phi}_0 - \omega = \beta \\ & \alpha,\, \beta\in\mathbb{R},\, \gamma^+,\gamma^-\in\mathbb{R}_+,\, \chi_j^q,\chi_j^0,\pi_j^q,\pi_j^0,\underline{\phi}_j,\overline{\phi}_j\in\mathbb{R}_+,\, j=0,1,\, \omega\in\mathbb{R}_+. \end{array} \tag{29}$$

is greater than or equal to $1-\epsilon$.

We now consider the problem $\sup_{\mathbb{P}\in\mathcal{P}_i} \mathbb{P}\text{-VaR}_{1-\epsilon}[\tilde{q}_i]$, which can be formulated as

$$\begin{aligned} \text{minimize} \quad & \tau \\ \text{subject to} \quad & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}[\tilde{q}_i] \leq \tau \\ & \tau \in \mathbb{R}. \end{aligned}$$

Our previous arguments imply that this problem is equivalent to

$$\begin{aligned} \text{minimize} \quad & \tau \\ \text{subject to} \quad & \alpha + \mu_i\beta - \sigma_i^+\gamma^+ - \sigma_i^-\gamma^- \geq 1-\epsilon \\ & \alpha - \mu_i(\chi_1^q - \pi_1^q) + \frac{1}{4\gamma^+}(\pi_1^q + \pi_1^0)^2 + \frac{1}{4\gamma^-}(\chi_1^q + \chi_1^0)^2 - \underline{q}_i\underline{\phi}_1 + \overline{q}_i\overline{\phi}_1 \leq 1 \\ & \alpha - \mu_i(\chi_0^q - \pi_0^q) + \frac{1}{4\gamma^+}(\pi_0^q + \pi_0^0)^2 + \frac{1}{4\gamma^-}(\chi_0^q + \chi_0^0)^2 - \underline{q}_i\underline{\phi}_0 + \overline{q}_i\overline{\phi}_0 - \tau\omega \leq 0 \\ & \pi_1^q - \chi_1^q + \overline{\phi}_1 - \underline{\phi}_1 = \beta, \;\; \pi_0^q - \chi_0^q + \overline{\phi}_0 - \underline{\phi}_0 - \omega = \beta \\ & \alpha,\, \beta \in \mathbb{R},\, \gamma^+,\, \gamma^- \in \mathbb{R}_+,\, \chi_j^q, \chi_j^0, \pi_j^q, \pi_j^0, \underline{\phi}_j, \overline{\phi}_j \in \mathbb{R}_+,\, j = 0,1,\, \omega \in \mathbb{R}_+,\, \tau \in \mathbb{R}. \end{aligned} \tag{30}$$

Similar manipulations as in the proof of Theorem 7 on page 24 of the main paper allow us to conclude that this problem has the same objective value as

$$\begin{aligned} \text{minimize} \quad & (\overline{q}_i - \mu_i)\overline{\phi}_0 + \left(\frac{1-\epsilon}{\epsilon}\right)(\mu_i - \underline{q}_i)\underline{\phi}_1 + \frac{1}{4\gamma^+}\left(\pi_0^q\right)^2 + \left(\frac{1-\epsilon}{\epsilon}\right)\frac{1}{4\gamma^-}\left(\chi_1^q\right)^2 + \mu_i + \frac{\sigma_i^+}{\epsilon}\gamma^+ + \frac{\sigma_i^-}{\epsilon}\gamma^- \\ \text{subject to} \quad & \pi_0^q + \chi_1^q + \overline{\phi}_0 + \underline{\phi}_1 = 1 \\ & \gamma^+,\, \gamma^- \in \mathbb{R}_+,\, \chi_1^q, \pi_0^q, \underline{\phi}_1, \overline{\phi}_0 \in \mathbb{R}_+. \end{aligned}$$

The unconstrained first-order optimality condition with respect to γ^+ gives $\gamma^+ = \pm\frac{1}{2}\sqrt{\frac{\epsilon}{\sigma_i^+}}\pi_0^q$. Since the second derivative $\frac{1}{2}\frac{(\pi_0^q)^2}{(\gamma^+)^3}$ is non-negative for $\gamma^+ \geq 0$, we thus conclude that $\gamma^+ = \frac{1}{2}\sqrt{\frac{\epsilon}{\sigma_i^+}}\pi_0^q$ is optimal in the above problem. A similar reasoning shows that $\gamma^- = \frac{1}{2}\sqrt{\frac{1-\epsilon}{\sigma_i^-}}\chi_1^q$ is optimal as well. We thus obtain the equivalent optimization problem

$$\begin{aligned} \text{minimize} \quad & \mu_i + (\overline{q}_i - \mu_i)\overline{\phi}_0 + \left(\frac{1-\epsilon}{\epsilon}\right)(\mu_i - \underline{q}_i)\underline{\phi}_1 + \sqrt{\frac{\sigma_i^+}{\epsilon}}\pi_0^q + \frac{\sqrt{(1-\epsilon)\sigma_i^-}}{\epsilon}\chi_1^q \\ \text{subject to} \quad & \pi_0^q + \chi_1^q + \overline{\phi}_0 + \underline{\phi}_1 = 1 \\ & \chi_1^q, \pi_0^q, \underline{\phi}_1, \overline{\phi}_0 \in \mathbb{R}_+. \end{aligned}$$

Since all objective coefficients are strictly positive, there is an optimal solution that sets one of the four decision variables with minimum objective coefficient to 1 and all other variables to 0. The statement then follows from a case distinction. □

Proof of Theorem 4. To prove the statement, we consider a distributionally robust CVRP instance with $n = 4$ customers, $m = 2$ vehicles of capacity $Q = 10$, a risk threshold $\epsilon = 0.1$ and a first-order generic moment ambiguity set of the form (12) with support $\mathcal{Q} = [1, 10]^4$, expected demands $\boldsymbol{\mu} = 4.6\mathbf{e}$ and the following dispersion constraints:

$$\mathbb{E}_{\mathbb{P}}\left[|\tilde{q}_1 - \mu_1| + |\tilde{q}_3 - \mu_3|\right] \leq 0.1, \quad \mathbb{E}_{\mathbb{P}}\left[|\tilde{q}_2 - \mu_2| + |\tilde{q}_4 - \mu_4|\right] \leq 0.1$$

Evaluating the worst-case values-at-risk of all customer subsets shows that the subsets $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{1, 3\}$ and $\{2, 4\}$ can all be served by a single vehicle, $\{1, 2\}$, $\{1, 4\}$, $\{2, 3\}$, $\{3, 4\}$, $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{1, 3, 4\}$ and $\{2, 3, 4\}$ require two vehicles, and the set of all customers also requires two vehicles. This implies that the set of feasible route sets consists of the permutations of $\{\{1, 3\}, \{2, 4\}\}$. We claim that there is no deterministic CVRP instance that has this set of feasible route sets.

