Effective Scenarios in Distributionally Robust Optimization with Wasserstein Distance

Chennan Zhou*[1] and Güzin Bayraksan†[1]

[1]Integrated Systems Engineering Department, The Ohio State University, Columbus OH, USA

Abstract

This paper studies effective scenarios in Distributionally Robust Optimization (DRO) problems defined on a finite number of realizations (also called scenarios) of the uncertain parameters. Effective scenarios are critical scenarios in DRO in the sense that their removal from the support of the considered distributions alters the optimal value. Ineffective scenarios are those whose removal do not alter the optimal value. In this paper, we first link the effectiveness of a scenario to its worst-case probability being always positive or uniquely zero under a general distance-based ambiguity set. We then narrow our focus to DROs with ambiguity sets formed via the Wasserstein distance (denoted DRO-W), and we provide easy-to-check sufficient conditions to identify the effectiveness of scenarios for this class of problems. When the Wasserstein distance is equivalent to the total variation distance (i.e., when the transportation cost between scenarios is zero if they are the same and one if they are different), the easy-to-check conditions for DRO-W presented in this paper recover the ones presented in the literature for DRO formed via the total variation distance as a special case. The numerical findings highlight the relationship between scenario effectiveness and the attributes of the transportation cost between scenarios that constitute the Wasserstein distance, revealing useful insights.

Keywords: Data-driven stochastic programming, distributionally robust optimization, Wasserstein distance, scenario analysis

1 Introduction

Distributionally Robust Optimization (DRO) has emerged as an alternative approach to traditional Stochastic Optimization (SO) for modeling decision making problems under uncertainty. Unlike traditional SO, DRO does not assume that the probability distribution of the underlying uncertainty is known. Instead, it forms an ambiguity set of probability distributions that is believed to contain the true distribution. This is especially useful for real-world problems where there is some data, but not all is known. Then, DRO optimizes the worst-case expectation from this ambiguity set of distributions. This way, DRO can effectively connect data with decision-making, addressing the risk and uncertainty arising from the unknown or ambiguous underlying distribution of uncertainty.

*zhou.2029@osu.edu

†bayraksan.1@osu.edu; Corresponding author

There are different ways to model the distributional ambiguity in DRO; see, e.g., the surveys [28, 41]. One common approach is *moment-based*, in which the ambiguity set contains all distributions whose moments (typically the first and second moments) satisfy certain properties [e.g., 16, 47]. Another popular approach is *distance-based*, where the ambiguity set is constructed by considering all probability distributions sufficiently close to a nominal probability distribution according to some measure of similarity or distance between distributions. A number of such measures/distances have been used, including the Prokhorov metric [19], ϕ-divergences [7, 5] (or specific cases of ϕ-divergences such as the Kullback-Leibler divergence [12, 23], χ^2 or modified χ^2 distances [26, 37], the total variation distance [25, 39]), and the Wasserstein distance [e.g., 20, 27, 36, 48, 33].

Starting with [36], there has been a significant growth in using the Wasserstein distance for modeling the distributional ambiguity in DROs. While we avoid an extensive literature review and direct the readers to [28, 41] for an in-depth exploration of the topic, we briefly mention that DROs formed with the Wasserstein distance (denoted DRO-W for short) can be difficult to solve in certain settings. Therefore, one line of work has focused on devising tractable reformulations [e.g., 20, 51, 21, 9, 49], another has investigated solution algorithms [e.g., 31, 18] or bounding techniques [e.g., 6, 14]. Others have studied DRO-W under chance constraints [e.g., 13] or in the presence of decision rules [e.g., 10]. DRO-W has also been successfully applied to solve problems arising in machine learning [e.g., 29, 30, 44] and energy systems [e.g., 50, 3], among others.

The focus of this paper is different from the aforementioned literature on DRO. Specifically, it studies *effective scenarios*, which are the scenarios that cause a change in the optimal value if removed from the support of the distributions in the ambiguity set. *Ineffective scenarios* can be removed without causing a change in the optimal value (formal definitions will be provided in Section 2.3). The study of effective scenarios can therefore be viewed as a type of sensitivity analysis with respect to the support of the probability distributions considered in a DRO.

The concepts of effective and ineffective scenarios were first defined in [39], where such scenarios were studied in detail for DROs formed via the total variation distance (denoted DRO-TV for short) with finite support. Later, these concepts were extended to continuous distributions in the context of newsvendor problems [40] and to multistage DRO problems under a finite stochastic process [38]. Both of these works again mainly focused on DRO-TV. Under certain conditions such as convex compact ambiguity sets and real-valued cost functions, DRO is equivalent to a risk-averse SO with a coherent risk measure [4, 45]. In cases where such a risk equivalence holds, [45] related the existence of ineffective scenarios to the corresponding risk measure of DRO not being strictly monotone, and [2] used effective scenarios for scenario reduction in a class of risk-averse SO. We note that related concepts of identifying important portions of a problem have appeared in other fields such as machine learning [e.g., 1] and for other classes of optimization problems [e.g., 11]; we refer the interested readers to the discussion in [38, Introduction].

The study of effective scenarios can yield multiple benefits. First, such a study reveals how the uncertainties affect optimization, helping decision makers (DMs) gain insights into their problems.

In multistage settings, it can also help reveal the time-dynamic aspects of critical uncertainties (for instance, which data uncertainty is more critical in the short term versus in the long term) [38]. It can direct DMs to collect more data around effective scenarios by alerting them to the importance of such scenarios for their problems. Effective scenarios have also been used to size ambiguity sets by relating effective scenarios to a DM's risk aversion for a class of inventory problems [40].

Sets of ineffective scenarios, on the other hand, can help with *problem-specific* scenario reduction. This is different than classical scenario reduction methods in SO, which minimize a distance between the original and the reduced probability distributions without using any other information specific to the problem [e.g., 35, 17]. In contrast, effective/ineffective scenarios explicitly utilize the specific problem's structure and its solution to determine which scenarios alter the optimal value. If a set of ineffective scenarios can be identified and removed from the problem, a much smaller problem with the same optimal value can be obtained. Based on this idea, the authors of [2] estimated the set of ineffective scenarios for problem-specific scenario reduction in risk-averse SO problems with Conditional Value-at-Risk (CVaR) objectives. They also showed conditions under which their estimated set of ineffective scenarios converge to the true set. In addition, examining effective/ineffective scenarios might uncover solution methods that solve DROs with better efficiency. As an example, [52] approximated the sets of effective and ineffective scenarios during a decomposition algorithm for solving DRO-TV in order to accelerate the algorithm by focusing the algorithm's effort on the effective scenarios.

Despite the above potential benefits, except for several basic properties established in [39, 38, 45], not much is known about effective/ineffective scenarios for general ambiguity sets outside of DRO-TV or risk-averse SO optimizing CVaR. One of the main contributions of this paper is that, it provides for the first time sufficient conditions on the optimal probabilities of DRO for a scenario to be categorized as effective or ineffective for *general* distance-based ambiguity sets. Such conditions can be immediately used for some DROs. For others, it provides a structured framework to study effective scenarios. We adopt this new framework to study effective scenarios in DRO-W.

One way to assess the effectiveness of a scenario is to remove it from the support of the considered distributions and solve the DRO problem again with this revised ambiguity set, and repeat this process for all scenarios. While such a method is guaranteed to determine the effectiveness of all scenarios, it is very costly. Therefore, easy-to-check conditions are desirable. The easy-to-check conditions for DRO-TV established in [39] mainly rely on an analysis of the subgradients associated with the objective function *specific* to DRO-TV (particularly, its equivalent risk-averse objective function, which is a convex combination of CVaR and worst-case cost). We centralize and generalize this analysis for convex DROs with general ambiguity sets defined on a finite support. This method not only extends the generality of our findings but also streamlines the proof structure. Particularly, this new framework removes the need to know the explicit form of the risk measure, which is often difficult to determine. It allows us to establish easy-to-check sufficient conditions for effectiveness of scenarios in DRO-W, whose induced risk measure is unknown.

The easy-to-check conditions are computationally cheap. However, they may not be able to determine the effectiveness of all scenarios. That said, in this paper, we uncover cases under which the effectiveness of *all* scenarios in DRO-W can be verified by our proposed easy-to-check conditions. It is well known that when the transportation cost between scenarios is one if the scenarios are different and zero if they are the same, the Wassestein distance becomes equivalent to the total variation distance. We show that the conditions established in this paper for DRO-W recover the ones proposed in the literature for DRO-TV as a special case; thereby significantly generalizing the results in [39]. Numerical results show that the proposed easy-to-check conditions often identify the effectiveness of all or majority of scenarios. The numerical experiments also highlight how the range and variability of the transportation cost between scenarios used to build the Wasserstein distance can have an impact on the behavior DRO-W, revealing further insights.

The rest of this paper is outlined as follows. In Section 2, we present the class of DRO problems considered in this study, review commonly used notation, and introduce background on DRO-W and effective scenarios. Section 3 characterizes conditions on the optimal probability distributions in general distance-based DROs sufficient to classify a scenario as effective or ineffective. Then, Section 4 focuses on DRO-W and establishes easy-to-check conditions to identify the effectiveness of scenarios for this class of problems. These easy-to-check conditions are shown to subsume the earlier ones established for DRO-TV as a special case in Section 5. Section 6 presents the numerical experiments, and Section 7 ends the paper with conclusions and discussion of future work.

2 Setting, Basic Notation, and Background

2.1 Problem Class

We consider convex DRO problems where the uncertainty is represented on a finite number of realizations, called *scenarios*. The set of scenarios is denoted by $\Omega = \{\omega_1, \ldots, \omega_{|\Omega|}\}$, where $|\mathcal{B}|$ is the cardinality of a given set $\mathcal{B}$. A generic scenario from this set is denoted by ω_i, $i = 1, 2, \ldots, |\Omega|$, or we simply use ω for this purpose. Throughout, we also use ω_j, $j = 1, \ldots, |\Omega|$ to denote a specific scenario (e.g., an effective scenario) or generic scenario depending on the context. Decisions x belong to a feasibility set $\mathcal{X} \subset \mathbb{R}^{d_x}$ (i.e., $x \in \mathcal{X}$), which is assumed to be a nonempty convex compact set. DRO can then be formulated as

$$\min_{x\in\mathcal{X}} \left\{ f(x) := \max_{\mathbf{p}\in\mathcal{P}} \mathbb{E}_{\mathbf{p}} \left[\mathbf{h}(x)\right] \right\}, \tag{1}$$

where, for a given $x \in \mathcal{X}$, $\mathbf{h}(x)$ denotes a random variable that takes value $h_{\omega_i}(x)$ for each scenario $\omega_i \in \Omega$. The so-called *cost functions* at scenario $\omega_i \in \Omega$, $h_{\omega_i}(\cdot) : \mathcal{X} \to \mathbb{R}$, are assumed to be real-valued and convex on an open set containing $\mathcal{X}$. This implies that $h_{\omega_i}(\cdot)$ are Lipschitz continuous [34, Theorem 1.26] and subdifferentiable [42, Theorem 23.4] on $\mathcal{X}$ for all $\omega_i \in \Omega$. Continuity, along with $\mathcal{X}$ being compact and $|\Omega| < \infty$, implies that $h_{\omega_i}(x)$ are bounded for every $\omega_i \in \Omega$ and there exists a constant $C < \infty$ such that $|h_{\omega_i}(x)| < C$ for all $x \in \mathcal{X}, \omega_i \in \Omega$.

Set $\mathcal{P}$ in DRO problem (1) denotes the *ambiguity set of distributions* (or simply the *ambiguity set*), which is a subset of all probability distributions on Ω. We refer to the inner maximization problem in (1) as the *worst-case expected problem* at $x \in \mathcal{X}$. Here, $\mathbf{p}$ represents a probability vector, with probability of scenario $\omega_i \in \Omega$ denoted as p_{ω_i}. Throughout the paper, we adopt bold notation to denote vectors or random variables that depend on scenarios $\omega_i \in \Omega$ like $\mathbf{p}$ and $\mathbf{h}(x)$. Because we assume $|\Omega| < \infty$, the expectation in (1), taken with respect to $\mathbf{p}$, can be written as $\sum_{\omega_i \in \Omega} p_{\omega_i} h_{\omega_i}(x)$. We refer to an optimal solution to the worst-case expected problem as a *worst-case distribution* and denote it as $\mathbf{p}^*$, where the components of $\mathbf{p}^*$ represent the worst-case probability $p^*_{\omega_i}$ at each scenario $\omega_i \in \Omega$. We suppress the dependence of $\mathbf{p}^*$ on x for simplicity. Assuming $\mathcal{P}$ is nonempty, DRO problem (1) satisfying the other assumptions stated above has a finite optimal value and finite optimal solution achieved on $\mathcal{X}$.

Before we review the Wasserstein distance, let us first introduce a general distance-based ambiguity set. Let $\mathbf{q}$ denote a *nominal distribution*, with q_{ω_i} representing the probability of scenario $\omega_i \in \Omega$. Such a nominal distribution can be obtained, for instance, by using historical data. Let $\Delta(\mathbf{p}, \mathbf{q})$ denote a measure of similarity or distance between two distributions $\mathbf{p}$ and $\mathbf{q}$. Then, a distance-based ambiguity set is formed by

$$\mathcal{P} := \left\{ \Delta(\mathbf{p}, \mathbf{q}) \le \rho, \ \sum_{\omega_i \in \Omega} p_{\omega_i} = 1, \quad p_{\omega_i} \ge 0 \ \forall \omega_i \in \Omega \right\}, \tag{2}$$

where ρ is called the *radius* of the ambiguity set, also referred to as the *level of robustness*. As ρ increases, DRO in (1) typically becomes more robust.

We end this section with notation that is frequently used throughout the paper. For a given set $\mathcal{F} \subseteq \Omega$, we denote the total nominal probability of that set as $q(\mathcal{F}) := \sum_{\omega \in \mathcal{F}} q_\omega$. Similarly, we denote the total worst-case probability of set $\mathcal{F}$ as $p^*(\mathcal{F}) := \sum_{\omega \in \mathcal{F}} p^*_\omega$, where we again suppress the dependence of $p^*(\mathcal{F})$ on x for simplicity.

2.2 Background on DRO-W

The Wasserstein distance between two finitely supported distributions $\mathbf{p}$ and $\mathbf{q}$ on Ω, denoted $W(\mathbf{p}, \mathbf{q})$, can be formulated as the following optimal transportation problem:

$$W(\mathbf{p}, \mathbf{q}) := \min_{\boldsymbol{\gamma} \ge 0} \left\{ \sum_{\omega_i, \omega_j \in \Omega} c_{\omega_i \omega_j} \gamma_{\omega_i \omega_j} : \ \sum_{\omega_i \in \Omega} \gamma_{\omega_i \omega_j} = p_{\omega_j} \ \ \forall \omega_j \in \Omega, \ \sum_{\omega_j \in \Omega} \gamma_{\omega_i \omega_j} = q_{\omega_i} \quad \forall \omega_i \in \Omega \right\},$$

where $\gamma_{\omega_i \omega_j}$ represents the probability mass that will be transported from scenario ω_i to scenario ω_j, $\boldsymbol{\gamma}$ represents the vector containing all $\gamma_{\omega_i \omega_j}$, and $c_{\omega_i \omega_j}$ denotes the *transportation cost* of moving probability mass from ω_i to ω_j. We assume the transportation cost $c_{\omega_i \omega_j}$ is a distance. That is, it satisfies (i) $c_{\omega_i \omega_i} = 0$ for all $\omega_i \in \Omega$, (ii) positivity (i.e., $c_{\omega_i \omega_j} > 0$ for all $\omega_i, \omega_j \in \Omega$ such that $\omega_i \neq \omega_j$), (iii) symmetry (i.e., $c_{\omega_i \omega_j} = c_{\omega_j \omega_i}$ for all $\omega_i, \omega_j \in \Omega$), and (iv) the triangular inequality (i.e., $c_{\omega_i\omega_j} \leq c_{\omega_i\omega_k} + c_{\omega_k\omega_j}$ for any $\omega_i, \omega_j, \omega_k \in \Omega$). A common transportation cost between scenarios is given by $c_{\omega_i\omega_j} = ||\xi_{\omega_i} - \xi_{\omega_j}||_\eta$ as the L_η norm (e.g., $\eta = 1, 2, \infty$), where ξ_ω denotes the vector of random parameters' realization for scenario ω. When the transportation cost is induced by a norm, this results in the well-studied 1-Wasserstein distance [e.g., 27], also known as the Kantorovich metric. We allow for more general transportation costs as long as they are distances, similar to the Optimal Transport (OT) based ambiguity sets [e.g., 10], although the OT based ambiguity sets can have more general transportation costs. Note that if the transportation cost is a distance, so is $W(\mathbf{p}, \mathbf{q})$. As a special case, when the transportation costs are given by $c_{\omega_i\omega_j} = 1$ if $\omega_i \neq \omega_j$ and 0 otherwise, the corresponding Wasserstein distance becomes the total variation distance.