Assume to the contrary that there is a demand vector $\boldsymbol{q}$ such that the associated deterministic CVRP instance has the aforementioned set of feasible route sets. In this case, we have $q_1 + q_2 > 10$ since $\{1, 2\}$ cannot be served by a single vehicle. Since the four customers together require two vehicles, we also have $10 < q_1 + q_2 + q_3 + q_4 \leq 20$. We thus conclude that $q_3 + q_4 \leq 10$. This is not possible, however, since the route $\{3, 4\}$ cannot be served by a single vehicle. $\square$

Proof of Theorem 5. For any fixed scalar $\tau \in \mathbb{R}$, we have $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\left[\mathbf{1}_S{}^\top \tilde{\boldsymbol{q}}\right] \leq \tau$ if and only if the optimal objective value of the moment problem

$$\begin{aligned}
\text{minimize} \quad & \int_{\mathcal{Q}} \mathbb{I}_{[\mathbf{1}_S{}^\top \boldsymbol{q} \leq \tau]} \, \mathbb{P}(\mathrm{d}\boldsymbol{q}) \\
\text{subject to} \quad & \int_{\mathcal{Q}} \mathbb{P}(\mathrm{d}\boldsymbol{q}) = 1 \\
& \int_{\mathcal{Q}} \boldsymbol{q} \, \mathbb{P}(\mathrm{d}\boldsymbol{q}) = \boldsymbol{\mu} \\
& \int_{\mathcal{Q}} \mathbf{1}_{S_i}^\top |\boldsymbol{q} - \boldsymbol{\mu}| \, \mathbb{P}(\mathrm{d}\boldsymbol{q}) \leq \nu_i \qquad \forall i = 1, \ldots, p \\
& \mathbb{P} \in \mathcal{M}_+(\mathcal{Q})
\end{aligned}$$

is greater than or equal to $1 - \epsilon$. A similar reasoning as in the proof of Theorem 2 shows that this is the case if and only if the optimal objective value of the problem

$$\begin{array}{ll} \text{maximize} & \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\nu}^\top \boldsymbol{\gamma} \\ \text{subject to} & \alpha + \boldsymbol{\mu}^\top (\boldsymbol{\pi}_1^+ - \boldsymbol{\pi}_1^-) + \overline{\boldsymbol{q}}^\top \overline{\boldsymbol{\phi}}_1 - \underline{\boldsymbol{q}}^\top \underline{\boldsymbol{\phi}}_1 \leq 1 \\ & \alpha + \boldsymbol{\mu}^\top (\boldsymbol{\pi}_0^+ - \boldsymbol{\pi}_0^-) + \overline{\boldsymbol{q}}^\top \overline{\boldsymbol{\phi}}_0 - \underline{\boldsymbol{q}}^\top \underline{\boldsymbol{\phi}}_0 - \tau\omega \leq 0 \\ & \boldsymbol{\pi}_1^+ - \boldsymbol{\pi}_1^- + \overline{\boldsymbol{\phi}}_1 - \underline{\boldsymbol{\phi}}_1 = \boldsymbol{\beta} \\ & \boldsymbol{\pi}_0^+ - \boldsymbol{\pi}_0^- + \overline{\boldsymbol{\phi}}_0 - \underline{\boldsymbol{\phi}}_0 - \mathbf{1}_S\omega = \boldsymbol{\beta} \\ & \boldsymbol{\pi}_1^+ + \boldsymbol{\pi}_1^- \leq \sum_{i=1}^p \gamma_i \mathbf{1}_{S_i}, \;\; \boldsymbol{\pi}_0^+ + \boldsymbol{\pi}_0^- \leq \sum_{i=1}^p \gamma_i \mathbf{1}_{S_i} \\ & \boldsymbol{\pi}_i^+, \boldsymbol{\pi}_i^-, \overline{\boldsymbol{\phi}}_i, \underline{\boldsymbol{\phi}}_i \in \mathbb{R}_+^n, \; i = 0, 1, \;\; \omega \in \mathbb{R}_+ \\ & \alpha \in \mathbb{R}, \;\; \boldsymbol{\beta} \in \mathbb{R}^n, \;\; \boldsymbol{\gamma} \in \mathbb{R}_+^p. \end{array}$$

is greater than or equal to $1 - \epsilon$.

We now consider the problem $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\mathbf{1}_S{}^\top \tilde{\boldsymbol{q}}\big]$, which can be formulated as

$$\begin{array}{ll} \text{minimize} & \tau \\ \text{subject to} & \sup\limits_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\mathbf{1}_S{}^\top \tilde{\boldsymbol{q}}\big] \leq \tau \\ & \tau \in \mathbb{R}. \end{array}$$

Our previous arguments imply that this problem is equivalent to

$$\begin{array}{ll} \text{minimize} & \tau \\ \text{subject to} & \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\nu}^\top \boldsymbol{\gamma} \geq 1 - \epsilon \\ & \alpha + \boldsymbol{\mu}^\top (\boldsymbol{\pi}_1^+ - \boldsymbol{\pi}_1^-) + \overline{\boldsymbol{q}}^\top \overline{\boldsymbol{\phi}}_1 - \underline{\boldsymbol{q}}^\top \underline{\boldsymbol{\phi}}_1 \leq 1 \\ & \alpha + \boldsymbol{\mu}^\top (\boldsymbol{\pi}_0^+ - \boldsymbol{\pi}_0^-) + \overline{\boldsymbol{q}}^\top \overline{\boldsymbol{\phi}}_0 - \underline{\boldsymbol{q}}^\top \underline{\boldsymbol{\phi}}_0 - \tau\omega \leq 0 \\ & \boldsymbol{\pi}_1^+ - \boldsymbol{\pi}_1^- + \overline{\boldsymbol{\phi}}_1 - \underline{\boldsymbol{\phi}}_1 = \boldsymbol{\beta} \\ & \boldsymbol{\pi}_0^+ - \boldsymbol{\pi}_0^- + \overline{\boldsymbol{\phi}}_0 - \underline{\boldsymbol{\phi}}_0 - \mathbf{1}_S\omega = \boldsymbol{\beta} \\ & \boldsymbol{\pi}_1^+ + \boldsymbol{\pi}_1^- \leq \sum_{i=1}^p \gamma_i \mathbf{1}_{S_i}, \;\; \boldsymbol{\pi}_0^+ + \boldsymbol{\pi}_0^- \leq \sum_{i=1}^p \gamma_i \mathbf{1}_{S_i} \\ & \boldsymbol{\pi}_i^+, \boldsymbol{\pi}_i^-, \overline{\boldsymbol{\phi}}_i, \underline{\boldsymbol{\phi}}_i \in \mathbb{R}_+^n, \; i = 0, 1, \;\; \omega \in \mathbb{R}_+ \\ & \alpha \in \mathbb{R}, \;\; \boldsymbol{\beta} \in \mathbb{R}^n, \;\; \boldsymbol{\gamma} \in \mathbb{R}_+^p, \;\; \tau \in \mathbb{R}. \end{array} \tag{31}$$

Note that in the absence of the first constraint, it would be optimal to choose α as small as possible. We can thus remove the first constraint and replace α with $(1-\epsilon) - \boldsymbol{\mu}^\top\big(\boldsymbol{\pi}_1^+ - \boldsymbol{\pi}_1^- + \overline{\boldsymbol{\phi}}_1 - \underline{\boldsymbol{\phi}}_1\big) + \boldsymbol{\nu}^\top\boldsymbol{\gamma}$ in the second constraint, resulting in

$$\big(\overline{\boldsymbol{q}} - \boldsymbol{\mu}\big)^\top \overline{\boldsymbol{\phi}}_1 - \big(\underline{\boldsymbol{q}} - \boldsymbol{\mu}\big)^\top \underline{\boldsymbol{\phi}}_1 + \boldsymbol{\nu}^\top\boldsymbol{\gamma} \leq \epsilon,$$

as well as with $(1-\epsilon) - \boldsymbol{\mu}^\top\big(\boldsymbol{\pi}_0^+ - \boldsymbol{\pi}_0^- + \overline{\boldsymbol{\phi}}_0 - \underline{\boldsymbol{\phi}}_0 - \mathbf{1}_S\omega\big) + \boldsymbol{\nu}^\top\boldsymbol{\gamma}$ in the third constraint, resulting in

$$\big(\overline{\boldsymbol{q}} - \boldsymbol{\mu}\big)^\top\overline{\boldsymbol{\phi}}_0 - \big(\underline{\boldsymbol{q}} - \boldsymbol{\mu}\big)^\top\underline{\boldsymbol{\phi}}_0 + \boldsymbol{\nu}^\top\boldsymbol{\gamma} + \boldsymbol{\mu}^\top\mathbf{1}_S\omega - \tau\omega \le -(1-\epsilon).$$

Moreover, since $\boldsymbol{\beta}$ is unrestricted in sign, we can remove it from the problem by replacing the fourth and fifth constraint in the above problem with the single constraint

$$\boldsymbol{\pi}_1^+ - \boldsymbol{\pi}_1^- + \overline{\boldsymbol{\phi}}_1 - \underline{\boldsymbol{\phi}}_1 = \boldsymbol{\pi}_0^+ - \boldsymbol{\pi}_0^- + \overline{\boldsymbol{\phi}}_0 - \underline{\boldsymbol{\phi}}_0 - \mathbf{1}_S\omega.$$

The optimization problem (31) is thus equivalent to

$$\begin{array}{ll}
\text{minimize} & \tau \\
\text{subject to} & \big(\overline{\boldsymbol{q}} - \boldsymbol{\mu}\big)^\top\overline{\boldsymbol{\phi}}_1 - \big(\underline{\boldsymbol{q}} - \boldsymbol{\mu}\big)^\top\underline{\boldsymbol{\phi}}_1 + \boldsymbol{\nu}^\top\boldsymbol{\gamma} \le \epsilon \\
& \big(\overline{\boldsymbol{q}} - \boldsymbol{\mu}\big)^\top\overline{\boldsymbol{\phi}}_0 - \big(\underline{\boldsymbol{q}} - \boldsymbol{\mu}\big)^\top\underline{\boldsymbol{\phi}}_0 + \boldsymbol{\nu}^\top\boldsymbol{\gamma} + \boldsymbol{\mu}^\top\mathbf{1}_S\omega - \tau\omega \le -(1-\epsilon) \\
& \boldsymbol{\pi}_1^+ - \boldsymbol{\pi}_1^- + \overline{\boldsymbol{\phi}}_1 - \underline{\boldsymbol{\phi}}_1 = \boldsymbol{\pi}_0^+ - \boldsymbol{\pi}_0^- + \overline{\boldsymbol{\phi}}_0 - \underline{\boldsymbol{\phi}}_0 - \mathbf{1}_S\omega \\
& \boldsymbol{\pi}_1^+ + \boldsymbol{\pi}_1^- \le \sum_{i=1}^p \gamma_i \mathbf{1}_{S_i}, \quad \boldsymbol{\pi}_0^+ + \boldsymbol{\pi}_0^- \le \sum_{i=1}^p \gamma_i \mathbf{1}_{S_i} \\
& \boldsymbol{\pi}_i^+, \boldsymbol{\pi}_i^-, \overline{\boldsymbol{\phi}}_i, \underline{\boldsymbol{\phi}}_i \in \mathbb{R}_+^n,\ i = 0,1, \quad \omega \in \mathbb{R}_+, \quad \boldsymbol{\gamma} \in \mathbb{R}_+^p, \quad \tau \in \mathbb{R}.
\end{array}$$

We claim that any feasible solution $(\boldsymbol{\pi}_i^+, \boldsymbol{\pi}_i^-, \overline{\boldsymbol{\phi}}_i, \underline{\boldsymbol{\phi}}_i, \omega, \boldsymbol{\gamma}, \tau)$ to this problem must satisfy $\omega > 0$. Indeed, if there was a feasible solution with $\omega = 0$, then the problem would be unbounded, which is impossible because $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\mathbf{1}_S{}^\top\tilde{\boldsymbol{q}}\big] \ge \sum_{i\in S} \underline{q}_i > -\infty$. We can thus conduct the substitutions $\boldsymbol{\pi}_i^+ \leftarrow \boldsymbol{\pi}_i^+/\omega$, $\boldsymbol{\pi}_i^- \leftarrow \boldsymbol{\pi}_i^-/\omega$, $\overline{\boldsymbol{\phi}}_i \leftarrow \overline{\boldsymbol{\phi}}_i/\omega$, $\underline{\boldsymbol{\phi}}_i \leftarrow \underline{\boldsymbol{\phi}}_i/\omega$, $i = 0,1$, $\boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma}/\omega$ and $\omega \leftarrow 1/\omega$ to obtain the equivalent problem

$$\begin{array}{ll}
\text{minimize} & \tau \\
\text{subject to} & \big(\overline{\boldsymbol{q}} - \boldsymbol{\mu}\big)^\top\overline{\boldsymbol{\phi}}_1 - \big(\underline{\boldsymbol{q}} - \boldsymbol{\mu}\big)^\top\underline{\boldsymbol{\phi}}_1 + \boldsymbol{\nu}^\top\boldsymbol{\gamma} \le \epsilon\omega \\
& \big(\overline{\boldsymbol{q}} - \boldsymbol{\mu}\big)^\top\overline{\boldsymbol{\phi}}_0 - \big(\underline{\boldsymbol{q}} - \boldsymbol{\mu}\big)^\top\underline{\boldsymbol{\phi}}_0 + \boldsymbol{\nu}^\top\boldsymbol{\gamma} + \boldsymbol{\mu}^\top\mathbf{1}_S - \tau \le -(1-\epsilon)\omega \\
& \boldsymbol{\pi}_1^+ - \boldsymbol{\pi}_1^- + \overline{\boldsymbol{\phi}}_1 - \underline{\boldsymbol{\phi}}_1 = \boldsymbol{\pi}_0^+ - \boldsymbol{\pi}_0^- + \overline{\boldsymbol{\phi}}_0 - \underline{\boldsymbol{\phi}}_0 - \mathbf{1}_S \\
& \boldsymbol{\pi}_1^+ + \boldsymbol{\pi}_1^- \le \sum_{i=1}^p \gamma_i \mathbf{1}_{S_i}, \quad \boldsymbol{\pi}_0^+ + \boldsymbol{\pi}_0^- \le \sum_{i=1}^p \gamma_i \mathbf{1}_{S_i} \\
& \boldsymbol{\pi}_i^+, \boldsymbol{\pi}_i^-, \overline{\boldsymbol{\phi}}_i, \underline{\boldsymbol{\phi}}_i \in \mathbb{R}_+^n,\ i = 0,1, \quad \omega \in \mathbb{R}_+, \quad \boldsymbol{\gamma} \in \mathbb{R}_+^p, \quad \tau \in \mathbb{R}.
\end{array}$$

Note that the second constraint in this problem must be binding at optimality. We can thus remove this constraint as well as the epigraphical variable τ to obtain the equivalent problem

$$\begin{array}{ll} \text{minimize} & \left(\overline{\boldsymbol{q}}-\boldsymbol{\mu}\right)^\top\overline{\boldsymbol{\phi}}_0-\left(\underline{\boldsymbol{q}}-\boldsymbol{\mu}\right)^\top\underline{\boldsymbol{\phi}}_0+\boldsymbol{\mu}^\top\mathbf{1}_S+\boldsymbol{\nu}^\top\boldsymbol{\gamma}+(1-\epsilon)\omega \\ \text{subject to} & \left(\overline{\boldsymbol{q}}-\boldsymbol{\mu}\right)^\top\overline{\boldsymbol{\phi}}_1-\left(\underline{\boldsymbol{q}}-\boldsymbol{\mu}\right)^\top\underline{\boldsymbol{\phi}}_1+\boldsymbol{\nu}^\top\boldsymbol{\gamma}\le\epsilon\omega \\ & \boldsymbol{\pi}_1^+-\boldsymbol{\pi}_1^-+\overline{\boldsymbol{\phi}}_1-\underline{\boldsymbol{\phi}}_1=\boldsymbol{\pi}_0^+-\boldsymbol{\pi}_0^-+\overline{\boldsymbol{\phi}}_0-\underline{\boldsymbol{\phi}}_0-\mathbf{1}_S \\ & \boldsymbol{\pi}_1^++\boldsymbol{\pi}_1^-\le\sum_{i=1}^p\gamma_i\mathbf{1}_{S_i},\quad \boldsymbol{\pi}_0^++\boldsymbol{\pi}_0^-\le\sum_{i=1}^p\gamma_i\mathbf{1}_{S_i} \\ & \boldsymbol{\pi}_i^+,\boldsymbol{\pi}_i^-,\overline{\boldsymbol{\phi}}_i,\underline{\boldsymbol{\phi}}_i\in\mathbb{R}^n_+,\ i=0,1,\quad \omega\in\mathbb{R}_+,\quad \boldsymbol{\gamma}\in\mathbb{R}^p_+. \end{array}$$