Following the definition in Section 2.1, DRO-W is formulated as problem (1) where $\Delta(\mathbf{p}, \mathbf{q})$ in the ambiguity set (2) is replaced by $W(\mathbf{p}, \mathbf{q})$. Consequently, the worst-case expected problem of (1) for DRO-W can be formulated as [18]

$$\max_{\mathbf{p}\geq 0, \boldsymbol{\gamma}\geq 0} \sum_{\omega_j\in\Omega} p_{\omega_j} h_{\omega_j}(x) \tag{3a}$$

$$\text{s.t.} \sum_{\omega_i\in\Omega} \gamma_{\omega_i\omega_j} - p_{\omega_j} = 0 \quad \forall \omega_j \in \Omega, \qquad (\alpha_{\omega_j}) \tag{3b}$$

$$\sum_{\omega_j\in\Omega} \gamma_{\omega_i\omega_j} = q_{\omega_i} \qquad \forall \omega_i \in \Omega, \qquad (\beta_{\omega_i}) \tag{3c}$$

$$\sum_{\omega_i\in\Omega} \sum_{\omega_j\in\Omega} c_{\omega_i\omega_j} \gamma_{\omega_i\omega_j} \leq \rho. \qquad (\lambda) \tag{3d}$$

Observe that $\sum_{\omega_j\in\Omega} p_{\omega_j} = 1$ in (2) is automatically satisfied by (3b) and (3c) because $\sum_{\omega_i\in\Omega} q_{\omega_i} = 1$. From this point on, we refer to the radius ρ of DRO-W that appears in equation (3d) as *transportation budget* or simply as *budget*. We denote the dual variables of the constraints (3b), (3c), and (3d) under a fixed $x \in \mathcal{X}$ as $\alpha_{\omega_j} := \alpha_{\omega_j}(x)$, $\beta_{\omega_i} := \beta_{\omega_i}(x)$, and $\lambda := \lambda(x)$, respectively, and as usual we suppress the dependence on x. We denote $\boldsymbol{\alpha}, \boldsymbol{\beta}$ as their corresponding vector form.

2.3 Background on Effective Scenarios

Consider the class of convex DRO problems (1)–(2) introduced above. A set of scenarios is called *effective* if the optimal value of DRO problem (1) changes when that set of scenarios is "removed" from the problem. Therefore, to formally define effective scenarios, we first need to mathematically describe the removal of scenarios.

By [39], removing a set of scenarios $\mathcal{F} \subset \Omega$ from a DRO problem means restricting the ambiguity set through constraints $p_{\omega_i} = 0$ for all $\omega_i \in \mathcal{F}$. This means that such scenarios can no longer be in the support of worst-case distributions. DRO problem with the restricted ambiguity set is called the *assessment problem* of scenarios in $\mathcal{F}$. Thus, the ambiguity set of the assessment problem is given by $\mathcal{P}^{\mathcal{A}}(\mathcal{F}) := \mathcal{P} \cap \{p_{\omega_i} = 0 \ \forall \omega_i \in \mathcal{F}\}$. Let $\mathcal{F}^c := \Omega \setminus \mathcal{F}$. Then, the assessment problem can be written as

$$\min_{x\in\mathcal{X}}\left\{f^{\mathcal{A}}(x;\mathcal{F}):=\max_{\mathbf{p}\in\mathcal{P}^{\mathcal{A}}(\mathcal{F})}\sum_{\omega_i\in\mathcal{F}^c}p_{\omega_i}h_{\omega_i}(x)\right\}. \tag{4}$$

If the ambiguity set of the assessment problem $\mathcal{P}^{\mathcal{A}}(\mathcal{F})$ is infeasible, we set $f^{\mathcal{A}}(x;\mathcal{F})=-\infty$. This can happen, for instance, when too many scenarios are removed so that it is no longer possible to find a distribution sufficiently close to the nominal distribution.

Suppose x^* solves (1) and $x^*_{\mathcal{F}}$ solves (4). The assessment problem of scenarios in $\mathcal{F}$, given in (4), has a more restrictive ambiguity set $\mathcal{P}^{\mathcal{A}}(\mathcal{F})$ compared to the ambiguity set $\mathcal{P}$ of its corresponding DRO given in (1). Therefore, for all $x\in\mathcal{X}$, we have $f^{\mathcal{A}}(x;\mathcal{F})\leq f(x)$. Consequently, we expect the optimal value of (4) to be always less than or equal to the optimal value of (1). If the optimal value changes, however, we call such set of scenarios effective. A formal definition follows.

Definition 1 (Effective and Ineffective Scenarios [39]). *A set of scenarios $\mathcal{F}\subset\Omega$ is called* effective *if* $\min_{x\in\mathcal{X}} f^{\mathcal{A}}(x;\mathcal{F})<\min_{x\in\mathcal{X}} f(x)$ *and called* ineffective *if* $\min_{x\in\mathcal{X}} f^{\mathcal{A}}(x;\mathcal{F})=\min_{x\in\mathcal{X}} f(x)$.

By the above definition, if $\mathcal{P}^{\mathcal{A}}(\mathcal{F})$ is infeasible, the set of scenarios $\mathcal{F}$ is trivially effective because $f^{\mathcal{A}}(x;\mathcal{F})=-\infty$.

3 Effective Scenarios in General Distance-Based Ambiguity Sets

Based on Definition 1 and the assessment problem (4), it may be tempting to think that a scenario ω_j with a positive worst-case probability ($p^*_{\omega_j}>0$) is an effective scenario, while a scenario with zero worst-case probability ($p^*_{\omega_j}=0$) is an ineffective scenario. For instance, if $p^*_{\omega_j}=0$, then it appears that solving the assessment problem with the additional constraint $p_{\omega_j}=0$ would not change anything. However, this is not always the case. Section 2.1 in [39] provides a counterexample. There are two main reasons for this counterintuitive outcome. First, the optimal decision x^* of problem (1) may not always be unique. Second, for a given x^*, the worst-case distribution $\mathbf{p}^*$ may not be unique. In fact, in the case of DRO-TV, [39] shows the existence of multiple worst-case probabilities $\mathbf{p}^*$ is a necessary condition for having an effective scenario with zero worst-case probability ($p^*_{\omega_j}=0$) or an ineffective scenario with positive worst-case probability ($p^*_{\omega_j}>0$). We will revisit this below.

In this section, we expand and generalize the results in [39], which focused on DROs formed with the total variation distance to DROs formed with a general distance-based ambiguity set. We only make the following minimal assumption on the ambiguity sets.

Assumption A1. *$\mathcal{P}$ given in (2) is nonempty and compact.*

$\mathcal{P}$ being nonempty is a natural condition for DRO to be well defined; furthermore, $\mathcal{P}$ at a minimum contains the nominal distribution by construction. $\mathcal{P}$ is also bounded by definition because it is a subset of a probability simplex. Many widely used ambiguity sets are also closed. In particular, the ambiguity sets of DRO-W and DRO-TV are all bounded and closed, hence compact.

For general distance-based ambiguity sets, we study the relationship between effectiveness of a scenario and its worst-case probability. Note that determination of whether a scenario is effective or ineffective does not solely depend on the *specific value* of its worst-case probability, which may be obtained by a numerical method. However, it is closely tied to the worst-case probability associated with that scenario being *always* zero or positive among *all* possible worst-case distributions. To show these results, below we focus on a particular optimal solution $x^* \in \arg\min_{x\in\mathcal{X}} f(x)$ of the DRO problem (1). The definition of effective scenarios does not depend on a given optimal solution. However, a DM, after solving the DRO problem by their preferred method, obtains a solution $x^* \in \mathcal{X}$. Because we are ultimately interested in identifying effective solutions easily—without having to *resolve* the assessment problems for each scenario $\omega_j \in \Omega$—as a post-optimality sensitivity analysis on the scenarios, we focus on information from an obtained solution $x^* \in \mathcal{X}$ to help with this goal.

Let $\mathbb{P}^* := \mathbb{P}^*(x^*)$ denote the set of all optimal worst-case distributions $\mathbf{p}^*$ in a DRO problem (1) under an optimal solution x^*. That is, $\mathbb{P}^* := \arg\max_{\mathbf{p}\in\mathcal{P}} \mathbb{E}_{\mathbf{p}}[\mathbf{h}(x^*)]$. Like an optimal worst-case distribution $\mathbf{p}^*$, we suppress the dependence of set $\mathbb{P}^*$ on x^* for simplicity. Propositions 1 and 2 establish that the worst-case probability for a given scenario being always strictly positive or always zero is sufficient to determine whether it is effective or ineffective, respectively.

Proposition 1. *Consider the DRO problem* (1)–(2) *with Assumption A1. Suppose x^* solves* (1)–(2)*, and consider scenario $\omega_j \in \Omega$. If among all worst-case distributions $\mathbf{p}^* \in \mathbb{P}^*$ at $x^* \in \mathcal{X}$, the worst-case probability of scenario ω_j is always positive, i.e., $p^*_{\omega_j} > 0$ for all $\mathbf{p}^* \in \mathbb{P}^*$, then it is effective.*

Proof. For the sake of contradiction, suppose $p^*_{\omega_j} > 0$ for all $\mathbf{p}^* \in \mathbb{P}^*$ but instead scenario ω_j is ineffective. With $\mathcal{F} = \{\omega_j\}$, let $x^*_{\mathcal{F}}$ solve (4). Then by $\mathcal{P}^{\mathcal{A}}(\mathcal{F})$ being more restrictive than $\mathcal{P}$, suboptimality of x^* to (4), and similarly by suboptimality of $x^*_{\mathcal{F}}$ to (1), we have $f^{\mathcal{A}}(x^*_{\mathcal{F}}; \mathcal{F}) \le f^{\mathcal{A}}(x^*; \mathcal{F}) \le f(x^*) \le f(x^*_{\mathcal{F}})$. This, combined with Definition 1, implies $f^{\mathcal{A}}(x^*; \mathcal{F}) = f(x^*)$. Suppose $\mathbf{p}^0 \in \arg\max_{\mathbf{p}\in\mathcal{P}^{\mathcal{A}}(\mathcal{F})} \mathbb{E}_{\mathbf{p}}[\mathbf{h}(x^*)]$. Because $\mathcal{P}^{\mathcal{A}}(\mathcal{F}) \subset \mathcal{P}$, we have $\mathbf{p}^0 \in \mathcal{P}$. Then, $f^{\mathcal{A}}(x^*; \mathcal{F}) = \sum_{\omega_i\in\Omega} p^0_{\omega_i} h_{\omega_i}(x^*) = f(x^*) = \max_{\mathbf{p}\in\mathcal{P}} \mathbb{E}_{\mathbf{p}}[\mathbf{h}(x)]$ implies that $\mathbf{p}^0 \in \mathbb{P}^*$. However, because $\mathbf{p}^0 \in \mathcal{P}^{\mathcal{A}}(\mathcal{F})$, it must have $p^0_{\omega_j} = 0$, which forms a contradiction. □

Note that Proposition 1 does not need the worst-case probability at scenario ω_j, $p^*_{\omega_j}$, to be unique at x^*.

Proposition 2. *Consider the DRO problem* (1)–(2) *with Assumption A1. Suppose x^* solves* (1)–(2)*, and consider scenario $\omega_j \in \Omega$. If the worst-case probability of scenario ω_j is uniquely zero, i.e., $p^*_{\omega_j} = 0$ for all $\mathbf{p}^* \in \mathbb{P}^*$ at $x^* \in \mathcal{X}$, then it is ineffective. The union of any such scenarios is also ineffective.*

Proof. Suppose the set $\left\{\omega_j \in \Omega : p^*_{\omega_j} = 0 \ \forall \mathbf{p}^* \in \mathbb{P}^*\right\}$ is nonempty, and let $\mathcal{F}_s$ be an arbitrary nonempty subset of it. We will prove the result by first showing (a) $f(x^*) = f^{\mathcal{A}}(x^*; \mathcal{F}_s)$ and

(b) $\partial f(x^*) = \partial f^{\mathcal{A}}(x^*; \mathcal{F}_s)$; then we will apply first-order optimality conditions to show (c) x^* is an optimal solution of the assessment problem at $\mathcal{F}_s$, i.e., $x^* \in \arg\min_{x\in\mathcal{X}} \; f^{\mathcal{A}}(x; \mathcal{F}_s)$. This, combined with (a), shows that $\mathcal{F}_s$ is ineffective. Condition (a) is straightforward since $\sum_{\omega_j \in \mathcal{F}_s} p^*_{\omega_j} = 0$ under all optimal solutions $\mathbf{p}^* \in \mathbb{P}^*$. This also implies $\mathbb{P}^* = \arg\max_{\mathbf{p}\in\mathcal{P}^{\mathcal{A}}(\mathcal{F}_s)} \sum_{\omega_i \in \mathcal{F}_s^c} p_{\omega_i} h_{\omega_i}(x^*)$. Now we show (b). Since $\mathcal{P}$ is nonempty compact and $\sum_{\omega_i\in\Omega} p_{\omega_i} h_{\omega_i}(x^*)$ is affine in $\mathbf{p}$, by [22, Theorem 4.4.2], we have

$$\partial f(x^*) = \text{Conv}\left(\bigcup_{\mathbf{p}^*\in\mathbb{P}^*} \partial \sum_{\omega_i\in\Omega} p^*_{\omega_i} h_{\omega_i}(x^*) \right), \tag{5}$$

where $\text{Conv}(\mathcal{S})$ denotes the convex hull of set $\mathcal{S}$. Now, for any $\mathbf{p}^* \in \mathbb{P}^*$, we have $\partial \sum_{\omega_i\in\Omega} p^*_{\omega_i} h_{\omega_i}(x^*) = \sum_{\omega_i\in\Omega} \partial p^*_{\omega_i} h_{\omega_i}(x^*)$ by [42, Theorem 23.8] and also $\sum_{\omega_i\in\Omega} \partial p^*_{\omega_i} h_{\omega_i}(x^*) = \sum_{\omega_i\in\Omega} p^*_{\omega_i} \partial h_{\omega_i}(x^*)$ with $p^*_{\omega_i} \geq 0$ because $\partial h_{\omega_i}(x^*) \neq \varnothing$ and bounded for all $\omega_i \in \Omega$ by our assumptions. Furthermore, $\sum_{\omega_i\in\Omega} p^*_{\omega_i} \partial h_{\omega_i}(x^*) = \sum_{\omega_i\in\Omega\setminus\mathcal{F}_s} p^*_{\omega_i} \partial h_{\omega_i}(x^*) = \partial \sum_{\omega_i\in\Omega\setminus\mathcal{F}_s} p^*_{\omega_i} h_{\omega_i}(x^*)$ by similar arguments and definition of set $\mathcal{F}_s$. Combining these with (5), we obtain

$$\partial f(x^*) = \text{Conv}\left(\bigcup_{\mathbf{p}^*\in\mathbb{P}^*} \partial \sum_{\omega_i\in\Omega\setminus\mathcal{F}_s} p^*_{\omega_i} h_{\omega_i}(x^*) \right) = \partial f^{\mathcal{A}}(x^*; \mathcal{F}_s).$$

The last equality above holds by another application of [22, Theorem 4.4.2] since $\mathcal{P}^{\mathcal{A}}(\mathcal{F}_s)$ is nonempty compact, $\sum_{\omega_i\in\Omega} p_{\omega_i\in\Omega\setminus\mathcal{F}_s} h_{\omega_i}(x^*)$ is affine in $\mathbf{p}$, and $\mathbb{P}^*$ forms the set of optimal solutions to the worst-case expected problem of the assessment problem at x^*. This completes (b).

Because x^* is an optimal solution of $\min_{x\in\mathcal{X}} f(x)$, there exists $s \in \partial f(x^*) = \partial f_{\mathcal{A}}(x^*; \mathcal{F}_s)$ such that $s(x - x^*) \geq 0$ for all $x \in \mathcal{X}$. Then, by convexity of $f_{\mathcal{A}}(x; \mathcal{F}_s)$ in x, $f_{\mathcal{A}}(x; \mathcal{F}_s) \geq f_{\mathcal{A}}(x^*; \mathcal{F}_s) + s(x - x^*) \geq f_{\mathcal{A}}(x^*; \mathcal{F}_s)$ for all $x \in \mathcal{X}$. This shows that x^* is also an optimal solution to the assessment problem at $\mathcal{F}_s$, $\min_{x\in\mathcal{X}} f_{\mathcal{A}}(x; \mathcal{F}_s)$. Then, because $f(x^*) = f_{\mathcal{A}}(x^*; \mathcal{F}_s)$ by (a), the set of scenarios $\mathcal{F}_s$ is ineffective by Definition 1. Since $\mathcal{F}_s$ is an arbitrary subset of $\left\{\omega_j \in \Omega : p^*_{\omega_j} = 0, \forall \mathbf{p}^* \in \mathbb{P}^*\right\}$, any union of such scenarios is ineffective. □

In general, it is possible to have scenarios that are individually ineffective, but when considered together, they become effective. Proposition 2 shows that all scenarios ω_j with uniquely zero worst-case probabilities (i.e., $p^*_{\omega_j} = 0, \forall \mathbf{p}^* \in \mathbb{P}^*$) are also collectively ineffective. If such scenarios can be identified, they can be safely removed from the problem, resulting in a reduced-size problem that is computationally more tractable; see, e.g., [2] for an application of this idea for scenario reduction.

Let us now consider the opposite cases of Propositions 1 and 2, i.e., when we observe an effective scenario with zero worst-case probability (i.e., $p^*_{\omega_j} = 0$) or an ineffective scenario with positive worst-case probability (i.e., $p^*_{\omega_j} > 0$). Theorem 5 of [39] shows that in DRO-TV, if scenario ω_j is a zero-probability effective scenario or a positive-probability ineffective scenario, then (i) the worst-case expected problem in (1)–(2) has multiple worst-case distributions and (ii) there exists at least one scenario ω_i with the same cost as scenario ω_j: $h_{\omega_i}(x^*) = h_{\omega_j}(x^*)$. This result can be partially extended to DROs under general ambiguity sets in the sense that, when such scenarios are observed, multiple worst-case distributions exist. However, in general, there may not be two scenarios having the same cost. We will provide a counterexample with DRO-W in Section 4. We present the generalized result in the following.

Proposition 3. *Consider the DRO problem* (1)–(2) *with Assumption A1. Suppose x^* solves* (1)–(2)*. If there exists a zero-probability effective scenario ω_j ($p^*_{\omega_j} = 0$) or a positive-probability ineffective scenario ω_j ($p^*_{\omega_j} > 0$), then the worst-case expected problem at x^* has multiple optimal solutions.*

Proof. First, consider an effective scenario ω_j and $\mathbf{p}^0 \in \mathbb{P}^*$ such that $p^0_{\omega_j} = 0$. Then, there must exist another optimal worst-case distribution, $\mathbf{p}^1 \in \mathbb{P}^*$ such that $p^1_{\omega_j} > 0$. Suppose not. Then, $p^*_{\omega_j} = 0$ is unique among all $\mathbf{p}^* \in \mathbb{P}^*$. This means ω_j is ineffective by Proposition 2, which forms a contradiction. Now consider an ineffective scenario ω_j with $\mathbf{p}^0 \in \mathbb{P}^*$ such that $p^0_{\omega_j} > 0$. Then there must exist at least one other optimal worst-case distribution, $\mathbf{p}^1 \in \mathbb{P}^*$ such that $p^1_{\omega_j} = 0$. Suppose not. Similarly, this means for all $\mathbf{p}^* \in \mathbb{P}^*$, $p^*_{\omega_j} > 0$. Proposition 1 then implies scenario ω_j must be effective, which is a contradiction. Therefore, under the conditions of the proposition, the worst-case expected problem has multiple optima. □

Let us now discuss the implications of these results. Propositions 1, 2, and 3 significantly generalize the analysis in [39], from DROs formed via total variation distance to general distance-based ambiguity sets. They can be very useful in some situations. For instance, some ϕ-divergences are known to be incapable of *popping* scenarios [5]. This means that if a scenario ω_j has zero nominal probability, the resulting ambiguity set cannot admit any distributions with positive probability for scenario ω_j. Thus, any optimal solution must have $p^*_{\omega_j} = 0$, and by Proposition 2, such scenarios can immediately be classified as ineffective without even solving the DRO problem. A similar result exists for positive-nominal-probability scenarios and ϕ-divergences that cannot admit zero probabilities for such scenarios in their ambiguity sets (called 'cannot *suppress* scenarios' [5], also related to strictly monotone risk measures [45]); thus immediately providing effective scenarios. As another example, if the optimal worst-case distribution at a given optimal solution x^* can be identified to be unique, then the worst-case probabilities being zero or positive automatically determine the effectiveness of scenarios. In other cases, these results may be somewhat difficult to use because determining unique or multiple optimal solutions, and in the case of multiple optima, determining all multiple solutions, can be difficult. Nevertheless, they provide a framework to identify the effectiveness of scenarios. They may also help uncover further structures present in these problems and reveal further insights into the behavior of DROs. In particular, we are able to use these results to identify the effectiveness of scenarios in DRO-W without needing to solve any additional optimization (e.g., assessment) problems.

Finally, we remark that in our experiments presented in Section 6, we never encountered a case where there is an easily identifiable unique solution to the worst-case expected problem of DRO-W given in (3). In fact, the obtained solutions were always degenerate (with a large number of nonbasic variables with zero reduced costs), which makes identifying unique/multiple optima more difficult. For instance, determining uniqueness of an optimal solution in such cases can be viewed equivalent to solving an optimization problem [8, p.130]. In the case of multiple optima, by Propositions 1 and 2, one also needs to determine if the worst-case probabilities are always positive or not. One could potentially resort to minimizing/maximizing p_ω for all $\omega \in \Omega$ with an additional constraint setting the objective function equal to the optimal value, but this is even more demanding than solving the assessment problems. In what follows, we exploit the structure of the worst-case expected problem of DRO-W given in (3) to determine the effectiveness of scenarios in a relatively efficient manner.

4 Effective Scenarios in DRO-W

For the rest of the paper, we narrow our focus to DRO-W. We first present notation and a categorization that will be used to identify the effectiveness of scenarios for this class of DRO. Then we present conditions to identify effective scenarios based on Propositions 1 and 2. We assume all scenarios in DRO-W have positive nominal probabilities, i.e., $q_{\omega_j} > 0$ for all $\omega_j \in \Omega$.

4.1 Categories with Respect to Probability Mass Inflows

Recall the transportation variables $\gamma_{\omega_i \omega_j}$ in worst-case expected problem of DRO-W given in (3). If $\gamma_{\omega_i \omega_j} > 0$, then there is an inflow of probability mass into scenario ω_j from scenario ω_i. If $\gamma_{\omega_i \omega_i} > 0$, this is an inflow into scenario ω_i from itself. To provide a categorization of scenarios with respect to their inflows, we examine the dual of (3). Because we study easy-to-check conditions at an optimal solution, from this point on, we focus on the worst-case expected problem of DRO-W (3) at an optimal solution $x^* \in \mathcal{X}$ to DRO-W. The dual of (3) can then be written as

$$\min_{\lambda \geq 0, \boldsymbol{\alpha}, \boldsymbol{\beta}} \ \rho\lambda + \sum_{\omega_i \in \Omega} q_{\omega_i} \beta_{\omega_i} \tag{6a}$$

$$\text{s.t.} \ -\alpha_{\omega_j} \geq h_{\omega_j}(x^*) \qquad \forall \omega_i \in \Omega, \qquad (p_{\omega_j}) \tag{6b}$$

$$\alpha_{\omega_j} + \beta_{\omega_i} + \lambda c_{\omega_i \omega_j} \geq 0 \qquad \forall \omega_i, \omega_j \in \Omega, \qquad (\gamma_{\omega_i \omega_j}) \tag{6c}$$

where the primal decisions $\boldsymbol{p}, \boldsymbol{\gamma}$ become the dual variables of the constraints (6b) and (6c), respectively. Both the primal problem (3) and its dual (6) always have feasible solutions. Therefore, both have finite optimal solutions and their optimal values are identical. As before, we use superscript * to denote an optimal dual solution to (6) such as λ^*, $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ at $x^* \in \mathcal{X}$. Our analysis specifically focuses on the optimal dual variable $\lambda^* := \lambda^*(x^*)$ corresponding to the transportation budget constraint (3d). We denote by $\Lambda^* := \Lambda^*(x^*)$ the set of all optimal dual solutions λ^* to (6) at x^*. As before, we suppress the dependence of Λ^*, λ^*, $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ on x^* for simplicity. Combining (6b) and (6c), we have $\beta^*_{\omega_i} \geq h_{\omega_j}(x^*) - \lambda^* c_{\omega_i \omega_j}$, for all $\omega_j \in \Omega$ for a given ω_i. Since problem (6) minimizes $q_{\omega_i}\beta_{\omega_i}$ with $q_{\omega_i} > 0$, we have at an optimal solution $\beta^*_{\omega_i} = \max_\omega \{h_\omega(x^*) - \lambda^* c_{\omega_i \omega}\}$. This motivates us to define the following sets.

The set $\mathcal{A}(\omega_i, \lambda^*) := \mathcal{A}(\omega_i, \lambda^*, x^*) := \arg\max_\omega\{h_\omega(x^*) - \lambda^* c_{\omega_i\omega}\}$ denotes the set of scenarios ω that maximizes the value of $h_\omega(x^*) - \lambda^* c_{\omega_i\omega}$ for a given scenario ω_i at a given dual optimal solution λ^*. This represents the set of scenarios that can be a destination for the nominal probability mass transferred out of scenario ω_i. Furthermore, let $\mathcal{M}(\omega_j, \lambda^*) := \mathcal{M}(\omega_j, \lambda^*, x^*) := \{\omega_i \in \Omega : \omega_j \in \mathcal{A}(\omega_i)\}$ represent the set of scenarios that scenario ω_j is a potential destination of at a given dual optimal solution λ^*. Note that $\omega_i \in \mathcal{M}(\omega_j, \lambda^*)$ if and only if $\omega_j \in \mathcal{A}(\omega_i, \lambda^*)$. While both sets $\mathcal{A}(\omega_i, \lambda^*)$ and $\mathcal{M}(\omega_j, \lambda^*)$ depend on x^*, for simplicity of presentation, we suppress this dependence from the notation as well.

At a given optimal solution $x^* \in \mathcal{X}$ and a corresponding optimal dual λ^*, we define a partition of the scenario set Ω into the following subsets (using the shortcuts N for Never a destination, U for Unique destination, and M for Multiple destinations):

- $\Omega_N(x^*, \lambda^*) := \{\omega_j \in \Omega : \mathcal{M}(\omega_j, \lambda^*) = \varnothing\}$, i.e., set of scenarios that are never a destination for probability mass inflows;
- $\Omega_U(x^*, \lambda^*) := \{\omega_j \in \Omega : \mathcal{A}(\omega_i, \lambda^*) = \{\omega_j\}$ for at least one $\omega_i \in \Omega\}$, i.e., set of scenarios that are a unique destination for probability mass inflow for at least one scenario;
- $\Omega_M(x^*, \lambda^*) := \{\omega_j \in \Omega : \mathcal{M}(\omega_j, \lambda^*) \neq \varnothing$ and $\{\omega_j\} \subsetneq \mathcal{A}(\omega_i, \lambda^*)$ for all $\omega_i \in \mathcal{M}(\omega_j, \lambda^*)\}$, i.e., set of scenarios that are always one of the multiple destinations for probability mass inflow from other scenarios.

Figure 1 shows an example of this partition with the set of scenarios $\Omega = \{\omega_1, \omega_2, \omega_3\}$, where an edge pointing from ω_i to ω_j means $\omega_i \in \mathcal{M}(\omega_j, \lambda^*)$ or equivalently $\omega_j \in \mathcal{A}(\omega_i, \lambda^*)$ at the given x^* and its corresponding optimal dual λ^*, indicating potential inflows into scenario ω_j. In this example, $\Omega_N(x^*, \lambda^*) = \{\omega_2\}$, $\Omega_U(x^*, \lambda^*) = \{\omega_1\}$ and $\Omega_M(x^*, \lambda^*) = \{\omega_3\}$.

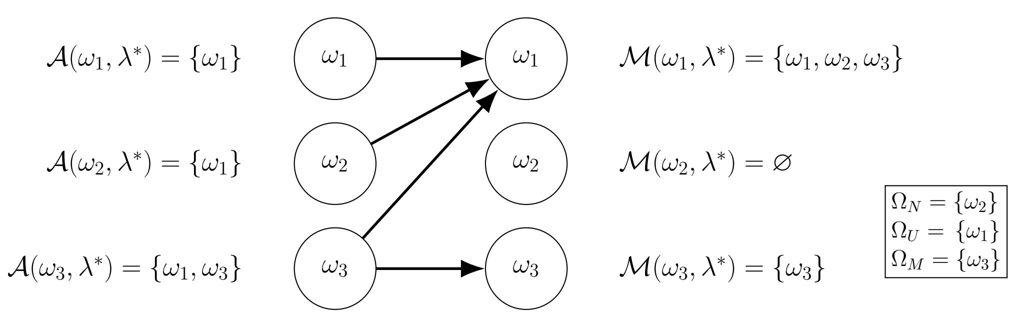


Figure 1: An example of the partition of the scenario set $\Omega = \{\omega_1, \omega_2, \omega_3\}$ given a fixed λ^*. An edge pointing from ω_i to ω_j means $\omega_i \in \mathcal{M}(\omega_j, \lambda^*)$ or equivalently $\omega_j \in \mathcal{A}(\omega_i, \lambda^*)$ at a given x^*.

4.2 Easy-to-Check Conditions for DRO-W

Based on the above partition of Ω, we first provide in this section a summary of easy-to-check conditions for identifying effective and ineffective scenarios in DRO-W. We will discuss details and present the proofs in later sections.

First observe that when $\rho = 0$, the ambiguity set contains only the nominal distribution and the worst-case distribution must be the same as the nominal distribution because we assume positivity in the transportation costs $c_{\omega_i \omega_j}$. Then, all scenarios are effective by Proposition 1. Therefore, for the rest of this section, we assume we are not in the trivial case of $\rho = 0$; i.e., we assume $\rho > 0$. Furthermore, to simplify the statements of the below theorems, for any $\omega_i \in \Omega_N(x^*, \lambda^*) \cup \Omega_M(x^*, \lambda^*)$, we use the notation $\underline{c}_{\omega_i}^{\lambda^*} := \min_{\omega_j \in \mathcal{A}(\omega_i, \lambda^*) \setminus \{\omega_i\}} c_{\omega_i \omega_j}$. That is, $\underline{c}_{\omega_i}^{\lambda^*}$ denotes the smallest positive unit transportation cost from scenario ω_i to any $\omega_j \in \mathcal{A}(\omega_i, \lambda^*)$, excluding the case $\omega_j = \omega_i$ if $\omega_i \in \mathcal{A}(\omega_i, \lambda^*)$ (because $c_{\omega_i \omega_i} = 0$). We are now ready to present the easy-to-check conditions.

Theorem 1 (Easy-to-check conditions for effective scenarios)**.** *Consider the worst-case expected problem of DRO-W (3) at an optimal solution $x^* \in \mathcal{X}$ with a given optimal dual variable $\lambda^* := \lambda^*(x^*)$ corresponding to constraint (3d). Scenario ω_j is effective if any of the following conditions hold:*

(i) $\rho < c_{\omega_j \omega_i} \cdot q_{\omega_j}$ *for all* $\omega_i \in \Omega \setminus \{\omega_j\}$*;*

(ii) $\omega_j \in \Omega_U(x^*, \lambda^*)$*;*

(iii) $\omega_j \in \Omega_M(x^*, \lambda^*)$ *and* $\rho - \sum_{\omega_k \in \Omega_N(x^*, \lambda^*)} q_{\omega_k} \cdot \underline{c}_{\omega_k}^{\lambda^*} < q_{\omega_j} \cdot \underline{c}_{\omega_j}^{\lambda^*}$.

Theorem 2 (Easy-to-check conditions for ineffective scenarios)**.** *Consider the worst-case expected problem of DRO-W (3) as described in Theorem 1. Scenario ω_j is ineffective if any of the following conditions hold:*

(i) $\omega_j \in \Omega_N(x^*, \lambda^*)$*;*

(ii) $\omega_j \in \Omega_M(x^*, \lambda^*)$ *and* $p^*(\Omega_M(x^*, \lambda^*)) = 0$*; also, for any scenario* $\omega_i \in \Omega_M(x^*, \lambda^*) \cup \Omega_N(x^*, \lambda^*)$*, one of the following conditions hold: (a)* $\mathcal{A}(\omega_i, \lambda^*) \setminus \{\omega_i\}$ *is a singleton, or (b)* $h_\omega(x^*)$ *is the same for all* $\omega \in \mathcal{A}(\omega_i, \lambda^*) \setminus \{\omega_i\}$*, or (c)* $\gamma^*_{\omega_i \omega} = 0$ *for any* $\omega \in \mathcal{A}(\omega_i, \lambda^*)$ *satisfying* $\omega \notin \arg\max_{\omega' \in \mathcal{A}(\omega_i, \lambda^*)} h_{\omega'(x^*)}$.

Furthermore, the union of any of the ineffective scenarios identified by (i)–(ii) above is ineffective.

Theorems 1 and 2 mainly consider the effectiveness of a single scenario. In addition, Theorem 2 establishes that any union of ineffective scenarios identified by Theorem 2 is also effective, which is not generally true of ineffective scenarios.

The above conditions, while relatively easy, are not computationally free to check. Conditions stated in Theorem 1(ii)–(iii) and Theorem 2 require one sorting for each scenario to determine the sets $\Omega_N(x^*, \lambda^*), \Omega_U(x^*, \lambda^*)$ and $\Omega_M(x^*, \lambda^*)$, thus resulting in a $O(|\Omega|^2 log(|\Omega|))$ complexity. Condition 1(i) can be checked a priori, using only the problem parameters. That said, these conditions are still significantly cheaper than solving the individual assessment problems for all scenarios $\omega_j \in \Omega$; they avoid solving any additional optimization problems.

As discussed earlier, the partition $\Omega_N(x^*, \lambda^*)$, $\Omega_U(x^*, \lambda^*)$ and $\Omega_M(x^*, \lambda^*)$—and hence the majority of the conditions stated in Theorems 1 and 2—depends on a given value of λ^*. However, the effectiveness of scenarios should in principle not depend on a particular $\lambda^* \in \Lambda^*$ (see Definition 1). In Sections 4.3–4.4 below, we will show that if any of the above conditions is satisfied at a given λ^*, we are able to identify the effectiveness of a scenario at *all* $\lambda^* \in \Lambda^*$. As such, the above conditions avoid the computational cost of finding all dual optimal solutions.

The remainder of this section is organized as follows. Section 4.3 investigates the worst-case distribution $\mathbf{p}^*$ at a fixed dual optimal solution λ^*. Next, Section 4.4 studies the changes to the partition $\Omega_N(x^*, \lambda^*)$, $\Omega_U(x^*, \lambda^*)$, and $\Omega_M(x^*, \lambda^*)$ under varying values of optimal dual solutions in the set Λ^*. Section 4.5 then proves the main results presented above. Finally, Section 4.6 discusses the implications of these results and establishes further details on the multiple optimal λ^* case. Particularly, it turns out, when Λ^* is *not* a singleton set, the conditions stated in Theorems 1 and 2 identify the effectiveness of *all* scenarios in Ω. This is not always the case when there is a unique optimum λ^*; however, our computational results show that these conditions can also identify the effectiveness of all or majority of the scenarios in this case.