In this problem, the left-hand side of the first constraint is nonnegative by construction. We thus conclude that the constraint is binding at optimality, which allows us to remove the constraint as well as the variable ω to obtain the equivalent problem

$$\begin{array}{ll} \text{minimize} & \left(\overline{\boldsymbol{q}}-\boldsymbol{\mu}\right)^\top\left(\overline{\boldsymbol{\phi}}_0+\frac{1-\epsilon}{\epsilon}\cdot\overline{\boldsymbol{\phi}}_1\right)-\left(\underline{\boldsymbol{q}}-\boldsymbol{\mu}\right)^\top\left(\underline{\boldsymbol{\phi}}_0+\frac{1-\epsilon}{\epsilon}\cdot\underline{\boldsymbol{\phi}}_1\right)+\frac{1}{\epsilon}\boldsymbol{\nu}^\top\boldsymbol{\gamma}+\boldsymbol{\mu}^\top\mathbf{1}_S \\ \text{subject to} & \left(\boldsymbol{\pi}_0^++\boldsymbol{\pi}_1^-\right)-\left(\boldsymbol{\pi}_0^-+\boldsymbol{\pi}_1^+\right)+\left(\overline{\boldsymbol{\phi}}_0+\underline{\boldsymbol{\phi}}_1\right)-\left(\underline{\boldsymbol{\phi}}_0+\overline{\boldsymbol{\phi}}_1\right)=\mathbf{1}_S \\ & \boldsymbol{\pi}_1^++\boldsymbol{\pi}_1^-\le\sum_{i=1}^p\gamma_i\mathbf{1}_{S_i},\quad \boldsymbol{\pi}_0^++\boldsymbol{\pi}_0^-\le\sum_{i=1}^p\gamma_i\mathbf{1}_{S_i} \\ & \boldsymbol{\pi}_i^+,\boldsymbol{\pi}_i^-,\overline{\boldsymbol{\phi}}_i,\underline{\boldsymbol{\phi}}_i\in\mathbb{R}^n_+,\ i=0,1,\quad \boldsymbol{\gamma}\in\mathbb{R}^p_+. \end{array}$$

The objective function and the second set of constraints imply that larger values of $\boldsymbol{\pi}_i^+$, $\boldsymbol{\pi}_i^-$, $\overline{\boldsymbol{\phi}}_i$ and $\underline{\boldsymbol{\phi}}_i$, $i=0,1$, are all detrimental to the objective function. We can thus assume that $\boldsymbol{\pi}_0^-=\boldsymbol{\pi}_1^+=\underline{\boldsymbol{\phi}}_0=\overline{\boldsymbol{\phi}}_1=\mathbf{0}$ at optimality. This leads to the simplified formulation

$$\begin{array}{ll} \text{minimize} & \left(\overline{\boldsymbol{q}}-\boldsymbol{\mu}\right)^\top\overline{\boldsymbol{\phi}}_0-\frac{1-\epsilon}{\epsilon}\left(\underline{\boldsymbol{q}}-\boldsymbol{\mu}\right)^\top\underline{\boldsymbol{\phi}}_1+\frac{1}{\epsilon}\boldsymbol{\nu}^\top\boldsymbol{\gamma}+\boldsymbol{\mu}^\top\mathbf{1}_S \\ \text{subject to} & \left(\boldsymbol{\pi}_0^++\boldsymbol{\pi}_1^-\right)+\left(\overline{\boldsymbol{\phi}}_0+\underline{\boldsymbol{\phi}}_1\right)=\mathbf{1}_S \\ & \boldsymbol{\pi}_1^-\le\sum_{i=1}^p\gamma_i\mathbf{1}_{S_i},\quad \boldsymbol{\pi}_0^+\le\sum_{i=1}^p\gamma_i\mathbf{1}_{S_i} \\ & \boldsymbol{\pi}_0^+,\boldsymbol{\pi}_1^-,\overline{\boldsymbol{\phi}}_0,\underline{\boldsymbol{\phi}}_1\in\mathbb{R}^n_+,\quad \boldsymbol{\gamma}\in\mathbb{R}^p_+. \end{array}$$

Since the constraints are symmetric in $\boldsymbol{\pi}_0^+$ and $\boldsymbol{\pi}_1^-$, we can replace both variable vectors with a single vector $\boldsymbol{\pi}\in\mathbb{R}^n_+$:

$$\begin{array}{ll} \text{minimize} & \left(\overline{\boldsymbol{q}}-\boldsymbol{\mu}\right)^\top\overline{\boldsymbol{\phi}}_0-\frac{1-\epsilon}{\epsilon}\left(\underline{\boldsymbol{q}}-\boldsymbol{\mu}\right)^\top\underline{\boldsymbol{\phi}}_1+\frac{1}{\epsilon}\boldsymbol{\nu}^\top\boldsymbol{\gamma}+\boldsymbol{\mu}^\top\mathbf{1}_S \\ \text{subject to} & 2\boldsymbol{\pi}+\left(\overline{\boldsymbol{\phi}}_0+\underline{\boldsymbol{\phi}}_1\right)=\mathbf{1}_S \\ & \boldsymbol{\pi}\le\sum_{i=1}^p\gamma_i\mathbf{1}_{S_i} \\ & \boldsymbol{\pi},\overline{\boldsymbol{\phi}}_0,\underline{\boldsymbol{\phi}}_1\in\mathbb{R}^n_+,\quad \boldsymbol{\gamma}\in\mathbb{R}^p_+. \end{array}$$

For a fixed value of $\boldsymbol{\gamma}$, the variable vector $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi} = \min\{\mathbf{1}_S/2 - (\overline{\boldsymbol{\phi}}_0 + \underline{\boldsymbol{\phi}}_1)/2,\ \sum_{i=1}^{p} \gamma_i \mathbf{1}_{S_i}\}$ at optimality. We can thus remove $\boldsymbol{\pi}$ from the problem and obtain the equivalent reformulation

$$\begin{array}{ll} \text{minimize} & (\overline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \overline{\boldsymbol{\phi}}_0 - \dfrac{1-\epsilon}{\epsilon}(\underline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \underline{\boldsymbol{\phi}}_1 + \dfrac{1}{\epsilon}\boldsymbol{\nu}^\top \boldsymbol{\gamma} + \boldsymbol{\mu}^\top \mathbf{1}_S \\ \text{subject to} & \overline{\boldsymbol{\phi}}_0 + \underline{\boldsymbol{\phi}}_1 = \max\left\{\overline{\boldsymbol{\phi}}_0 + \underline{\boldsymbol{\phi}}_1,\ \mathbf{1}_S - 2\displaystyle\sum_{i=1}^{p} \gamma_i \mathbf{1}_{S_i}\right\} \\ & \overline{\boldsymbol{\phi}}_0, \underline{\boldsymbol{\phi}}_1 \in \mathbb{R}^n_+,\ \ \boldsymbol{\gamma} \in \mathbb{R}^p_+. \end{array}$$

The constraint in this problem is equivalent to $\overline{\boldsymbol{\phi}}_0 + \underline{\boldsymbol{\phi}}_1 \geq \mathbf{1}_S - 2\sum_{i=1}^{p} \gamma_i \mathbf{1}_{S_i}$. Since both $\overline{\boldsymbol{\phi}}_0$ and $\underline{\boldsymbol{\phi}}_1$ are penalized in the objective function, we thus conclude that $\overline{\boldsymbol{\phi}}_0 + \underline{\boldsymbol{\phi}}_1 = \left[\mathbf{1}_S - 2\sum_{i=1}^{p} \gamma_i \mathbf{1}_{S_i}\right]_+$ at optimality. The statement then follows since we can assume that $\overline{\boldsymbol{\phi}}_0{}^\top \underline{\boldsymbol{\phi}}_1 = 0$ at optimality. □

Proof of Corollary 2. Under the assumption that $\bigcup_{i=1}^{p-1} S_i = V_C$, problem (13) can be written as

$$\begin{array}{ll} \text{minimize} & \mathbf{1}_S{}^\top \boldsymbol{\mu} + \displaystyle\sum_{i=1}^{p-1} \sum_{j \in S \cap S_i} \hat{q}_j \Big[1 - 2(\gamma_i + \gamma_p)\Big]_+ + \dfrac{1}{\epsilon}\boldsymbol{\nu}^\top \boldsymbol{\gamma} \\ \text{subject to} & \boldsymbol{\gamma} \in [\mathbf{0}, \mathbf{e}/2]. \end{array}$$

For a fixed value of γ_p, an optimal choice of γ_i, $i = 1, \ldots, p-1$, is $\gamma_i = 0$ if $\mathbf{1}_{S \cap S_i}{}^\top \hat{\boldsymbol{q}} \leq \nu_i/(2\epsilon)$ and $\gamma_i = \frac{1}{2} - \gamma_p$ otherwise. The problem thus simplifies to the one-dimensional problem

$$\begin{array}{ll} \text{minimize} & \mathbf{1}_S{}^\top \boldsymbol{\mu} + \displaystyle\sum_{i=1}^{p-1} \min\left\{(1 - 2\gamma_p)\mathbf{1}_{S \cap S_i}{}^\top \hat{\boldsymbol{q}},\ \dfrac{\nu_i}{\epsilon}\left[\dfrac{1}{2} - \gamma_p\right]\right\} + \dfrac{\nu_p}{\epsilon}\gamma_p \\ \text{subject to} & \gamma_p \in [0, 1/2]. \end{array}$$

Since the objective function is concave, its minimum is attained at $\gamma_p^\star \in \{0, 1/2\}$. □

Proof of Corollary 3. The statement immediately follows from Corollary 2 if we use the definitions of the sets S_i in (14) and reorder the summation terms. □

Proof of Theorem 6. To prove the statement, we consider the same distributionally robust CVRP instance as in the proof of Theorem 4, with the exception that the expected demands satisfy $\boldsymbol{\mu} = 4.5\mathbf{e}$ and the demand dispersion is bounded from above by the covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.1 & 0 & -0.05 & 0 \\ 0 & 0.1 & 0 & -0.05 \\ -0.05 & 0 & 0.1 & 0 \\ 0 & -0.05 & 0 & 0.1 \end{bmatrix}.$$

An evaluation of the worst-case values-at-risk for all customer subsets reveals that the set of feasible route sets is exactly the same as in the distributionally robust CVRP instance from the proof of Theorem 4. Thus, we can use the same argument as in that proof to conclude that there is no deterministic CVRP instance that has the same set of feasible route sets. □

Proof of Theorem 7. For a fixed scalar $\tau \in \mathbb{R}$, we have $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\mathbf{1}_S{}^\top \tilde{\boldsymbol{q}}\big] \leq \tau$ if and only if the optimal objective value of the moment problem

$$\begin{aligned}
\text{minimize} \quad & \int_{\mathcal{Q}} \mathbb{I}_{[\mathbf{1}_S{}^\top \boldsymbol{q} \leq \tau]} \, \mathbb{P}(\mathrm{d}\boldsymbol{q}) \\
\text{subject to} \quad & \int_{\mathcal{Q}} \mathbb{P}(\mathrm{d}\boldsymbol{q}) = 1 \\
& \int_{\mathcal{Q}} \boldsymbol{q} \, \mathbb{P}(\mathrm{d}\boldsymbol{q}) = \boldsymbol{\mu} \\
& \int_{\mathcal{Q}} (\boldsymbol{q} - \boldsymbol{\mu})(\boldsymbol{q} - \boldsymbol{\mu})^\top \, \mathbb{P}(\mathrm{d}\boldsymbol{q}) \preceq \boldsymbol{\Sigma} \\
& \mathbb{P} \in \mathcal{M}_+(\mathbb{R}^n)
\end{aligned}$$

is greater than or equal to $1 - \epsilon$. A similar reasoning as in the proof of Theorem 2 shows that this is the case if and only if the optimal objective value of the problem

$$\begin{aligned}
\text{maximize} \quad & \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \langle \boldsymbol{\Sigma}, \boldsymbol{\Gamma} \rangle \\
\text{subject to} \quad & \alpha + \frac{1}{4}(\boldsymbol{\beta} + \underline{\boldsymbol{\phi}}_1 - \overline{\boldsymbol{\phi}}_1)^\top \boldsymbol{\Gamma}^{-1} (\boldsymbol{\beta} + \underline{\boldsymbol{\phi}}_1 - \overline{\boldsymbol{\phi}}_1) \\
& \quad + (\overline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \overline{\boldsymbol{\phi}}_1 - (\underline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \underline{\boldsymbol{\phi}}_1 + \boldsymbol{\mu}^\top \boldsymbol{\beta} \leq 1 \\
& \alpha + \frac{1}{4}(\boldsymbol{\beta} + \underline{\boldsymbol{\phi}}_0 - \overline{\boldsymbol{\phi}}_0 + \mathbf{1}_S \omega)^\top \boldsymbol{\Gamma}^{-1} (\boldsymbol{\beta} + \underline{\boldsymbol{\phi}}_0 - \overline{\boldsymbol{\phi}}_0 + \mathbf{1}_S \omega) \\
& \quad + (\overline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \overline{\boldsymbol{\phi}}_0 - (\underline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \underline{\boldsymbol{\phi}}_0 + \boldsymbol{\mu}^\top (\boldsymbol{\beta} + \mathbf{1}_S \omega) - \tau\omega \leq 0 \\
& \underline{\boldsymbol{\phi}}_i, \overline{\boldsymbol{\phi}}_i \in \mathbb{R}^n_+, \; i = 0, 1, \;\; \omega \in \mathbb{R}_+, \;\; \alpha \in \mathbb{R}, \;\; \boldsymbol{\beta} \in \mathbb{R}^n, \;\; \boldsymbol{\Gamma} \in \mathbb{S}^{n \times n}_+.
\end{aligned}$$

is greater than or equal to $1 - \epsilon$.