4.3 Results for a Fixed λ^*

Let us begin by discussing the worst-case probability distribution for a fixed dual optimal value λ^*. Note that for a fixed λ^*, we may have multiple (or unique) primal optimal solutions $(\mathbf{p}^*, \boldsymbol{\gamma}^*)$ to problem (3). We first state three simple but important lemmas that are routinely used.

Lemma 1. *Consider the worst-case expected problem of DRO-W (3) as described in Theorem 1. If $\omega_j \notin \mathcal{A}(\omega_i, \lambda^*)$ for a given scenario ω_i, then we must have $\gamma^*_{\omega_i\omega_j} = 0$ at that λ^*.*

Proof. We show this by contradiction. Suppose $\gamma^*_{\omega_i\omega_j} > 0$. Then, by complementary slackness, $\beta^*_{\omega_i} = -\alpha^*_{\omega_j} - \lambda^* c_{\omega_i\omega_j}$. Also, since $p^*_{\omega_j} \geq \gamma^*_{\omega_i\omega_j} > 0$ by constraint (3b), using complementary slackness again, we obtain $-\alpha^*_{\omega_j} = h_{\omega_j}(x^*)$. Therefore, $\beta^*_{\omega_i} = h_{\omega_j}(x^*) - \lambda^* c_{\omega_i\omega_j}$. Because, $\beta^*_{\omega_i}$ must satisfy (6c) for all scenarios, this means scenario ω_j must belong to the set $\mathcal{A}(\omega_i, \lambda^*)$, which results in a contradiction. □

Lemma 2. *Consider the worst-case expected problem of DRO-W (3) as described in Theorem 1. If $\omega_j \in \mathcal{A}(\omega_i, \lambda^*)$ for any scenario $\omega_i \neq \omega_j$, then we must have $\omega_j \in \mathcal{A}(\omega_j, \lambda^*)$ at that λ^*.*

Proof. Suppose $\omega_j \notin \mathcal{A}(\omega_j, \lambda^*)$; instead, suppose $\omega_k \in \mathcal{A}(\omega_j, \lambda^*)$ for some scenario $\omega_k \neq \omega_j$. Then, we have $h_{\omega_k}(x^*) - \lambda^* c_{\omega_j\omega_k} > h_{\omega_j}(x^*)$. Also, since $\omega_j \in \mathcal{A}(\omega_i, \lambda^*)$ for a scenario $\omega_i \neq \omega_j$, we have $h_{\omega_j}(x^*) - \lambda^* c_{\omega_i\omega_j} \geq h_{\omega_k}(x^*) - \lambda^* c_{\omega_i\omega_k}$. When $\lambda^* > 0$, combining the two inequalities gives $c_{\omega_i\omega_j} + c_{\omega_j\omega_k} < c_{\omega_i\omega_k}$, which is a contradiction to the triangular inequality. When $\lambda^* = 0$, by combining the inequalities we get $h_{\omega_k}(x^*) > h_{\omega_k}(x^*)$, which is impossible. □

From Lemma 2, we know that if $\omega_j \in \Omega_M(x^*, \lambda^*)$, we must have $\omega_j \in \mathcal{A}(\omega_j, \lambda^*)$. However, it is possible to say more when $\omega_j \in \Omega_U(x^*, \lambda^*)$, which is presented below.

Lemma 3. *Consider the worst-case expected problem of DRO-W (3) as described in Theorem 1. If scenario $\omega_j \in \Omega_U(x^*, \lambda^*)$, we must have $\mathcal{A}(\omega_j, \lambda^*) = \{\omega_j\}$ at that λ^*.*

Proof. If $\mathcal{A}(\omega_j, \lambda^*) = \{\omega_j\}$, then $\omega_j \in \Omega_U(x^*, \lambda^*)$ and the proposition is automatically satisfied. Otherwise, by Lemma 2, we have $\omega_j \in \mathcal{A}(\omega_j, \lambda^*)$. We now show $\mathcal{A}(\omega_j, \lambda^*)$ is a singleton set. Since $\omega_j \in \Omega_U(x^*, \lambda^*)$, let $\{\omega_j\} = \mathcal{A}(\omega_i, \lambda^*)$ for some $\omega_i \neq \omega_j$. Then, we have $h_{\omega_j}(x^*) - \lambda^* c_{\omega_i \omega_j} > h_\omega(x^*) - \lambda^* c_{\omega_i \omega}$ for any $\omega \neq \omega_j$. Suppose set $\mathcal{A}(\omega_j, \lambda^*)$ also contains ω_k. This means $h_{\omega_k}(x^*) - \lambda^* c_{\omega_j \omega_k} = h_{\omega_j}(x^*)$. When $\lambda^* > 0$, combining the two, we get $c_{\omega_i \omega_j} + c_{\omega_j \omega_k} < c_{\omega_i \omega_k}$, which is a contradiction to the triangular inequality. When $\lambda^* = 0$, on the other hand, we get $h_{\omega_j}(x^*) > h_{\omega_j}(x^*)$, which is impossible. □

The implications of Lemmas 1–3 with respect to the partition $\Omega_N(x^*, \lambda^*)$, $\Omega_U(x^*, \lambda^*)$, $\Omega_M(x^*, \lambda^*)$ are as follows. (i) If $\omega_j \in \Omega_N(x^*, \lambda^*)$, then we must have $\gamma^*_{\omega_i \omega_j} = 0$ for all $\omega_i \in \Omega$ at every primal optimal solution at that λ^*. Also, $\omega_j \notin \mathcal{A}(\omega_j, \lambda^*)$. (ii) If $\omega_j \in \Omega_M(x^*, \lambda^*)$, then we have $\omega_j \in \mathcal{A}(\omega_j, \lambda^*)$ and also we must have at least one other scenario $\omega_k \neq \omega_j$ belonging to the set $\mathcal{A}(\omega_j, \lambda^*)$; i.e., $\omega_k \in \mathcal{A}(\omega_j, \lambda^*)$. In this case, when $\lambda^* > 0$, $h_{\omega_k}(x^*) > h_{\omega_j}(x^*)$ because $h_{\omega_j}(x^*) = h_{\omega_k}(x^*) - \lambda^* c_{\omega_j \omega_k}$. On the other hand, when $\lambda^* = 0$, $h_{\omega_k}(x^*) = h_{\omega_j}(x^*) = \sup[\mathbf{h}(x^*)] := \sup_{\omega \in \Omega} h_\omega(x^*)$. (iii) Finally, when $\omega_j \in \Omega_U(x^*, \lambda^*)$, we must have $\mathcal{A}(\omega_j, \lambda^*) = \{\omega_j\}$ and $\gamma^*_{\omega_j \omega_j} = q_{\omega_j}$ at every optimal solution at that λ^*. We will use these results as the foundation of our proofs in the rest of this section.

As discussed before, for a fixed λ^*, we may have multiple or unique ($\mathbf{p}^*$, $\boldsymbol{\gamma}^*$). Propositions 4–7 below identify portions of the worst-case probabilities $\mathbf{p}^*$ that must be always zero or positive among *all* possible optimal $(\mathbf{p}^*, \boldsymbol{\gamma}^*)$ at that λ^*. For scenarios in $\Omega_N(x^*, \lambda^*)$ and $\Omega_U(x^*, \lambda^*)$, the results readily follow from above, which we present first.

Proposition 4. *Consider the worst-case expected problem of DRO-W (3) as described in Theorem 1. Any scenario $\omega_j \in \Omega_N(x^*, \lambda^*)$ must have $p^*_{\omega_j} = 0$ at that λ^*.*

Proof. By Lemma 1, we must have $\gamma^*_{\omega_i \omega_j} = 0$ for all $\omega_i \in \Omega$ at all optimal solutions to (3). Therefore, we must also have $p^*_{\omega_j} = \sum_{\omega_i \in \Omega} \gamma^*_{\omega_i \omega_j} = 0$. □

Proposition 5. *Consider the worst-case expected problem of DRO-W (3) as described in Theorem 1. Any scenario $\omega_j \in \Omega_U(x^*, \lambda^*)$ must have $p^*_{\omega_j} > 0$ at that λ^*.*

Proof. By Lemma 3, $\mathcal{A}(\omega_j) = \{\omega_j\}$. Therefore, by Lemma 1, we must have $\gamma^*_{\omega_j \omega_i} = 0$ for all $\omega_i \neq \omega_j$. Then, by constraints (3b) and (3c), we have $p^*_{\omega_j} \geq \gamma^*_{\omega_j \omega_j} > 0$. □

We now discuss the case when $\omega_j \in \Omega_M(x^*, \lambda)$, which is more complicated. We begin with a lemma that will be useful to study this case, followed by a result (Proposition 6) which provides a set of conditions such that all scenarios in $\Omega_M(x^*, \lambda^*)$ always have zero worst-case probabilities.

Lemma 4. *Consider the worst-case expected problem of DRO-W (3) as described in Theorem 1. Suppose $\lambda^* > 0$. Then, for any scenario $\omega_i \in \Omega$, the highest-cost scenarios in $\mathcal{A}(\omega_i, \lambda^*)$ must belong to $\Omega_U(x^*, \lambda^*)$, where the cost of scenario ω is given by $h_\omega(x^*)$.*

Proof. When $\omega_i \in \Omega_U(x^*, \lambda^*)$, because $\mathcal{A}(\omega_i, \lambda^*) = \{\omega_i\}$ is a singleton, the result is automatically satisfied. Now let $\omega_i \in \Omega_N(x^*, \lambda^*) \cup \Omega_M(x^*, \lambda^*)$. For the sake of contradiction, let $\omega_j \in \mathcal{A}(\omega_i, \lambda^*)$ with $\omega_j \in \arg\max_{\omega' \in \mathcal{A}(\omega_i,\lambda^*)} h_{\omega'}(x^*)$ but $\omega_j \in \Omega_M(x^*, \lambda^*)$. Note that $\omega_j \notin \Omega_N(x^*, \lambda^*)$ because $\omega_j \in \mathcal{A}(\omega_i, \lambda^*)$. Then, since $\omega_j \in \Omega_M(x^*, \lambda^*)$, there must be another scenario $\omega_s \in \mathcal{A}(\omega_j, \lambda^*)$ satisfying $h_{\omega_s}(x^*) - \lambda^* c_{\omega_j \omega_s} = h_{\omega_j}(x^*)$ and because $\lambda^* > 0$, we have $h_{\omega_s}(x^*) > h_{\omega_j}(x^*)$. Also, since $\omega_j \in \mathcal{A}(\omega_i, \lambda^*)$, we have $h_{\omega_j}(x^*) - \lambda^* c_{\omega_i \omega_j} \geq h_{\omega_s}(x^*) - \lambda^* c_{\omega_i \omega_s}$. Then, by these and the triangular inequality, we get $c_{\omega_i \omega_j} + c_{\omega_j \omega_s} = c_{\omega_i \omega_s}$. This indicates that $h_{\omega_j}(x^*) - \lambda^* c_{\omega_i \omega_j} = h_{\omega_s}(x^*) - \lambda^* c_{\omega_i \omega_s}$ and so $\omega_s \in \mathcal{A}(\omega_i, \lambda^*)$. This causes a contradiction to ω_j being the highest-cost scenario in $\mathcal{A}(\omega_i, \lambda^*)$ because $h_{\omega_s}(x^*) > h_{\omega_j}(x^*)$. $\square$

Proposition 6. *Consider the worst-case expected problem of DRO-W (3) as described in Theorem 1. All scenarios $\omega_j \in \Omega_M(x^*, \lambda^*)$ must have $p^*_{\omega_j} = 0$ at that λ^* if both of the below conditions hold simultaneously: (i) $p^*(\Omega_M(x^*, \lambda^*)) = 0$, and (ii) for any scenario $\omega_i \in \Omega_M(x^*, \lambda^*) \cup \Omega_N(x^*, \lambda^*)$, one of the following conditions hold: (a) $\mathcal{A}(\omega_i, \lambda^*) \setminus \{\omega_i\}$ is a singleton, or (b) $h_\omega(x^*)$ is the same for all $\omega \in \mathcal{A}(\omega_i, \lambda^*) \setminus \{\omega_i\}$, or (c) $\gamma^*_{\omega_i \omega} = 0$ for any $\omega \in \mathcal{A}(\omega_i, \lambda^*)$ such that $\omega \notin \arg\max_{\omega' \in \mathcal{A}(\omega_i,\lambda^*)} h_{\omega'}(x^*)$. In this case, $\lambda^* > 0$.*

Proof. If $\lambda^* = 0$, $\mathcal{A}(\omega_k, \lambda^*)$ only contains the highest-cost scenarios for any $\omega_k \in \Omega$. Then either $\Omega_M(x^*, \lambda^*)$ is empty and we have nothing to show or $\Omega_M(x^*, \lambda^*)$ contains all the highest-cost scenarios. In the latter case, $\Omega_U(x^*, \lambda^*)$ is empty; therefore by Lemma 1 and constraint (3c), for any $\omega_k \in \Omega$, we must have $\gamma_{\omega_k \omega_j} > 0$ for some $\omega_j \in \Omega_M(x^*, \lambda^*)$. Thus, condition (i) cannot be satisfied. So, the conditions of the proposition can only be satisfied when $\lambda^* > 0$.

Now suppose $\lambda^* > 0$. By (i), there exists an optimal solution to (3), denoted $(\mathbf{p}^{*,(0)}, \boldsymbol{\gamma}^{*,(0)})$, that satisfies $\gamma^{*,(0)}_{\omega_i \omega_j} = 0$ for all $\omega_j \in \Omega_M(x^*, \lambda^*)$ and all $\omega_i \in \Omega$. Furthermore by (ii), for all $\omega_i \in \Omega$, $\gamma^{*,(0)}_{\omega_i \omega} > 0$ only happens when ω has the highest cost among all scenarios in $\mathcal{A}(\omega_i, \lambda^*)$.

Let us fix an arbitrary $\omega_j \in \Omega_M(x^*, \lambda^*)$. In the subsequent discussion, we will show that at *any* optimal solution to (3), $\gamma^*_{\omega_l \omega_j} = 0$ must always hold for all $\omega_l \in \Omega$. Thus, at any optimal solution, we must have $p^*_{\omega_j} = 0$. Let $0 < \varepsilon \leq \min\{\gamma^{*,(0)}_{\omega_i \omega_s} : \gamma^{*,(0)}_{\omega_i \omega_s} > 0, \forall \omega_i \in \Omega, \forall \omega_s \in \Omega\}$. Now, compared to $(\mathbf{p}^{*,(0)}, \boldsymbol{\gamma}^{*,(0)})$, where $\gamma^{*,(0)}_{\omega_i \omega_j} = 0$ for all $\omega_i \in \Omega$, we want to move to another *optimal* solution, denoted $(\mathbf{p}^{(1)}, \boldsymbol{\gamma}^{(1)})$, where $\gamma^{(1)}_{\omega_l \omega_j} = \varepsilon > 0$ for some $\omega_l \in \Omega$ with $\omega_j \in \mathcal{A}(\omega_l, \lambda^*)$. In this solution, $p^{(1)}_{\omega_j} \geq \varepsilon > 0$. Note that $\omega_l \notin \Omega_U(x^*, \lambda^*)$ because if $\omega_l \in \Omega_U(x^*, \lambda^*)$, then $\mathcal{A}(\omega_l, \lambda^*) = \{\omega_l\}$ and so we must have $\gamma^*_{\omega_l \omega_j} = 0$ by Lemma 1. So, $\omega_l \in \Omega_N(x^*, \lambda^*) \cup \Omega_M(x^*, \lambda^*)$. Observe that we can pick $\omega_l = \omega_j$; so there exists at least one such ω_l to construct $(\mathbf{p}^{(1)}, \boldsymbol{\gamma}^{(1)})$. In either case, we must have another scenario ω_k such that $\{\omega_j, \omega_k\} \subseteq \mathcal{A}(\omega_l, \lambda^*)$, $h_{\omega_k}(x^*) > h_{\omega_j}(x^*)$, and $\omega_k \in \Omega_U(x^*, \lambda^*)$ by Lemma 4 satisfying $\gamma^{*,(0)}_{\omega_l \omega_k} > 0$ by condition (ii). We already know the existence of at least one such scenario ω_k if $\omega_l \in \Omega_M(x^*, \lambda^*)$ (see the paragraph after the proof of Lemma 3). If $\omega_l \in \Omega_N(x^*, \lambda^*)$, then if $\mathcal{A}(\omega_l, \lambda^*) = \{\omega_j\}$, this means that $\omega_j \in \Omega_U(x^*, \lambda^*)$, which contradicts with $\omega_j \in \Omega_M(x^*, \lambda^*)$. So, such a scenario ω_k also exists if $\omega_l \in \Omega_N(x^*, \lambda^*)$.