We now consider the problem $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\big[\mathbf{1}_S{}^\top \tilde{\boldsymbol{q}}\big]$, which can be formulated as

$$
\begin{aligned}
\text{minimize} \quad & \tau \\
\text{subject to} \quad & \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \langle \boldsymbol{\Sigma}, \boldsymbol{\Gamma} \rangle \geq 1 - \epsilon \\
& \alpha + \frac{1}{4}(\boldsymbol{\beta} + \underline{\boldsymbol{\phi}}_1 - \overline{\boldsymbol{\phi}}_1)^\top \boldsymbol{\Gamma}^{-1} (\boldsymbol{\beta} + \underline{\boldsymbol{\phi}}_1 - \overline{\boldsymbol{\phi}}_1) \\
& \quad + (\overline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \overline{\boldsymbol{\phi}}_1 - (\underline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \underline{\boldsymbol{\phi}}_1 + \boldsymbol{\mu}^\top \boldsymbol{\beta} \leq 1 \\
& \alpha + \frac{1}{4}(\boldsymbol{\beta} + \underline{\boldsymbol{\phi}}_0 - \overline{\boldsymbol{\phi}}_0 + \mathbf{1}_S \omega)^\top \boldsymbol{\Gamma}^{-1} (\boldsymbol{\beta} + \underline{\boldsymbol{\phi}}_0 - \overline{\boldsymbol{\phi}}_0 + \mathbf{1}_S \omega) \\
& \quad + (\overline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \overline{\boldsymbol{\phi}}_0 - (\underline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \underline{\boldsymbol{\phi}}_0 + \boldsymbol{\mu}^\top (\boldsymbol{\beta} + \mathbf{1}_S \omega) - \tau\omega \leq 0 \\
& \underline{\boldsymbol{\phi}}_i, \overline{\boldsymbol{\phi}}_i \in \mathbb{R}^n_+, \; i = 0, 1, \;\; \omega \in \mathbb{R}_+, \;\; \alpha \in \mathbb{R}, \;\; \boldsymbol{\beta} \in \mathbb{R}^n, \;\; \boldsymbol{\Gamma} \in \mathbb{S}^{n \times n}_+, \;\; \tau \in \mathbb{R}.
\end{aligned}
$$

As in the proof of Theorem 5, we can substitute out α and remove the first constraint, conclude that ω is strictly positive and replace all remaining decision variables (except for τ) with their divisions by ω and remove the variables τ and ω to obtain the equivalent problem

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{\epsilon}\langle \boldsymbol{\Sigma}, \boldsymbol{\Gamma} \rangle + \frac{1}{4}(\boldsymbol{\beta} + \underline{\boldsymbol{\phi}}_0 - \overline{\boldsymbol{\phi}}_0 + \mathbf{1}_S)^\top \boldsymbol{\Gamma}^{-1} (\boldsymbol{\beta} + \underline{\boldsymbol{\phi}}_0 - \overline{\boldsymbol{\phi}}_0 + \mathbf{1}_S) \\
& + \frac{1}{4} \cdot \frac{1-\epsilon}{\epsilon} (\boldsymbol{\beta} + \underline{\boldsymbol{\phi}}_1 - \overline{\boldsymbol{\phi}}_1)^\top \boldsymbol{\Gamma}^{-1} (\boldsymbol{\beta} + \underline{\boldsymbol{\phi}}_1 - \overline{\boldsymbol{\phi}}_1) \\
& + (\overline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \left[\overline{\boldsymbol{\phi}}_0 + \frac{1-\epsilon}{\epsilon} \cdot \overline{\boldsymbol{\phi}}_1 \right] - (\underline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \left[\underline{\boldsymbol{\phi}}_0 + \frac{1-\epsilon}{\epsilon} \cdot \underline{\boldsymbol{\phi}}_1 \right] + \boldsymbol{\mu}^\top \mathbf{1}_S \\
\text{subject to} \quad & \underline{\boldsymbol{\phi}}_i, \overline{\boldsymbol{\phi}}_i \in \mathbb{R}^n_+, \; i = 0, 1, \;\; \boldsymbol{\beta} \in \mathbb{R}^n, \;\; \boldsymbol{\Gamma} \in \mathbb{S}^{n \times n}_+.
\end{aligned}
$$

We can now replace $\boldsymbol{\beta}$ with its optimal value $\boldsymbol{\beta} = \epsilon(\overline{\boldsymbol{\phi}}_0 - \underline{\boldsymbol{\phi}}_0 - \mathbf{1}_S) + (1-\epsilon)(\overline{\boldsymbol{\phi}}_1 - \underline{\boldsymbol{\phi}}_1)$ from the first-order unconstrained optimality condition to obtain the equivalent reformulation

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{\epsilon}\langle \boldsymbol{\Sigma}, \boldsymbol{\Gamma} \rangle + \frac{1}{4}(1-\epsilon)(\overline{\boldsymbol{\phi}}_1 - \underline{\boldsymbol{\phi}}_1 + \underline{\boldsymbol{\phi}}_0 - \overline{\boldsymbol{\phi}}_0 + \mathbf{1}_S)^\top \boldsymbol{\Gamma}^{-1} (\overline{\boldsymbol{\phi}}_1 - \underline{\boldsymbol{\phi}}_1 + \underline{\boldsymbol{\phi}}_0 - \overline{\boldsymbol{\phi}}_0 + \mathbf{1}_S) \\
& + (\overline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \left[\overline{\boldsymbol{\phi}}_0 + \frac{1-\epsilon}{\epsilon} \cdot \overline{\boldsymbol{\phi}}_1 \right] - (\underline{\boldsymbol{q}} - \boldsymbol{\mu})^\top \left[\underline{\boldsymbol{\phi}}_0 + \frac{1-\epsilon}{\epsilon} \cdot \underline{\boldsymbol{\phi}}_1 \right] + \boldsymbol{\mu}^\top \mathbf{1}_S \\
\text{subject to} \quad & \underline{\boldsymbol{\phi}}_i, \overline{\boldsymbol{\phi}}_i \in \mathbb{R}^n_+, \; i = 0, 1, \;\; \boldsymbol{\Gamma} \in \mathbb{S}^{n \times n}_+.
\end{aligned}
$$

Since $\underline{\boldsymbol{\phi}}_i$ and $\overline{\boldsymbol{\phi}}_i$, $i = 0, 1$, are all penalized in the second row of the objective function, we can assume that $\overline{\boldsymbol{\phi}}_1{}^\top \underline{\boldsymbol{\phi}}_0 = 0$ and $\underline{\boldsymbol{\phi}}_1{}^\top \overline{\boldsymbol{\phi}}_0 = 0$ at optimality. This leads to the simplified formulation

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{\epsilon}\langle \boldsymbol{\Sigma}, \boldsymbol{\Gamma} \rangle + \frac{1}{4}(1-\epsilon)(\boldsymbol{\phi}^+ - \boldsymbol{\phi}^- + \mathbf{1}_S)^\top \boldsymbol{\Gamma}^{-1} (\boldsymbol{\phi}^+ - \boldsymbol{\phi}^- + \mathbf{1}_S) \\
& + \boldsymbol{q}^{+\top} \boldsymbol{\phi}^+ + \boldsymbol{q}^{-\top} \boldsymbol{\phi}^- + \boldsymbol{\mu}^\top \mathbf{1}_S \\
\text{subject to} \quad & \boldsymbol{\phi}^+, \boldsymbol{\phi}^- \in \mathbb{R}^n_+, \;\; \boldsymbol{\Gamma} \in \mathbb{S}^{n \times n}_+,
\end{aligned}
$$

where $\boldsymbol{q}^+ = \min\left\{\frac{1-\epsilon}{\epsilon}(\overline{\boldsymbol{q}} - \boldsymbol{\mu}),\, -(\underline{\boldsymbol{q}} - \boldsymbol{\mu})\right\}$ and $\boldsymbol{q}^- = \min\left\{-\frac{1-\epsilon}{\epsilon}(\underline{\boldsymbol{q}} - \boldsymbol{\mu}),\, \overline{\boldsymbol{q}} - \boldsymbol{\mu}\right\}$. We apply an epigraph reformulation to obtain the equivalent problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{\epsilon}\langle \boldsymbol{\Sigma}, \boldsymbol{\Gamma}\rangle + \frac{1}{4}(1-\epsilon)\kappa + {\boldsymbol{q}^+}^\top \boldsymbol{\phi}^+ + {\boldsymbol{q}^-}^\top \boldsymbol{\phi}^- + \boldsymbol{\mu}^\top \mathbf{1}_S \\ \text{subject to} & \kappa \geq (\boldsymbol{\phi}^+ - \boldsymbol{\phi}^- + \mathbf{1}_S)^\top \boldsymbol{\Gamma}^{-1}(\boldsymbol{\phi}^+ - \boldsymbol{\phi}^- + \mathbf{1}_S) \\ & \boldsymbol{\phi}^+, \boldsymbol{\phi}^- \in \mathbb{R}^n_+, \;\; \boldsymbol{\Gamma} \in \mathbb{S}^{n\times n}_+, \kappa \in \mathbb{R}, \end{array} \tag{32}$$

and an application of Schur's complement allows us to reformulate the constraint in (32) as

$$\begin{pmatrix} \kappa & (\boldsymbol{\phi}^+ - \boldsymbol{\phi}^- + \mathbf{1}_S)^\top \\ (\boldsymbol{\phi}^+ - \boldsymbol{\phi}^- + \mathbf{1}_S) & \boldsymbol{\Gamma} \end{pmatrix} \succeq \mathbf{0}.$$

Strong conic duality, which holds since the primal problem (32) is strictly feasible, implies that (32) attains the same optimal objective value as its associated dual problem, which—after some minor simplifications—can be expressed as

$$\begin{array}{ll} \text{maximize} & {\mathbf{1}_S}^\top \boldsymbol{\mu} - 2 \cdot {\mathbf{1}_S}^\top \boldsymbol{\varphi} \\ \text{subject to} & \theta \leq \frac{1}{4}(1-\epsilon) \\ & \boldsymbol{\varphi} \in \left[-\boldsymbol{q}^-/2,\, \boldsymbol{q}^+/2\right] \\ & \boldsymbol{\Lambda} \preceq \frac{1}{\epsilon}\boldsymbol{\Sigma} \\ & \begin{pmatrix} \theta & \boldsymbol{\varphi}^\top \\ \boldsymbol{\varphi} & \boldsymbol{\Lambda} \end{pmatrix} \in \mathbb{S}_+^{(n+1)\times(n+1)} \\ & \theta \in \mathbb{R}, \;\; \boldsymbol{\varphi} \in \mathbb{R}^n, \;\; \boldsymbol{\Lambda} \in \mathbb{S}^{n\times n}. \end{array}$$

Applying Schur's complement to the last constraint in this problem shows that the last two constraints are satisfied if and only if $\frac{1}{\theta}\boldsymbol{\varphi}\boldsymbol{\varphi}^\top \preceq \boldsymbol{\Lambda} \preceq \frac{1}{\epsilon}\boldsymbol{\Sigma}$ for some $\boldsymbol{\Lambda} \in \mathbb{S}^{n\times n}$, that is, if and only if $\frac{1}{\theta}\boldsymbol{\varphi}\boldsymbol{\varphi}^\top \preceq \frac{1}{\epsilon}\boldsymbol{\Sigma}$. Since the first constraint imposes the only upper bound on θ, we can replace θ with the right-hand side of that constraint, $\theta = \frac{1}{4}(1-\epsilon)$, to obtain the equivalent reformulation

$$\begin{array}{ll} \text{maximize} & {\mathbf{1}_S}^\top \boldsymbol{\mu} - 2 \cdot {\mathbf{1}_S}^\top \boldsymbol{\varphi} \\ \text{subject to} & \boldsymbol{\varphi} \in \left[-\boldsymbol{q}^-/2,\, \boldsymbol{q}^+/2\right] \\ & \boldsymbol{\varphi}\boldsymbol{\varphi}^\top \preceq \frac{1-\epsilon}{4\epsilon} \cdot \boldsymbol{\Sigma} \\ & \boldsymbol{\varphi} \in \mathbb{R}^n. \end{array}$$

Finally, two further applications of Schur's complement yield

$$\frac{1-\epsilon}{4\epsilon} \cdot \boldsymbol{\Sigma} - \boldsymbol{\varphi}\boldsymbol{\varphi}^\top \succeq \mathbf{0} \iff \begin{pmatrix} 1 & \boldsymbol{\varphi}^\top \\ \boldsymbol{\varphi} & \frac{1-\epsilon}{4\epsilon} \cdot \boldsymbol{\Sigma} \end{pmatrix} \succeq \mathbf{0} \iff 1 - \frac{4\epsilon}{1-\epsilon} \cdot \boldsymbol{\varphi}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\varphi} \geq 0,$$

which simplifies the problem to

$$\begin{aligned}
\text{maximize} \quad & \mathbf{1}_S{}^\top \boldsymbol{\mu} - 2 \cdot \mathbf{1}_S{}^\top \boldsymbol{\varphi} \\
\text{subject to} \quad & \boldsymbol{\varphi} \in \left[-\boldsymbol{q}^-/2,\, \boldsymbol{q}^+/2\right] \\
& \boldsymbol{\varphi}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\varphi} \leq \frac{1-\epsilon}{4\epsilon} \\
& \boldsymbol{\varphi} \in \mathbb{R}^n.
\end{aligned}$$

The statement now follows from the variable transformation $\boldsymbol{q} \leftarrow -2\boldsymbol{\varphi}$. □

Proof of Corollary 4. The statement follows from Theorem 7 as well as an adaptation of Lemma 2 in Pessoa and Poss (2015) that replaces the box constraints $\boldsymbol{q} \in [-\mathbf{e}, +\mathbf{e}]$ from that paper with $\boldsymbol{q} \in [\boldsymbol{q}^\ell,\, \boldsymbol{q}^\mathrm{u}]$ and the ellipsoidal constraint $\boldsymbol{q}^\top \boldsymbol{q} \leq \kappa_2$ with $\boldsymbol{q}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{q} \leq \frac{1-\epsilon}{\epsilon}$. □

References

Y. Adulyasak and P. Jaillet. Models and algorithms for stochastic and robust vehicle routing with deadlines. *Transportation Science*, 50(2):363–761, 2016.

P. Augerat, J. M. Belenguer, E. Benavent, A. Corberán, and D. Naddef. Separating capacity constraints in the CVRP using tabu search. *European Journal of Operational Research*, 106(2–3):546–557, 1998.

C. Bandi and D. Bertsimas. Tractable stochastic analysis in high dimensions via robust optimization. *Mathematical Programming*, 134(1):23–70, 2012.

A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

A. Beck and M. Teboulle. Smoothing and first order methods: A unified framework. *SIAM Journal on Optimization*, 22(2):557–580, 2012.

A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.

A. Ben-Tal, D. den Hertog, A. de Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Operations Research*, 59(2):341–357, 2013.

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

D. Bertsimas and D. Simchi-Levi. A new generation of vehicle routing research: Robust algorithms, addressing uncertainty. *Operations Research*, 44(2):286–304, 1996.

D. Bertsimas, V. Gupta, and N. Kallus. Data-driven robust optimization. *Mathematical Programming*, 167 (2):235–292, 2018.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

J. G. Carlsson and E. Delage. Robust partitioning for stochastic multivehicle routing. *Operations Research*, 61(3):546–557, 2013.

J. G. Carlsson, M. Behroozi, and K. Mihic. Wasserstein distance and the distributionally robust TSP. Forthcoming in Operations Research, 2017.