Because constraint (3c) must be satisfied at ω_l, in order to have $\gamma^{(1)}_{\omega_l \omega_j} = \varepsilon = \varepsilon + \gamma^{*,(0)}_{\omega_l \omega_j} > 0$, we must take ε mass away from $\gamma^{*,(0)}_{\omega_l\omega}$ for scenarios ω that have the highest-cost in $\mathcal{A}(\omega_l, \lambda^*)$ by condition (ii). Without loss of generality, let $\gamma^{(1)}_{\omega_l\omega_k} = \gamma^{*,(0)}_{\omega_l\omega_k} - \varepsilon$. If all other components of $\boldsymbol{\gamma}^{(1)}$ are left the same as $\boldsymbol{\gamma}^{*,(0)}$, then the resulting $(\mathbf{p}^{(1)}, \boldsymbol{\gamma}^{(1)})$ would constitute a feasible but suboptimal solution. In this case, $p^{(1)}_{\omega_j} = p^{*,(0)}_{\omega_j} + \varepsilon$, $p^{(1)}_{\omega_k} = p^{*,(0)}_{\omega_k} - \varepsilon$, and all other $p^{(1)}_{\omega}$ are the same as $p^{*,(0)}_{\omega}$. Then, constraints (3b) and (3c) are satisfied. To see why (3d) is satisfied, first note that because $h_{\omega_k}(x^*) > h_{\omega_j}(x^*)$, $\{\omega_k, \omega_j\} \subseteq \mathcal{A}(\omega_l, \lambda^*)$, and $\lambda^* > 0$, we have $c_{\omega_l\omega_k} > c_{\omega_l\omega_j}$. So, the left-hand side of constraint (3d) is $\varepsilon \cdot (c_{\omega_l\omega_k} - c_{\omega_l\omega_j})$ less in the resulting $(\mathbf{p}^{(1)}, \boldsymbol{\gamma}^{(1)})$ than that in $(\mathbf{p}^{*,(0)}, \boldsymbol{\gamma}^{*,(0)})$. It is also suboptimal because the objective function value (3a) with the resulting $(\mathbf{p}^{(1)}, \boldsymbol{\gamma}^{(1)})$ is $\varepsilon \cdot (h_{\omega_k}(x^*) - h_{\omega_j}(x^*)) > 0$ less than that of $(\mathbf{p}^{*,(0)}, \boldsymbol{\gamma}^{*,(0)})$.

To get the same optimal value with $(\mathbf{p}^{(1)}, \boldsymbol{\gamma}^{(1)})$, we have to compensate for the loss in the objective $(\varepsilon \cdot (h_{\omega_k}(x^*) - h_{\omega_j}(x^*) > 0)$. This means some probability mass in other parts of $(\mathbf{p}^{*,(0)}, \boldsymbol{\gamma}^{*,(0)})$ should be transferred from lower-cost scenario(s) to higher cost scenario(s) while obeying all constraints of (3). By Lemma 1, under a fixed λ^*, such transfers must happen between scenarios with different costs but those that belong to the same $\mathcal{A}(\omega_s, \lambda^*)$ for some $\omega_s \in \Omega$. Note that $\omega_s \notin \Omega_U(x^*, \lambda^*)$ because otherwise $\mathcal{A}(\omega_s, \lambda^*) = \{\omega_s\}$. Now, if $\omega_s = \omega_l$, we must change $(\mathbf{p}^{(1)}, \boldsymbol{\gamma}^{(1)})$ to $(\mathbf{p}^{*,(0)}, \boldsymbol{\gamma}^{*,(0)})$ to compensate for the loss in the objective because condition (ii) indicates that, in $(\mathbf{p}^{*,(0)}, \boldsymbol{\gamma}^{*,(0)})$, there does not exist a scenario $\omega_u \neq \omega_j$ that is not a highest-cost scenario among $\mathcal{A}(\omega_s, \lambda^*)$ with $\gamma^{*,(0)}_{\omega_l\omega_u} > 0$. If $\omega_s \neq \omega_l$, such a scenario ω_s cannot exist. This is similarly because condition (ii) indicates that in $(\mathbf{p}^{*,(0)}, \boldsymbol{\gamma}^{*,(0)})$, there does not exist a scenario ω_u that is not a highest-cost scenario among $\mathcal{A}(\omega_s, \lambda^*)$ with $\gamma^{*,(0)}_{\omega_s\omega_u} > 0$. This means we cannot compensate for the loss in the objective for $(\mathbf{p}^{(1)}, \boldsymbol{\gamma}^{(1)})$ and therefore it is not optimal. As the above holds for any chosen ω_l, $p^*_{\omega_j}$ must be zero at λ^*. Because $\omega_j \in \Omega_M(x^*, \lambda^*)$ was chosen arbitrarily, $p^*_{\omega_j} = 0$ for all $\omega_j \in \Omega_M(x^*, \lambda^*)$. □

Proposition 6 provides a number of conditions that must be satisfied so that all $\omega_j \in \Omega_M(x^*, \lambda^*)$ have uniquely zero worst-case probabilities, $p^*_{\omega_j} = 0$, at all primal optimal solutions at that λ^*. However, even when these conditions are not met, we are able to deduce additional information for some scenarios in $\Omega_M(x^*, \lambda^*)$ by looking at the remaining transportation budget. Specifically, we know from Lemma 1 that for any $\omega_i \in \Omega_N(x^*, \lambda^*)$, we must have $\gamma^*_{\omega_i\omega_i} = 0$, and by constraint (3c), we also must have $\sum_{\omega \in \mathcal{A}(\omega_i, \lambda^*)} \gamma^*_{\omega_i\omega} = q_{\omega_i}$. This means that there exists a positive *minimum* transportation cost to satisfy constraint (3c). If we subtract such minimum transportation costs for every $\omega_i \in \Omega_N(x^*, \lambda^*)$ from the right-hand side of (3d), and the remaining transportation budget is not sufficient to move all probability mass from a scenario $\omega_j \in \Omega_M(x^*, \lambda^*)$, then ω_j must have a positive worst-case probability at λ^*. Observe that for any scenario $\omega_i \in \Omega_U(x^*, \lambda^*)$, because $\mathcal{A}(\omega_i, \lambda^*) = \{\omega_i\}$, we have $\gamma^*_{\omega_i\omega_i} = q_{\omega_i}$ and this does not contribute to the transportation cost because $c_{\omega_i\omega_i} = 0$. Proposition 7 formalizes this discussion. Recall the notation $\underline{c}^{\lambda^*}_{\omega_i} = \min_{\omega_j \in \mathcal{A}(\omega_i, \lambda^*) \setminus \{\omega_i\}} c_{\omega_i\omega_j}$.

Proposition 7. *Consider the worst-case expected problem of DRO-W (3) as described in Theorem 1. Let ρ_r be the minimum transportation budget needed to move all probability mass from scenarios in $\Omega_N(x^*, \lambda^*)$ at λ^*, that is, $\rho_r = \sum_{\omega_i \in \Omega_N(x^*,\lambda^*)} q_{\omega_i} \cdot \underline{c}_{\omega_i}^{\lambda^*}$. Any scenario $\omega_j \in \Omega_M(x^*, \lambda^*)$ satisfying $\rho - \rho_r < q_{\omega_j} \cdot \underline{c}_{\omega_j}^{\lambda^*}$ must have $p^*_{\omega_j} > 0$ at that λ^*.*

Proof. At a fixed λ^*, (3d) can be reformulated as $\sum_{\omega_i \in \Omega_N(x^*,\lambda^*)} \sum_{\omega_k \in \mathcal{A}(\omega_i,\lambda^*)} c_{\omega_i\omega_k} \gamma_{\omega_i\omega_k} + \sum_{\omega_i \in \Omega_M(x^*,\lambda^*)} \sum_{\omega_k \in \mathcal{A}(\omega_i,\lambda^*)} c_{\omega_i\omega_k} \gamma_{\omega_i\omega_k} \leq \rho$ by Lemma 1. Since $\sum_{\omega_i \in \Omega_N(x^*,\lambda^*)} \sum_{\omega_k \in \mathcal{A}(\omega_i,\lambda^*)} c_{\omega_i\omega_k} \gamma^*_{\omega_i\omega_k} \geq \rho_r$, $\sum_{\omega_i \in \Omega_M(x^*,\lambda^*)} \sum_{\omega_k \in \mathcal{A}(\omega_i,\lambda^*)} c_{\omega_i\omega_k} \gamma^*_{\omega_i\omega_k} \leq \rho - \rho_r$ must hold for any $\boldsymbol{\gamma}^*$ at λ^*. Now consider scenario $\omega_j \in \Omega_M(x^*, \lambda^*)$ satisfying $\rho - \rho_r < q_{\omega_j} \cdot \underline{c}_{\omega_j}^{\lambda^*}$. Recall $\omega_j \in \mathcal{A}(\omega_j, \lambda^*)$, and note that constraint (3c) for ω_j is equivalent to $\sum_{\omega_k \in \mathcal{A}(\omega_j,\lambda^*)} \gamma_{\omega_j\omega_k} = q_{\omega_j}$ by Lemma 1. Since $q_{\omega_j} > 0$, $\gamma^*_{\omega_j\omega_j} > 0$ must hold. Otherwise, $\sum_{\omega_k \in \mathcal{A}(\omega_j,\lambda^*)\setminus\{\omega_j\}} c_{\omega_j\omega_k} \gamma^*_{\omega_j\omega_k} > \rho - \rho_r$ and $\boldsymbol{\gamma}^*$ is infeasible to (3d). Thus, we must have $p^*_{\omega_j} \geq \gamma^*_{\omega_j\omega_j} > 0$ at λ^*. □

This concludes the results for a fixed λ^*. We next study the case when Λ^* is not a singleton.

4.4 Results for Multiple λ^*

4.4.1 Changes in Categories

We first discuss how the categorization of a given scenario changes within the partition $\Omega_N(x^*, \lambda^*)$, $\Omega_U(x^*, \lambda^*)$, and $\Omega_M(x^*, \lambda^*)$ when there are multiple optimal dual solutions λ^* corresponding to constraint (3d). For problems we consider, the set of dual solutions Λ^* is bounded. Therefore, we have $\Lambda^* = [\lambda^*_{\min}, \lambda^*_{\max}]$ with $\lambda^*_{\max} > 0$. Also, in this case, $\lambda^*_{\min}$ is the so-called *right shadow price* of constraint (3d) and $\lambda^*_{\max}$ is the *left shadow price* of constraint (3d); see [46, Theorem 5.6.2]. The right shadow price of (3d), $\lambda^*_{\min}$, is defined as the rate at which the optimal objective function value of (3) increases by a small positive perturbation of the right-hand side of (3d); i.e., when ρ is changed to $\rho + \varepsilon$ for some small $\varepsilon > 0$. Similarly, the left shadow price of (3d), $\lambda^*_{\max}$, is defined as the rate at which the optimal value decreases by a negative perturbation of the right-hand side of (3d); i.e., when ρ is changed to $\rho - \varepsilon$ for some small $\varepsilon > 0$. Recall also that the optimal value of (3), as a function of ρ, is a piecewise linear concave function. At a point where two "pieces" of this function intersect, there are multiple optimal λ^* and the right/left shadow prices are the slopes of the corresponding pieces [46, Theorem 5.4.2 and Figure 5.16]. Note that the worst-case expected problem (3) does not become infeasible (recall $\rho > 0$) or unbounded when the right-hand side of (3d) is perturbed by a small amount in our setting.

We begin with Lemmas 5 and 6, which form the foundation of all results in this section.

Lemma 5. *Consider the worst-case expected problem of DRO-W (3) as described in Theorem 1. Let the set of optimal dual solutions corresponding to constraint (3d) be $\Lambda^* := \Lambda^*(x^*) = [\lambda^*_{\min}, \lambda^*_{\max}]$ with $0 \leq \lambda^*_{\min} < \lambda^*_{\max}$. Consider any $(\mathbf{p}^*, \boldsymbol{\gamma}^*)$ that solve (3) at the given x^* and $\lambda^* \in \Lambda^*$. Then, for any $\omega_i, \omega_j \in \Omega$ such that $\gamma^*_{\omega_i\omega_j} > 0$ (ω_i, ω_j can be the same), $\frac{h_{\omega_j}(x^*) - h_{\omega_k}(x^*)}{c_{\omega_i\omega_j} - c_{\omega_i\omega_k}} \geq \lambda^*_{\max}$ must happen for any $\omega_k \in \Omega$ such that $h_{\omega_k}(x^*) < h_{\omega_j}(x^*)$ and $c_{\omega_i\omega_k} < c_{\omega_i\omega_j}$.*

Proof. For the sake of contradiction, assume there exists $\omega_i, \omega_j, \omega_k$ such that $h_{\omega_k}(x^*) < h_{\omega_j}(x^*)$ and $c_{\omega_i\omega_k} < c_{\omega_i\omega_j}$ but $\frac{h_{\omega_j}(x^*)-h_{\omega_k}(x^*)}{c_{\omega_i\omega_j}-c_{\omega_i\omega_k}} = \lambda_0 < \lambda^*_{\max}$. Then, if we decrease the right-hand side of constraint (3d) by a small $\varepsilon > 0$ such that $\frac{\varepsilon}{c_{\omega_i\omega_j}-c_{\omega_i\omega_k}} = \delta < \gamma^*_{\omega_i\omega_j}$, we can have another set of primal feasible solutions to (3), denoted $(\mathbf{p}^1, \boldsymbol{\gamma}^1)$, such that $p^1_{\omega_j} = p^*_{\omega_j} - \delta$, $\gamma^1_{\omega_i\omega_j} = \gamma^*_{\omega_i\omega_j} - \delta$, $p^1_{\omega_k} = p^*_{\omega_k} + \delta$, $\gamma^1_{\omega_i\omega_k} = \gamma^*_{\omega_i\omega_k} + \delta$, and all other components of $(\mathbf{p}^1, \boldsymbol{\gamma}^1)$ are the same as $(\mathbf{p}^*, \boldsymbol{\gamma}^*)$. Then, in $\mathbf{p}^1$, compared to $\mathbf{p}^*$, the objective function value of (3) decreases by $\varepsilon \cdot \lambda_0$. By Theorem 5.6.2 of [46] and the fact that $(\mathbf{p}^1, \boldsymbol{\gamma}^1)$ is a feasible but not necessarily optimal solution at $\rho + \varepsilon$, we have $\lambda^*_{\max} \le \lambda_0$, which results in a contradiction. □

Lemma 6. *Consider the setting of the worst-case expected problem of DRO-W* (3) *with the notation described in Lemma 5. Then, for any $\omega_i, \omega_j \in \Omega$ such that $\gamma^*_{\omega_i\omega_j} > 0$ (ω_i, ω_j can be the same), $\frac{h_{\omega_k}(x^*)-h_{\omega_j}(x^*)}{c_{\omega_i\omega_k}-c_{\omega_i\omega_j}} \le \lambda^*_{\min}$ must happen for any $\omega_k \in \Omega$ such that $h_{\omega_k}(x^*) > h_{\omega_j}(x^*)$.*

Proof. Since $\gamma^*_{\omega_i\omega_j} > 0$, by Lemma 1, $\omega_j \in \mathcal{A}(\omega_i, \lambda^*)$. Therefore, we must have both $c_{\omega_i\omega_k} > c_{\omega_i\omega_j}$ and $\lambda^* > 0$ because $h_{\omega_k}(x^*) > h_{\omega_j}(x^*)$. Otherwise, if $\lambda^* = 0$, since $h_{\omega_k}(x^*) > h_{\omega_j}(x^*)$, $\omega_j \notin \mathcal{A}(\omega_i, \lambda^*)$. Else, if $\lambda^* > 0$, we have $h_{\omega_k}(x^*) - \lambda^* c_{\omega_i\omega_k} > h_{\omega_j}(x^*) - \lambda^* c_{\omega_i\omega_j}$, which means again $\omega_j \notin \mathcal{A}(\omega_i, \lambda^*)$. Now for the sake of contradiction, assume there exists $\omega_i, \omega_j, \omega_k$ such that $h_{\omega_k}(x^*) > h_{\omega_j}(x^*)$ but $\frac{h_{\omega_k}(x^*)-h_{\omega_j}(x^*)}{c_{\omega_i\omega_k}-c_{\omega_i\omega_j}} = \lambda_0 > \lambda^*_{\min}$. Then, if we increase the right-hand side of constraint (3d) by a small $\varepsilon > 0$ such that $\frac{\varepsilon}{c_{\omega_i\omega_k}-c_{\omega_i\omega_j}} = \delta < \gamma^*_{\omega_i\omega_j}$, we can have another set of primal feasible solutions to (3), denoted $(\mathbf{p}^1, \boldsymbol{\gamma}^1)$, such that $p^1_{\omega_j} = p^*_{\omega_j} - \delta$, $\gamma^1_{\omega_i\omega_j} = \gamma^*_{\omega_i\omega_j} - \delta$, $p^1_{\omega_k} = p^*_{\omega_k} + \delta$, $\gamma^1_{\omega_i\omega_k} = \gamma^*_{\omega_i\omega_k} + \delta$, and all other components of $(\mathbf{p}^1, \boldsymbol{\gamma}^1)$ are the same as $(\mathbf{p}^*, \boldsymbol{\gamma}^*)$. Then, in $\mathbf{p}^1$, compared to $\mathbf{p}^*$, the objective function value of (3) increases by $\varepsilon \cdot \lambda_0$. By Theorem 5.6.2 of [46] and the fact that $(\mathbf{p}^1, \boldsymbol{\gamma}^1)$ is a feasible but not necessarily optimal solution of $\rho - \varepsilon$, we have $\lambda^*_{\min} \ge \lambda_0$, which results in a contradiction. □

Armed with the above two lemmas, we now show how the categorization of a scenario belonging to $\Omega_N(x^*, \lambda^*)$, $\Omega_U(x^*, \lambda^*)$, or $\Omega_M(x^*, \lambda^*)$ for a fixed λ^* changes as λ^* varies in Propositions 8, 9, and 10, respectively.

Proposition 8. *Let $\Lambda^* = [\lambda^*_{\min}, \lambda^*_{\max}]$ with $0 \le \lambda^*_{\min} < \lambda^*_{\max}$ be the set of optimal dual solutions to* (3d) *in the worst-case expected problem of DRO-W* (3) *at an optimal solution $x^* \in \mathcal{X}$. Given $\lambda^*_1 \in \Lambda^*$, for any $\omega_i \in \Omega_N(x^*, \lambda^*_1)$, all the following conditions hold:*

(i) For any $\lambda^*_2 \in \Lambda^*$ such that $\lambda^*_2 < \lambda^*_1$, $\omega_i \in \Omega_N(x^*, \lambda^*_2)$;

(ii) If there exists a $\lambda^*_2 \in \Lambda^*$ such that $\omega_i \in \Omega_M(x^*, \lambda^*_2)$, then $\lambda^*_2 = \lambda^*_{\max}$;

(iii) For any $\lambda^*_2 \in \Lambda^*$, $\omega_i \notin \Omega_U(x^*, \lambda^*_2)$.

Proof. We first show part (i). Since $\omega_i \in \Omega_N(x^*, \lambda^*_1)$, there must exist a scenario $\omega_j \in \mathcal{A}(\omega_i, \lambda^*_1)$ such that $h_{\omega_j}(x^*) - \lambda^*_1 c_{\omega_i\omega_j} > h_{\omega_i}(x^*)$. This results in $h_{\omega_j}(x^*) - \lambda^*_2 c_{\omega_i\omega_j} > h_{\omega_j}(x^*) - \lambda^*_1 c_{\omega_i\omega_j} > h_{\omega_i}(x^*)$, which means $\omega_i \notin \mathcal{A}(\omega_i, \lambda_2^*)$. Then by Lemma 2, $\omega_i \in \Omega_N(x^*, \lambda_2^*)$; see also the paragraph after the proof of Lemma 3. Now we show parts (ii) and (iii). From part (i), we have $\lambda_2^* > \lambda_1^* \geq 0$. Also, since $\omega_i \in \Omega_N(x^*, \lambda_1^*)$ and $q_{\omega_i} > 0$, there must exist an optimal $(\mathbf{p}^*, \boldsymbol{\gamma}^*)$ at λ_1^* where $\gamma^*_{\omega_i\omega_j} > 0$ for some scenario $\omega_j \in \mathcal{A}(\omega_i, \lambda_1^*)$. From the proof of part (i), we also have $h_{\omega_j}(x^*) > h_{\omega_i}(x^*)$. Then by Lemma 5, $\frac{h_{\omega_j}(x^*) - h_{\omega_i}(x^*)}{c_{\omega_i\omega_j}} \geq \lambda^*_{\max}$. Now, if $\omega_i \in \Omega_M(x^*, \lambda_2^*)$, we have $h_{\omega_i}(x^*) \geq h_{\omega_j}(x^*) - \lambda_2^* c_{\omega_i\omega_j}$. This means $\lambda_2^* \geq \frac{h_{\omega_j}(x^*) - h_{\omega_i}(x^*)}{c_{\omega_i\omega_j}} \geq \lambda^*_{\max}$ and hence $\lambda_2^* = \lambda^*_{\max}$, which shows part (ii). On the other hand, if $\omega_i \in \Omega_U(x^*, \lambda_2^*)$, from Lemma 3, $\{\omega_i\} = \mathcal{A}(\omega_i, \lambda_2^*)$. Therefore we have $h_{\omega_i}(x^*) > h_{\omega_j}(x^*) - \lambda_2^* c_{\omega_i\omega_j}$. This means $\lambda_2^* > \frac{h_{\omega_j}(x^*) - h_{\omega_i}(x^*)}{c_{\omega_i\omega_j}} \geq \lambda^*_{\max}$, which forms a contradiction. This shows part (iii). □

Proposition 9. *Consider the conditions of Proposition 8. Given $\lambda_1^* \in \Lambda^*$, for any $\omega_i \in \Omega_U(x^*, \lambda_1^*)$, all the following conditions hold:*

(i) For any $\lambda_2^* \in \Lambda^*$ such that $\lambda_2^* > \lambda_1^*$, $\omega_i \in \Omega_U(x^*, \lambda_2^*)$;

(ii) If there exists a $\lambda_2^* \in \Lambda^*$ such that $\omega_i \in \Omega_M(x^*, \lambda_2^*)$, then $\lambda_2^* = \lambda^*_{\min}$;

(iii) For any $\lambda_2^* \in \Lambda^*$, $\omega_i \notin \Omega_N(x^*, \lambda_2^*)$.

Proof. We only need to show parts (i) and (ii) because part (iii) is indicated by Proposition 8(iii). (i) Since $\omega_i \in \Omega_U(x^*, \lambda_1^*)$, by Lemma 3, $\{\omega_i\} = \mathcal{A}(\omega_i, \lambda_1^*)$. This combined with $\lambda_2^* > \lambda_1^*$ indicates that $h_{\omega_i}(x^*) > h_{\omega_j}(x^*) - \lambda_1^* c_{\omega_i\omega_j} > h_{\omega_j}(x^*) - \lambda_2^* c_{\omega_i\omega_j}$ for any $\omega_j \in \Omega \setminus \{\omega_i\}$. This means $\{\omega_i\} = \mathcal{A}(\omega_i, \lambda_2^*)$ and hence $\omega_i \in \Omega_U(x^*, \lambda_2^*)$. (ii) From part (i), we know $\lambda_2^* < \lambda_1^*$. When $\lambda_2^* = 0$, then $\lambda_2^* = \lambda^*_{\min}$. Now let $\lambda_2^* > 0$. Since $\omega_i \in \Omega_M(x^*, \lambda_2^*)$, there exists a scenario ω_j satisfying $h_{\omega_i}(x^*) = h_{\omega_j}(x^*) - \lambda_2^* c_{\omega_i\omega_j}$ and $h_{\omega_j}(x^*) > h_{\omega_i}(x^*)$. This means $\frac{h_{\omega_j}(x^*) - h_{\omega_i}(x^*)}{c_{\omega_i\omega_j}} = \lambda_2^*$. Also, we have $\{\omega_i\} = \mathcal{A}(\omega_i, \lambda_1^*)$. This means there must exist an optimal $(\mathbf{p}^*, \boldsymbol{\gamma}^*)$ at λ_1^* where $\gamma^*_{\omega_i\omega_i} > 0$. By Lemma 6, $\frac{h_{\omega_j}(x^*) - h_{\omega_i}(x^*)}{c_{\omega_i\omega_j}} \leq \lambda^*_{\min}$ must hold and therefore we have $\lambda_2^* = \frac{h_{\omega_j}(x^*) - h_{\omega_i}(x^*)}{c_{\omega_i\omega_j}} = \lambda^*_{\min}$. □

Proposition 10. *Consider the conditions of Proposition 8. Given $\lambda_1^* \in \Lambda^*$ and the set $\Omega_M(x^*, \lambda_1^*)$, all the following conditions hold:*

(i) $\Omega_M(x^*, \lambda_1^*) = \varnothing$ for any $\lambda_1^* \in \left(\lambda^*_{\min}, \lambda^*_{\max}\right)$;

(ii) If scenario $\omega_i \in \Omega_M(x^*, \lambda^*_{\max})$, then $\omega_i \in \Omega_N(x^*, \lambda_1^*)$ for any $\lambda_1^* \in \Lambda^* \setminus \{\lambda^*_{\max}\}$;

(iii) If scenario $\omega_i \in \Omega_M(x^*, \lambda^*_{\min})$, then $\omega_i \in \Omega_U(x^*, \lambda_1^*)$ for any $\lambda_1^* \in \Lambda^* \setminus \{\lambda^*_{\min}\}$.

Proof. We only need to show part (i) because, given part (i), part (ii) is indicated by Proposition 8(ii) and part (iii) is indicated by Proposition 9(ii). Let scenario $\omega_i \in \Omega_M(x^*, \lambda_1^*)$ for some λ_1^*. Since $q_{\omega_i} > 0$, there must exist an optimal $(\mathbf{p}^*, \boldsymbol{\gamma}^*)$ at λ_1^* where either $\gamma^*_{\omega_i\omega_i} > 0$ or $\gamma^*_{\omega_i\omega_j} > 0$ for some scenario $\omega_j \in \mathcal{A}(\omega_i, \lambda_1^*)$ such that $h_{\omega_j}(x^*) > h_{\omega_i}(x^*)$. If $\gamma^*_{\omega_i\omega_i} > 0$, from Lemma 6, we have $\frac{h_{\omega_j}(x^*) - h_{\omega_i}(x^*)}{c_{\omega_i\omega_j}} \leq \lambda^*_{\min}$. On the other hand, if $\gamma^*_{\omega_i\omega_j} > 0$, from Lemma 5, we have

$\frac{h_{\omega_j}(x^*)-h_{\omega_i}(x^*)}{c_{\omega_i\omega_j}} \geq \lambda^*_{\text{max}}$. Note that both $\gamma^*_{\omega_i\omega_i} > 0$ and $\gamma^*_{\omega_i\omega_j} > 0$ cannot happen simultaneously; otherwise $\lambda^*_{\text{max}} \leq \lambda^*_{\text{min}}$ (but we have $\lambda^*_{\text{min}} < \lambda^*_{\text{max}}$). Also, since $\{\omega_i, \omega_j\} \subseteq \mathcal{A}(\omega_i, \lambda^*_1)$, we have $h_{\omega_i}(x^*) = h_{\omega_j}(x^*) - \lambda^*_1 c_{\omega_i\omega_j}$, which means $\lambda^*_1 = \frac{h_{\omega_j}(x^*)-h_{\omega_i}(x^*)}{c_{\omega_i\omega_j}}$. Therefore, either $\lambda^*_1 = \lambda^*_{\text{min}}$ or $\lambda^*_1 = \lambda^*_{\text{max}}$ must happen. □

Table 1 summarizes the results of Propositions 8–10. Given $\Lambda^* = [\lambda^*_{\text{min}}, \lambda^*_{\text{max}}]$ and any $\lambda^*_{\text{mid}} \in (\lambda^*_{\text{min}}, \lambda^*_{\text{max}})$, scenarios $\omega_j \in \Omega_N(x^*, \lambda^*_{\text{mid}})$ must still belong to $\Omega_N(x^*, \lambda^*)$ at any $\lambda^* \in \Lambda^* \setminus \{\lambda^*_{\text{max}}\}$ including $\lambda^* = \lambda^*_{\text{min}}$. At λ^*_{max}, these scenarios can either still belong to $\Omega_N(x^*, \lambda^*_{\text{max}})$ or switch to $\Omega_M(x^*, \lambda^*_{\text{max}})$. On the other hand, scenarios $\omega_j \in \Omega_U(x^*, \lambda^*_{\text{mid}})$ must still belong to $\Omega_U(x^*, \lambda^*)$ at any $\lambda^* \in \Lambda^* \setminus \{\lambda^*_{\text{min}}\}$ including $\lambda^* = \lambda^*_{\text{max}}$. At λ^*_{min}, these scenarios can either belong to $\Omega_U(x^*, \lambda^*_{\text{min}})$ or switch to $\Omega_M(x^*, \lambda^*_{\text{min}})$. Finally, $\Omega_M(x^*, \lambda^*_{\text{mid}})$ is always an empty set.

λ^*_{min}		$\lambda^*_{\text{mid}} \in (\lambda^*_{\text{min}}, \lambda^*_{\text{max}})$		λ^*_{max}
$\Omega_N(x^*, \lambda^*_{\text{min}})$	$\longleftrightarrow$	$\Omega_N(x^*, \lambda^*_{\text{mid}})$	$\longleftrightarrow$	$\Omega_N(x^*, \lambda^*_{\text{max}})$
$\Omega_M(x^*, \lambda^*_{\text{min}})$		$\varnothing$		$\Omega_M(x^*, \lambda^*_{\text{max}})$
$\Omega_U(x^*, \lambda^*_{\text{min}})$	$\longleftrightarrow$	$\Omega_U(x^*, \lambda^*_{\text{mid}})$	$\longleftrightarrow$	$\Omega_U(x^*, \lambda^*_{\text{max}})$

Table 1: How categorization of a scenario into sets $\Omega_{\mathfrak{c}}(x^*, \lambda^*)$, $\mathfrak{c} = N, M, U$ changes as λ^* changes (summary of Propositions 8–10).

4.4.2 Results at Smallest and Largest λ^*

Before we present the proofs of Theorems 1–2, we first show two additional results that form a bridge between the conclusions for multiple λ^* in Section 4.4.1 and those for a fixed λ^* in Section 4.3, focusing on λ^*_{max} and λ^*_{min}. Specifically, Proposition 11 shows that all scenarios belonging to $\Omega_N(x^*, \lambda^*)$ for some $\lambda^* \in \Lambda^* \setminus \{\lambda^*_{\text{max}}\}$ but switch to $\Omega_M(x^*, \lambda^*)$ at $\lambda^* = \lambda^*_{\text{max}}$ satisfy Proposition 6 at λ^*_{max}. Similarly, Proposition 12 shows that any scenarios belonging to $\Omega_U(x^*, \lambda^*)$ for some $\lambda^* \in \Lambda^* \setminus \{\lambda^*_{\text{min}}\}$ and switch to $\Omega_M(x^*, \lambda^*)$ at $\lambda^* = \lambda^*_{\text{min}}$ satisfy Proposition 7 at λ^*_{min}.

Proposition 11. *Let $\Lambda^* = [\lambda^*_{\text{min}}, \lambda^*_{\text{max}}]$ with $0 \leq \lambda^*_{\text{min}} < \lambda^*_{\text{max}}$ be the set of the optimal dual solutions to (3d) in the worst-case expected problem of DRO-W (3) at an optimal solution $x^* \in \mathcal{X}$. All scenarios ω_j such that $\omega_j \in \Omega_N(x^*, \lambda^*)$ at all $\lambda^* \in [\lambda^*_{\text{min}}, \lambda^*_{\text{max}})$ and $\omega_j \in \Omega_M(x^*, \lambda^*)$ at $\lambda^* = \lambda^*_{\text{max}}$ satisfy the conditions of Proposition 6 and thus have $p^*_{\omega_j} = 0$ at λ^*_{max}.*

Proof. From Proposition 10(ii), all scenarios in $\Omega_M(x^*, \lambda^*_{\text{max}})$ must belong to $\Omega_N(x^*, \lambda^*)$ for $\lambda^*_{\text{min}} \leq \lambda^* < \lambda^*_{\text{max}}$. We first show at λ^*_{max}, Proposition 6(i) must be satisfied. For the sake of contradiction, suppose $p^*_{\omega_j} > 0$ at λ^*_{max} for some $\omega_j \in \Omega_M(x^*, \lambda^*_{\text{max}})$. Then, by (3b) and Lemma 1, there must be a scenario ω_i such that $\omega_j \in \mathcal{A}(\omega_i, \lambda^*_{\text{max}})$ and $\gamma^*_{\omega_i\omega_j} > 0$ at λ^*_{max}. Also, by Lemma 4, there must be a scenario $\omega_k \in \Omega_U(x^*, \lambda^*_{\text{max}})$ satisfying $\omega_k \in \mathcal{A}(\omega_i, \lambda^*_{\text{max}})$ and $h_{\omega_k}(x^*) > h_{\omega_j}(x^*)$. Also, since $\{\omega_j, \omega_k\} \subseteq \mathcal{A}(\omega_i, \lambda^*_{\text{max}})$, we have $h_{\omega_k}(x^*) - \lambda^*_{\text{max}} c_{\omega_i\omega_k} = h_{\omega_j}(x^*) - \lambda^*_{\text{max}} c_{\omega_i\omega_j}$, which means

$\lambda^*_{\max} = \frac{h_{\omega_k}(x^*) - h_{\omega_j}(x^*)}{c_{\omega_i\omega_k} - c_{\omega_i\omega_j}}$. Also, since $\gamma^*_{\omega_i\omega_j} > 0$, by Lemma 6, we have $\frac{h_{\omega_k}(x^*) - h_{\omega_j}(x^*)}{c_{\omega_i\omega_k} - c_{\omega_i\omega_j}} \leq \lambda^*_{\min}$. This means $\lambda^*_{\min} = \lambda^*_{\max}$, which forms a contradiction to $\lambda^*_{\min} < \lambda^*_{\max}$. Following the same logic, for any $\omega_i \in \Omega_N(x^*, \lambda^*_{\max}) \cup \Omega_M(x^*, \lambda^*_{\max})$, if there exists $\{\omega_s, \omega_k\} \subseteq \mathcal{A}(\omega_i, \lambda^*_{\max})$ with $h_{\omega_k}(x^*) > h_{\omega_s}(x^*)$, we must have $\gamma^*_{\omega_i\omega_s} = 0$; otherwise we get $\lambda^*_{\min} = \lambda^*_{\max}$. This shows Proposition 6(ii) must also be satisfied at $\lambda^*_{\max}$, and therefore $p^*_{\omega_j} = 0$ must happen. □

Proposition 12. *Let $\Lambda^* = [\lambda^*_{\min}, \lambda^*_{\max}]$ with $0 \leq \lambda^*_{\min} < \lambda^*_{\max}$ be the set of the optimal dual solutions to (3d) in the worst-case expected problem of DRO-W (3) at an optimal solution $x^* \in \mathcal{X}$. All scenarios ω_j such that $\omega_j \in \Omega_U(x^*, \lambda^*)$ at all $\lambda^* \in (\lambda^*_{\min}, \lambda^*_{\max}]$ and $\omega_j \in \Omega_M(x^*, \lambda^*)$ at $\lambda^* = \lambda^*_{\min}$ satisfy the conditions of Proposition 7 and thus have $p^*_{\omega_j} > 0$ at $\lambda^*_{\min}$.*

Proof. Consider any λ^* satisfying $\lambda^*_{\min} < \lambda^* < \lambda^*_{\max}$. We first show that given any $\omega_i \in \Omega_N(x^*, \lambda^*)$, for any $\omega_k \in \mathcal{A}(\omega_i, \lambda^*)$ and any $\omega_s \in \mathcal{A}(\omega_i, \lambda^*_{\min})$, we must have $h_{\omega_s}(x^*) \geq h_{\omega_k}(x^*)$ and $c_{\omega_i\omega_s} \geq c_{\omega_i\omega_k}$.