G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Thomson Learning, 2nd edition, 2002.

M. R. Chernick. *Bootstrap Methods: A Guide for Practitioners and Researchers*. Wiley-Blackwell, 2nd edition, 2007.

J. F. Cordeau, G. Laporte, M. W. P. Savelsbergh, and D. Vigo. Vehicle routing. In C. Barnhart and G. Laporte, editors, *Handbooks in Operations Research and Management Science: Transportation*, volume 14, pages 367–428. 2006.

G. B. Dantzig and J. H. Ramser. The truck dispatching problem. *Management Science*, 6(1):80–91, 1959.

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

J. Dhaene, M. Denuit, M. J. Goovaerts, R. Kaas, and D. Vyncke. The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, 31(1):3–33, 2002.

B. D. Díaz. The VRP web. http://www.bernabe.dorronsoro.es/vrp/, 2006. Online; accessed August 2018.

T. Dinh, R. Fukasawa, and J. Luedtke. Exact algorithms for the chance-constrained vehicle routing problem. pages 89–101, 2016.

T. Dinh, R. Fukasawa, and J. Luedtke. Exact algorithms for the chance-constrained vehicle routing problem. *Mathematical Programming*, pages 1–34, 2017. doi: https://doi.org/10.1007/s10107-017-1151-6.

M. Dror and P. Trudeau. Stochastic vehicle routing with modified savings algorithm. *European Journal of Operational Research*, 23(2):228–235, 1986.

M. Dror, G. Laporte, and F. V. Louveaux. Vehicle routing with stochastic demands and restricted failures. *Methods and Models of Operations Research*, 37:273–283, 1993.

M. Dyer and L. Stougie. Computational complexity of stochastic programming problems. *Mathematical Programming*, 106(3):423–432, 2006.

L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.

P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* (forthcoming), 2017.

A. Flajolet, S. Blandin, and P. Jaillet. Robust adaptive routing under uncertainty. *Operations Research*, 66 (1):210–229, 2018.

R. Fukasawa, H. Longo, J. Lysgaard, M. Poggi de Aragão, M. Reis, E. Uchoa, and R. F. Werneck. Robust branch-and-cut-and-price for the capacitated vehicle routing problem. *Mathematical Programming*, 106 (3):491–511, 2006.

M. Gendreau, G. Laporte, and R. Séguin. Stochastic vehicle routing. *European Journal of Operational Research*, 88(1):3–12, 1996.

B. L. Golden, S. Raghavan, and E. A. Wasil, editors. *The Vehicle Routing Problem: Latest Advances and New Challenges*. Springer, New York, NY, USA, 2008.

B.L. Golden and J.R. Yee. A framework for probabilistic vehicle routing. *AIIE Transactions*, 11(2):109–112, 1979.

D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28 (1):1–38, 2003.

C.E. Gounaris, W. Wiesemann, and C.A. Floudas. The robust capacitated vehicle routing problem under demand uncertainty. *Operations Research*, 61(3):677–693, 2013.

G. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*, 65(3):751–767, 2017.

G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151(1): 35–62, 2015.

G. A. Hanasusanto, D. Kuhn, and W. Wiesemann. A comment on "Computational complexity of stochastic programming problems". *Mathematical Programming*, 159(1):557–569, 2016.

Z. Hu and L. J. Hong. Kullback-leibler divergence constrained distributionally robust optimization. Available on Optimization Online, 2013.

P. Jaillet, J. Qi, and M. Sim. Routing optimization under uncertainty. *Operations Research*, 64(1):186–200, 2016.

R. Jiang and Y. Guan. Risk-averse two-stage stochastic program with distributional ambiguity. Available on Optimization Online, 2015.

G. Laporte. Fifty years of vehicle routing. *Transportation Science*, 43(4):408–416, 2009.

G. Laporte, Y. Norbert, and M. Desrochers. Optimal routing under capacity and distance restrictions. *Operations Research*, 33(5):1050–1073, 1985.

G. Laporte, F. Louveaux, and H. Mercure. The vehicle routing problem with stochastic travel times. *Transportation Science*, 26(3):161–170, 1992.

Q. Li, A. M. C. So, and W.-K. Ma. Distributionally robust chance-constrained transmit beamforming for multiuser MISO downlink. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3479–3483, 2014.

J. Luedtke and S. Ahmed. A sample average approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19(2):674–699, 2008.

J. Lysgaard, A. N. Letchford, and R. W. Eglese. A new branch-and-cut algorithm for the capacitated vehicle routing problem. *Mathematical Programming*, 100(2):423–445, 2004.

F. Meng, J. Qi, M. Zhang, J. Ang, S. Chu, and M. Sim. A robust optimization model for managing elective admission in a public hospital. *Operations Research*, 63(6):1452–1467, 2015.

A. Nemirovski. On safe tractable approximations of chance constraints. *European Journal of Operational Research*, 219(3):707–718, 2012.

B. O'Donoghue and E. Candès. Adaptive restart for accelerated gradient schemes. *Foundations of Computational Mathematics*, 15(3):715–732, 2015.

D. Pecin, A. Pessoa, M. Poggi, and E. Uchoa. Improved branch-cut-and-price for capacitated vehicle routing. *Mathematical Programming Computation*, 9(1):61–100, 2017.

A. A. Pessoa and M. Poss. Robust network design with uncertain outsourcing cost. *INFORMS Journal on Computing*, 27(3):507–524, 2015.

G. C. Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In S. P. Uryasev, editor, *Probabilistic Constrained Optimization*, pages 272–281. Kluwer Academic Publishers, 2000.

T. Pham-Gia and T.L. Hung. The mean and median absolute deviations. *Mathematical and Computer Modelling*, 34(7-8):921–936, 2001.

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

R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer, 1997.

J. Segers. On the asymptotic distribution of the mean absolute deviation about the mean. Available on arXiv, 2014.

F. Semet, P. Toth, and D. Vigo. Classical exact algorithms for the capacitated vehicle routing problem. In P. Toth and D. Vigo, editors, *Vehicle Routing: Problems, Methods, and Applications*, pages 37–57. MOS-SIAM, 2nd edition, 2014.

A. Shapiro. On duality theory of conic linear problems. In M. A. Goberna and M. A. López, editors, *Semi-Infinite Programming*, pages 135–165. Kluwer Academic Publishers, 2001.

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory.* MOS-SIAM, 2nd edition, 2014.

W. R. Stewart and B. L. Golden. Stochastic vehicle routing: A comprehensive approach. *European Journal of Operational Research*, 14(4):371–385, 1983.

I. Sungur and F. Ordonez. A robust optimization approach for the capacitated vehicle routing problem with demand uncertainty. *IIE Transactions*, 40(5):509–523, 2008.

P. Toth and D. Vigo, editors. *Vehicle Routing: Problems, Methods, and Applications.* MOS-SIAM, 2nd edition, 2014.

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

W. Xie and S. Ahmed. On deterministic reformulations of distributionally robust joint chance constrained optimization problems. Forthcoming in SIAM Journal on Optimization, 2017.

W-H. Yang, K. Mathur, and R.H. Ballou. Stochastic vehicle routing problem with restocking. *Transportation Science*, 34(1):99–112, 2000.

Y. Zhang, R. Baldacci, M. Sim, and J. Tang. Routing optimization with time windows under uncertainty. Forthcoming in Mathematical Programming, 2018.

C. Zhao and Y. Guan. Data-driven risk-averse stochastic optimization with Wasserstein metric. *Operations Research Letters*, 46(2):262–267, 2018.

C. Zhao and R. Jiang. Distributionally robust contingency-constrained unit commitment. *IEEE Transactions on Power Systems*, 33(1):94–102, 2018.

S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1–2):167–198, 2013.