Let ω_u be one of the highest-cost scenarios in $\mathcal{A}(\omega_i, \lambda^*)$. Since $\lambda^* > 0$, $c_{\omega_i\omega_u} \geq c_{\omega_i\omega_k}$ for any $\omega_k \in \mathcal{A}(\omega_i, \lambda^*)$. Then any scenario ω_l such that $h_{\omega_l}(x^*) < h_{\omega_u}(x^*)$ will not belong to $\mathcal{A}(\omega_i, \lambda^*_{\min})$. The reason is the following. First, if $c_{\omega_i\omega_l} \geq c_{\omega_i\omega_u}$, we have $h_{\omega_l}(x^*) - \lambda^*_{\min}c_{\omega_i\omega_l} < h_{\omega_u}(x^*) - \lambda^*_{\min}c_{\omega_i\omega_u}$. So we have $\omega_l \notin \mathcal{A}(\omega_i, \lambda^*_{\min})$. Second, if $c_{\omega_i\omega_l} < c_{\omega_i\omega_u}$, because $\omega_u \in \mathcal{A}(\omega_i, \lambda^*)$, we have $h_{\omega_l}(x^*) - \lambda^* c_{\omega_i\omega_l} \leq h_{\omega_u}(x^*) - \lambda^* c_{\omega_i\omega_u}$. This indicates $h_{\omega_l}(x^*) - \lambda^*_{\min}c_{\omega_i\omega_l} < h_{\omega_u}(x^*) - \lambda^*_{\min}c_{\omega_i\omega_u}$. So again $\omega_l \notin \mathcal{A}(\omega_i, \lambda^*_{\min})$. Now instead consider any scenario ω_l such that $h_{\omega_l}(x^*) \geq h_{\omega_u}(x^*)$. Then, it is impossible to have $h_{\omega_l}(x^*) \geq h_{\omega_u}(x^*)$ but $c_{\omega_i\omega_l} < c_{\omega_i\omega_u}$. Otherwise we have $h_{\omega_l}(x^*) - \lambda^* c_{\omega_i\omega_l} > h_{\omega_u}(x^*) - \lambda^* c_{\omega_i\omega_u}$ and $\omega_u \notin \mathcal{A}(\omega_i, \lambda^*)$, which is a contradiction. In the remaining case (i.e., $h_{\omega_l}(x^*) \geq h_{\omega_u}(x^*)$ and $c_{\omega_i\omega_l} \geq c_{\omega_i\omega_u}$), we are not able to say if $\omega_l \in \mathcal{A}(\omega_i, \lambda^*_{\min})$ hold but in any case, for any $\omega_s \in \mathcal{A}(\omega_i, \lambda^*_{\min})$, we must have $h_{\omega_s}(x^*) \geq h_{\omega_u}(x^*)$ and $c_{\omega_i\omega_s} \geq c_{\omega_i\omega_u}$.

At λ^*, by Proposition 10(i), $\Omega_M(x^*, \lambda^*) = \varnothing$. Also, since $\lambda^*_{\max} > \lambda^* > 0$, constraint (3d) holds with equality. Therefore, constraint (3d) can be reformulated at λ^* as $\rho = \sum_{\omega_i \in \Omega_N(x^*, \lambda^*)} \sum_{\omega_k \in \mathcal{A}(\omega_i, \lambda^*)} c_{\omega_i\omega_k}\gamma^*_{\omega_i\omega_k}$. Then,

$$\begin{aligned}
\rho = \sum_{\omega_i \in \Omega_N(x^*, \lambda^*)} \sum_{\omega_k \in \mathcal{A}(\omega_i, \lambda^*)} c_{\omega_i\omega_k}\gamma^*_{\omega_i\omega_k} &\leq \sum_{\omega_i \in \Omega_N(x^*, \lambda^*)} q_{\omega_i} \cdot \max_{\omega_k \in \mathcal{A}(\omega_i, \lambda^*)} c_{\omega_i\omega_k} \\
&= \sum_{\omega_i \in \Omega_N(x^*, \lambda^*_{\min})} q_{\omega_i} \cdot \max_{\omega_k \in \mathcal{A}(\omega_i, \lambda^*)} c_{\omega_i\omega_k} \\
&\leq \sum_{\omega_i \in \Omega_N(x^*, \lambda^*_{\min})} q_{\omega_i} \cdot \min_{\omega_k \in \mathcal{A}(\omega_i, \lambda^*_{\min})} c_{\omega_i\omega_k} = \rho_r,
\end{aligned}$$

where the first inequality holds because $\sum_{\omega_k \in \mathcal{A}(\omega_i, \lambda^*)} \gamma^*_{\omega_i\omega_k} = q_{\omega_i}$ (see (3c) and Lemma 1), the next equality holds because $\Omega_N(x^*, \lambda^*_{\min}) = \Omega_N(x^*, \lambda^*)$ for $\lambda^*_{\min} < \lambda^* < \lambda^*_{\max}$ (see Proposition 8(i)), and the last inequality holds because of the earlier result shown in the proof.

This means $\rho_r = \rho$ at $\lambda^*_{\min}$. Therefore, any scenario $\omega_j \in \Omega_M(x^*, \lambda^*_{\min})$ satisfies Proposition 7 because $q_{\omega_j} > 0$. Moreover, because $\rho_r = \rho$ and constraint (3c) must be satisfied at $\omega_j \in \Omega_M(x^*, \lambda^*_{\min})$, we have $\gamma^*_{\omega_j\omega_j} = q_{\omega_j}$ at $\lambda^*_{\min}$. This means $p^*_{\omega_j} \geq q^*_{\omega_j} > 0$ at $\lambda^*_{\min}$ by constraint (3b). □

4.5 Proof of Main Results

We are now ready to present the proofs of Theorems 1 and 2.

Proof of Theorem 1. (i) Let $\underline{c} = \min_{\Omega\setminus\{\omega_j\}} c_{\omega_j\omega_i}$. Since $\rho < c_{\omega_j\omega_i} q_{\omega_j}$ for any $\omega_i \in \Omega \setminus \{\omega_j\}$ we have $\rho < \underline{c} q_{\omega_j}$. We now show $p^*_{\omega_j}$ is always positive and so effective by Proposition 1. Suppose not; i.e., there exists an optimal solution $\mathbf{p}^*$ to (3) such that $p^*_{\omega_j} = 0$. Then, in this solution, we have $\gamma^*_{\omega_j\omega_j} = 0$ by (3b). Furthermore, by constraint (3c), $\sum_{\omega_i\in\Omega\setminus\{\omega_j\}} \gamma^*_{\omega_j\omega_i} = q_{\omega_j}$. As a result, $\sum_{\omega_i\in\Omega\setminus\{\omega_j\}} c_{\omega_j\omega_i}\gamma^*_{\omega_j\omega_i} \geq \sum_{\omega_i\in\Omega\setminus\{\omega_j\}} \gamma^*_{\omega_j\omega_i} \cdot \underline{c} = \underline{c} \cdot q_{\omega_j} > \rho$. This means (3d) is violated and results in a contradiction.

(ii) When Λ^* is a singleton, part (ii) is a direct result of Proposition 5 and Proposition 1. Now we discuss the case when Λ^* is not a singleton. At any $\bar{\lambda}^* \in \Lambda^*$, if $\omega_j \in \Omega_U(x^*, \bar{\lambda}^*)$, from Proposition 9, either $\omega_j \in \Omega_U(x^*, \lambda^*)$ for all $\lambda^* \in \Lambda^*$, or $\omega_j \in \Omega_U(x^*, \lambda^*)$ for any $\lambda^* > \lambda^*_{\min}$ and $\omega_j \in \Omega_M(x^*, \lambda^*_{\min})$. If $\omega_j \in \Omega_U(x^*, \lambda^*)$ for all $\lambda^* \in \Lambda^*$, Proposition 5 shows $p^*_{\omega_j} > 0$ at every $\lambda^* \in \Lambda^*$, and hence ω_j is effective by Proposition 1. If $\omega_j \in \Omega_U(x^*, \lambda^*)$ for any $\lambda^* > \lambda^*_{\min}$ and $\omega_j \in \Omega_M(x^*, \lambda^*_{\min})$, Proposition 5 shows $p^*_{\omega_j} > 0$ at every $\lambda^* > \lambda^*_{\min}$, while Proposition 12 shows $p^*_{\omega_j} > 0$ at $\lambda^* = \lambda^*_{\min}$. Therefore, ω_j is effective by Proposition 1.

(iii) When Λ^* is a singleton, part (iii) is a direct result of Proposition 7 and then Proposition 1. Now we discuss the case when Λ^* is not a singleton. By Propositions 10, 11 and 12, part (iii) can only happen when $\lambda^* = \lambda^*_{\min}$, where Proposition 7 shows $p^*_{\omega_j} > 0$. Also by Proposition 10(iii), $\omega_j \in \Omega_U(x^*, \lambda^*)$ at any $\lambda^* > \lambda^*_{\min}$ and Proposition 5 shows $p^*_{\omega_j} > 0$ at every $\lambda^* > \lambda^*_{\min}$. Therefore ω_j is effective by Proposition 1. □

Proof of Theorem 2. (i) When Λ^* is a singleton, part (i) is a direct result of Proposition 4 and Proposition 2. Now we discuss the case when Λ^* is not a singleton. At any $\bar{\lambda}^* \in \Lambda^*$, if $\omega_j \in \Omega_N(x^*, \bar{\lambda}^*)$, from Proposition 8, either $\omega_j \in \Omega_N(x^*, \lambda^*)$ all any $\lambda^* \in \Lambda^*$, or $\omega_j \in \Omega_N(x^*, \lambda^*)$ for any $\lambda^* < \lambda^*_{\max}$ and $\omega_j \in \Omega_M(x^*, \lambda^*_{\max})$. If $\omega_j \in \Omega_N(x^*, \lambda^*)$ for all $\lambda^* \in \Lambda^*$, Proposition 4 shows $p^*_{\omega_j} = 0$ at every $\lambda^* \in \Lambda^*$ and hence ω_j is effective by Proposition 2. If $\omega_j \in \Omega_N(x^*, \lambda^*)$ for any $\lambda^* < \lambda^*_{\max}$ and $\omega_j \in \Omega_M(x^*, \lambda^*_{\max})$, Proposition 4 shows $p^*_{\omega_j} = 0$ at every $\lambda^* < \lambda^*_{\max}$, while Proposition 11 shows $p^*_{\omega_j} = 0$ at $\lambda^* = \lambda^*_{\max}$. Therefore ω_j is ineffective by Proposition 2.

(ii) When Λ^* is a singleton, part (ii) is a direct result of Proposition 6 and Proposition 2. Now we discuss the case when Λ^* is not a singleton. By Propositions 10, 11 and 12, part (ii) can only happen when $\lambda^* = \lambda^*_{\max}$, where Proposition 6 shows $p^*_{\omega_j} = 0$. Also by Proposition 10(ii), $\omega_j \in \Omega_N(x^*, \lambda^*)$ at any $\lambda^* < \lambda^*_{\max}$, and Proposition 4 shows $p^*_{\omega_j} = 0$ at every $\lambda^* < \lambda^*_{\max}$. Therefore ω_j is ineffective by Proposition 2.

(iii) From the proof of (i)–(ii), all scenarios being identified as ineffective are identified by Proposition 2. Therefore they are collectively ineffective by Proposition 2. □

4.6 Discussion

The above results imply that when Λ^* is not a singleton set, i.e., when there are multiple optimal λ^*, we can identify the effectiveness of *all* scenarios. By Proposition 10, $\Omega_M(x^*, \lambda^*)$ is an empty set for all $\lambda^* \in (\lambda^*_{\text{min}}, \lambda^*_{\text{max}})$ and we only have $\Omega_U(x^*, \lambda^*)$ and $\Omega_N(x^*, \lambda^*)$. These scenarios are effective and ineffective, respectively by Theorem 1(ii) and Theorem 2(i). The scenarios that switch from $\Omega_N(x^*, \lambda^*)$ for $\lambda^* \in [\lambda^*_{\text{min}}, \lambda^*_{\text{max}})$ to $\Omega_M(x^*, \lambda^*)$ at $\lambda^* = \lambda^*_{\text{max}}$ are fully captured by condition (ii) in Theorem 2 (see Proposition 11) and hence remain ineffective. Similarly, scenarios that switch from $\Omega_U(x^*, \lambda^*)$ for $\lambda^* \in (\lambda^*_{\text{min}}, \lambda^*_{\text{max}}]$ to $\Omega_M(x^*, \lambda^*)$ at $\lambda^* = \lambda^*_{\text{min}}$ are fully captured by condition (iii) in Theorem 1 (see Proposition 12) and hence remain effective. We formalize this in Corollary 1.

Corollary 1. *When Λ^* is not a singleton, Theorems 1 and 2 identify the effectiveness of all scenarios in Ω.*

However, when Λ^* is singleton, i.e., when there is a unique optimal λ^* (which constitutes majority of the cases), while the conditions in Theorem 1(iii) and Theorem 2(ii) help us identify the effectiveness of *some* scenarios in $\Omega_M(x^*, \lambda^*)$, they may not be able to identify all scenarios. In our numerical results, these conditions perform quite successfully even in these cases.

Another case when the effectiveness of *all* scenarios can be identified by the proposed conditions is whenever $\Omega_M(x^*, \lambda^*)$ is an empty set, which may also happen when there is a unique λ^* (in addition to multiple optimal λ^*, e.g., at all $\lambda^*_{\text{mid}} \in (\lambda^*_{\text{min}}, \lambda^*_{\text{max}})$). Our numerical experiments show that this case can occur occasionally. Before presenting the numerical experiments, we first compare the results of this paper to those that were obtained for DRO-TV in [39].

5 Comparison between DRO-W and DRO-TV

Recall that when the transportation costs in DRO-W are defined as $c_{\omega_i\omega_j} = 1$ if $\omega_i \neq \omega_j$ and 0 otherwise, DRO-W becomes equivalent to DRO-TV. In this section, we examine Theorems 1–2 with this definition of transportation costs and show their equivalence to the easy-to-check conditions proposed in [39] for DRO-TV. To begin, let us provide a brief overview of DRO-TV and revisit the conditions proposed in [39].

5.1 Review of Effective Scenarios in DRO-TV

DRO-TV is formulated as the DRO problem (1) with the ambiguity set $\mathcal{P}$ defined as

$$\mathcal{P}_{\text{TV}} = \left\{ \mathbf{p} : \sum_{\omega_i \in \Omega} \frac{1}{2} |p_{\omega_i} - q_{\omega_i}| \leq \rho, \sum_{\omega_i \in \Omega} p_{\omega_i} = 1, \mathbf{p} \geq 0 \right\}. \tag{7}$$

Above, the total variation distance between $\mathbf{p}$ and $\mathbf{q}$, $\sum_{\omega_i \in \Omega} \frac{1}{2}|p_{\omega_i} - q_{\omega_i}|$, is at most 1. Therefore, the radius ρ typically has a value between 0 and 1. For any $\rho \geq 1$, DRO-TV is equivalent to a traditional robust problem of minimizing the worst-case cost.

For a given $x \in \mathcal{X}$, it is well known that the optimal value of the worst-case expected problem of DRO-TV, i.e., $f(x)$ in (1), is equivalent to a convex combination of supremum and Conditional Value-at-Risk (CVaR) at level ρ taken with respect to the nominal distribution $\mathbf{q}$ [25]:

$$f(x) = \rho \sup[\mathbf{h}(x)] + (1-\rho)\mathrm{CVaR}_\rho[\mathbf{h}(x)],$$

where $\mathrm{CVaR}_\rho[\mathbf{h}(x)] := \inf_\eta \left\{ \eta + \frac{1}{1-\rho}\mathbb{E}\left[(\mathbf{h}(x)-\eta)_+\right] \right\}$ and $(\cdot)_+ = \max\{0, \cdot\}$. By convention, $\mathrm{CVaR}_0[\mathbf{h}(x)] = \mathbb{E}\left[\mathbf{h}(x)\right] = \sum_{\omega_i \in \Omega} q_{\omega_i} h_{\omega_i}(x)$ and $\mathrm{CVaR}_1[\mathbf{h}(x)] = \sup[\mathbf{h}(x)] = \sup_{\omega_i \in \Omega} h_{\omega_i}(x)$. For a fixed $\eta \in \mathbb{R}$ and a fixed $x \in \mathcal{X}$, we define the cumulative distribution function of $\mathbf{h}(x)$ as $\Psi(x, \eta) := \sum_{\{\omega_i : h_{\omega_i}(x) \le \eta\}} q_{\omega_i}$. Also, given $\zeta \in [0,1]$, let $\mathrm{VaR}_\zeta[\mathbf{h}(x)]$ denote the left-side ζ-quantile of $\mathbf{h}(x)$, which is referred to as the Value-at-Risk (VaR) of $\mathbf{h}(x)$ at level ζ: $\mathrm{VaR}_\zeta[\mathbf{h}(x)] := \inf\{\eta : \Psi(x,\eta) \ge \zeta\}$ [43]. Again by convention, $\mathrm{VaR}_0[\mathbf{h}(x)] = -\infty$ and $\mathrm{VaR}_1[\mathbf{h}(x)] = \sup[\mathbf{h}(x)]$.

[39] defines the following sets that partition the scenario set Ω (while these sets are defined in [39] for a given $x \in \mathcal{X}$, we present them below for a given optimal solution of DRO-TV, $x^* \in \mathcal{X}$):

- $\Omega_1(x^*) := \{\omega \in \Omega : h_\omega(x^*) < \mathrm{VaR}_\rho[\mathbf{h}(x^*)]\}$, i.e., the set of scenarios with costs strictly below $\mathrm{VaR}_\rho[\mathbf{h}(x^*)]$;
- $\Omega_2(x^*) := \{\omega \in \Omega : h_\omega(x^*) = \mathrm{VaR}_\rho[\mathbf{h}(x^*)]\}$, i.e., the set of scenarios with costs at $\mathrm{VaR}_\rho[\mathbf{h}(x^*)]$;
- $\Omega_3(x^*) := \{\omega \in \Omega : \mathrm{VaR}_\rho[\mathbf{h}(x^*)] < h_\omega(x^*) < \sup[\mathbf{h}(x^*)]\}$, i.e., the set of scenarios with costs strictly between $\mathrm{VaR}_\rho[\mathbf{h}(x^*)]$ and $\sup[\mathbf{h}(x^*)]$;
- $\Omega_4(x^*) := \{\omega \in \Omega : h_\omega(x^*) = \sup[\mathbf{h}(x^*)]\}$, i.e., the set of scenarios with highest costs at x^*.

At x^*, an optimal dual variable corresponding to the first constraint in (7) in the worst-case expected problem of DRO-TV is given by [39]

$$\lambda^*_{\mathrm{TV}} = \sup[\mathbf{h}(x^*)] - \mathrm{VaR}_\rho[\mathbf{h}(x^*)]. \tag{8}$$

Note that λ^*_{TV} corresponds to λ^* in DRO-W, the dual of constraint (3d). As before, we suppress the dependence of λ^*_{TV} on x^* for simplicity. When $\lambda^*_{\mathrm{TV}} = 0$, then $\sup[\mathbf{h}(x^*)] = \mathrm{VaR}_\rho[\mathbf{h}(x^*)]$ and $\Omega_2(x^*) = \Omega_4(x^*)$. In this case, we also have $\Omega_3(x^*) = \varnothing$. Therefore, when $\lambda^*_{\mathrm{TV}} = 0$, we only have the categories $\Omega_1(x^*)$ and $\Omega_4(x^*)$. Otherwise when $\lambda^*_{\mathrm{TV}} > 0$, we have all four possible categories: $\Omega_1(x^*)$, $\Omega_2(x^*)$, $\Omega_3(x^*)$, and $\Omega_4(x^*)$.

Theorems 3–5 below restate the easy-to-check conditions categorizing effective and ineffective scenarios in DRO-TV from [39], adapted to our setting.

Theorem 3 (Theorem 2 in [39] with $\mathbf{q} > 0$)**.** *Suppose $(x^*, \mathbf{p}^*)$ solves DRO-TV, and let λ^*_{TV} be defined as in (8). When $\lambda^*_{\mathrm{TV}} > 0$, scenario ω_j is **effective** if any the following conditions hold: (i)*

$q_{\omega_j} > \rho$, *(ii)* $\omega_j \in \Omega_3(x^*)$*, (iii)* $\Omega_2(x^*) = \{\omega_j\}$ *and* $p^*_{\omega_j} > 0$*, or (iv)* $\omega_j \in \Omega_4(x^*)$*. When* $\lambda^*_{\mathrm{TV}} = 0$*, scenario* ω_j *is effective if condition (i) or (v)* $\Omega_4(x^*) = \{\omega_j\}$ *holds.*

Theorem 4 (Theorem 1 in [39] with $\mathbf{q} > 0$)**.** *Consider the notation defined for DRO-TV in Theorem 3. Further, consider a scenario* ω_j *with* $q_{\omega_j} \leq \rho$*. When* $\lambda^*_{\mathrm{TV}} > 0$*, scenario* ω_j *is* ***ineffective*** *if any of the following conditions hold: (i)* $\omega_j \in \Omega_1(x^*)$ *or (ii)* $\omega_j \in \Omega_2(x^*)$ *and* $\sum_{\omega \in \Omega_2(x^*)} p^*_{\omega_j} = 0$*. When* $\lambda^*_{\mathrm{TV}} = 0$*, scenario* ω_j *is ineffective if condition (i) holds.*

Before presenting the next result, let us define some additional notation. For a set of scenarios $\mathcal{F} \subseteq \Omega$, recall $q(\mathcal{F}) = \sum_{\omega_i \in \mathcal{F}} q_{\omega_i}$ and now define $\rho_{\mathcal{F}} := \frac{\rho - q(\mathcal{F})}{1 - q(\mathcal{F})}$. Also, for fixed $\eta \in \mathbb{R}$ and fixed $x^* \in \mathcal{X}$, recall the cumulative distribution function of $\mathbf{h}(x^*)$, $\Psi(x^*, \eta) = \sum_{\{\omega_i : h_{\omega_i}(x^*) \leq \eta\}} q_{\omega_i}$, and now define $\psi_{|\mathcal{F}^c}(x^*, \eta)$ as the conditional version of $\Psi(x^*, \eta)$. That is, $\Psi_{|\mathcal{F}^c}(x^*, \eta) := \sum_{\omega_i \in \mathcal{F}^c \cap \{\omega : h_\omega(x^*) \leq \eta\}} q_{\omega_i|\mathcal{F}^c}$, where $q_{\omega_i|\mathcal{F}^c} = \frac{q_{\omega_i}}{1 - q(\mathcal{F})}$ is the probability of scenario ω_i conditioned on $\mathcal{F}^c$. Finally, let $\inf\{\eta : \Psi_{|\mathcal{F}^c}(x^*, \eta) \geq \rho_{\mathcal{F}}\}$ be the VaR of $\mathbf{h}(x^*)$ at level $0 \leq \rho_{\mathcal{F}} \leq 1$ conditioned on $\mathcal{F}^c$, denoted by $\mathrm{VaR}_{\rho_{\mathcal{F}}}[\mathbf{h}(x^*)|\mathcal{F}^c]$. Note that if $q(\mathcal{F}) \leq \rho \leq 1$, we have $0 \leq \rho_{\mathcal{F}} \leq 1$.

Theorem 5 (Theorem 3 in [39])**.** *Consider the notation defined for DRO-TV in Theorem 3. Suppose scenario* $\omega_j \in \Omega_2(x^*)$ *cannot be categorized by Theorem 3 or 4. Then, scenario* ω_j *is* ***effective*** *if both conditions hold: (i)* $\mathrm{VaR}_{\rho_{\mathcal{F}}}[\mathbf{h}(x^*)|\mathcal{F}^c] < \mathrm{VaR}_\rho[\mathbf{h}(x^*)]$ *and (ii) either there exists a scenario* ω_i *satisfying* $\mathrm{VaR}_{\rho_{\mathcal{F}}}[\mathbf{h}(x^*)|\mathcal{F}^c] < h_{\omega_i}(x^*) < \mathrm{VaR}_\rho[\mathbf{h}(x^*)]$ *or* $\Psi_{|\mathcal{F}^c}\big(x^*, \mathrm{VaR}_\rho[\mathbf{h}(x^*)]\big) > \rho_{\mathcal{F}}$*.*

5.2 Equivalence of Easy-to-Check Conditions

In this section, we show that the two sets of easy-to-check conditions—the ones proposed in Section 4.2 and the ones in [39] summarized in Section 5.1—are the same when DRO-W is constructed to be equivalent to DRO-TV by the appropriate transportation costs $c_{\omega_i \omega_j}$. We first discuss the case when $\Lambda^* = \{\lambda^*\}$ is a singleton.

Unique optimal λ^*. In this case, the unique λ^* of the equivalent DRO-W is equal to λ^*_{TV} defined in (8). In Lemma 7 below, we establish relationships between the two partitions of Ω under a unique λ^*. Recall that we assume $\rho > 0$.

Lemma 7. *Suppose the transportation costs in DRO-W are defined as* $c_{\omega_i \omega_j} = 1$ *if* $\omega_i \neq \omega_j$*, and* $c_{\omega_i \omega_j} = 0$ *otherwise. Let* x^* *be an optimal solution of DRO-W (or, DRO-TV), and let* $\lambda^* = \lambda^*_{\mathrm{TV}}$ *given in (8) be the unique optimal dual solution corresponding to (3d) (equivalently the first constraint in (7)) in the worst-case expected problem of DRO-W (or, DRO-TV). Recall the partitions of* Ω *defined for DRO-W* $\Big(\Omega_N(x^*, \lambda^*), \Omega_U(x^*, \lambda^*), \Omega_M(x^*, \lambda^*)\Big)$ *and for DRO-TV* $\Big(\Omega_1(x^*)\text{–}\Omega_4(x^*)\Big)$*.*

- *(i) When* $\lambda^* = 0$*,* $\Omega_1(x^*) = \Omega_N(x^*, \lambda^*)$*. Furthermore,* $\Omega_4(x^*)$ *is a singleton if and only if* $\Omega_4(x^*) = \Omega_U(x^*, \lambda^*)$*, and in this case* $\Omega_M(x^*, \lambda^*) = \varnothing$*. Otherwise, if* $\Omega_4(x^*)$ *is not a singleton, then* $\Omega_4(x^*) = \Omega_M(x^*, \lambda^*)$ *and* $\Omega_U(x^*, \lambda^*) = \varnothing$*.*

(ii) When $\lambda^* > 0$, $\Omega_1(x^*) = \Omega_N(x^*, \lambda^*)$, $\Omega_2(x^*) = \Omega_M(x^*, \lambda^*)$, $\Omega_3(x^*) \cup \Omega_4(x^*) = \Omega_U(x^*, \lambda^*)$. In addition, for a scenario $\omega_j \in \Omega_2(x^*) \cup \Omega_3(x^*)$, we have $\mathcal{M}(\omega_j) = \{\omega_j\}$, and for a scenario $\omega_j \in \Omega_4(x^*)$, we have $\mathcal{M}(\omega_j) = \Omega_1(x^*) \cup \Omega_2(x^*) \cup \{\omega_j\}$.

Proof. (i) When $\lambda^* = 0$, we have $\mathrm{VaR}_\rho[\mathbf{h}(x^*)] = \sup[\mathbf{h}(x^*)]$. Recall that in this case only $\Omega_1(x^*)$ and $\Omega_4(x^*)$ exist. For any scenario ω_i, $\mathcal{A}(\omega_i, \lambda^*)$ can only contain scenarios in $\Omega_4(x^*)$ because with $\lambda^* = 0$, $\mathcal{A}(\omega_i, \lambda^*) = \arg\max_\omega\{h_\omega(x^*)\}$. Therefore, $\Omega_1(x^*) = \Omega_N(x^*, \lambda^*)$. Now let us show if $\Omega_4(x^*) = \Omega_U(x^*, \lambda^*)$, then $\Omega_4(x^*)$ is a singleton; i.e., $\Omega_4(x^*) = \{\omega_j\}$ for some $\omega_j \in \Omega$. For the sake of contradiction, suppose $\Omega_4(x^*) = \Omega_U(x^*, \lambda^*)$ but $\Omega_4(x^*)$ contains more than one scenario. Then any scenario $\omega_k \in \Omega_4(x^*)$ with $\omega_k \neq \omega_j$ will also satisfy $\omega_k \in \mathcal{A}(\omega_i, \lambda^*)$ for all $\omega_i \in \Omega$. This means $\omega_k \in \Omega_M(x^*, \lambda^*)$, which is a contradiction to $\Omega_4(x^*) = \Omega_U(x^*, \lambda^*)$. This also means when $\Omega_4(x^*)$ is not a singleton, $\Omega_4(x^*) = \Omega_M(x^*, \lambda^*)$ and $\Omega_U(x^*, \lambda^*) = \varnothing$. On the other hand, if $\Omega_4(x^*) = \{\omega_j\}$, then $\{\omega_j\} = \arg\max_\omega\{h_\omega(x)\}$ and hence $\{\omega_j\} = \mathcal{A}(\omega_i, \lambda^*)$ for all $\omega_i \in \Omega$. Therefore, by definition, $\Omega_4(x^*) = \Omega_U(x^*, \lambda^*)$ and $\Omega_M(x^*, \lambda^*) = \varnothing$.

(ii) When $\lambda^* > 0$, $\lambda^* = \sup[\mathbf{h}(x^*)] - \mathrm{VaR}_\rho[\mathbf{h}(x^*)] > 0$. Plugging in this value of λ^* and using the fact that $c_{\omega_i\omega_j} = 1$ if $\omega_i \neq \omega_j$ and 0 otherwise, we obtain the following for $\mathcal{A}(\omega_i, \lambda^*) = \arg\max_\omega\{h_\omega(x^*) - \lambda^* c_{\omega_i\omega}\}$. If $\omega_i \in \Omega_1(x^*)$, then $\mathcal{A}(\omega_i, \lambda^*) = \Omega_4(x^*)$. If $\omega_i \in \Omega_2(x^*)$, then $\mathcal{A}(\omega_i, \lambda^*) = \Omega_4(x^*) \cup \{\omega_i\}$. If $\omega_i \in \Omega_3(x^*)$, then $\mathcal{A}(\omega_i, \lambda^*) = \{\omega_i\}$, and finally if $\omega_i \in \Omega_4(x^*)$, we again have $\mathcal{A}(\omega_i, \lambda^*) = \{\omega_i\}$. Therefore, based on these, we obtain $\Omega_N(x^*, \lambda^*) = \Omega_1(x^*)$, $\Omega_M(x^*, \lambda^*) = \Omega_2(x^*)$ and $\Omega_U(x^*, \lambda^*) = \Omega_3(x^*) \cup \Omega_4(x^*)$. The other statements can also be derived similarly form the above. □

We now formally present the equivalence of the two sets of easy-to-check conditions—Theorems 1–2 for DRO-W and Theorems 3–5 for DRO-TV—under a unique optimal λ^*.

Proposition 13. *Consider the DRO-W setting as described in Lemma 7. Then,*

(a) Theorem 3(i) is equivalent to Theorem 1(i);

(b) Theorem 3(ii), (iv), and (v) combined are equivalent to Theorem 1(ii);

(c) Theorem 4(i) is equivalent to Theorem 2(i);

(d) Theorem 4(ii) is equivalent to Theorem 2(ii);

(e) Theorem 3(iii) and Theorem 5 combined are equivalent to Theorem 1(iii).

Proof. Recall the notation $q(\mathcal{F}) := \sum_{\omega\in\mathcal{F}} q_\omega$ and $p^*(\mathcal{F}) := \sum_{\omega\in\mathcal{F}} p^*_\omega$ for a given subset $\mathcal{F} \subseteq \Omega$. We first show parts (a) to (d) by arguing that the conditions in their respective theorems are equivalent. Part (a) holds because $c_{\omega_i\omega_j} = 1$ if $\omega_i \neq \omega_j$ and $c_{\omega_i\omega_j} = 0$ if $\omega_i = \omega_j$. Now we show part (b) starting with $\lambda^* = 0$ case. When $\lambda^* = 0$ and $\Omega_4(x^*) = \{\omega_j\}$, we must have $\{\omega_j\} = \Omega_U(x^*, \lambda^*)$ by Lemma 7(i). On the other hand, when $\lambda^* > 0$, by Lemma 7(ii), we have $\Omega_3(x^*) \cup \Omega_4(x^*) = \Omega_U(x^*, \lambda^*)$. For parts (c) and (d), note that Theorem 4 considers scenarios ω_j with $q_{\omega_j} \leq \rho$. We have shown in part (a) that the complement of this condition ($q_{\omega_j} > \rho$) is equivalent to Theorem 1(i). Any scenario that satisfies the condition in 1(i) cannot satisfy the conditions in Theorem 2. This is because if $\omega_j \in \Omega_N(x^*, \lambda^*)$ or $\omega_j \in \Omega_M(x^*, \lambda^*)$ and $p^*(\Omega_M(x^*, \lambda^*)) = 0$, then moving all probability masses out of ω_j will violate the constraint (3d). Therefore, we automatically only consider scenarios with $q_{\omega_j} \leq \rho$ in Theorem 2. Then, part (c) is implied by Lemma 7 because $\Omega_1(x^*) = \Omega_N(x^*, \lambda^*)$ always holds regardless $\lambda^* = 0$ or $\lambda^* > 0$. For part (d), we only need to consider $\lambda^* > 0$. Since in this case $\Omega_2(x^*) = \Omega_M(x^*, \lambda^*)$ by Lemma 7, condition (ii) in Theorem 4 is exactly the same as the first condition of Theorem 2(ii). Also, for all $\omega_j \in \Omega_N(x^*, \lambda^*)$, $\mathcal{A}(\omega_j, \lambda^*)$ contains only the highest-cost scenarios, which must be in $\Omega_U(x^*, \lambda^*)$ by Lemma 4. For all $\omega_j \in \Omega_M(x^*, \lambda^*)$, $\mathcal{A}(\omega_j, \lambda^*)$ contains all the largest-cost scenarios and itself. Since $p^*(\Omega_M(x^*, \lambda^*)) = 0$, we have $\gamma^*_{\omega_i\omega_j} = 0$ for any $\omega_i \in \Omega_N(x^*, \lambda^*) \cup \Omega_M(x^*, \lambda^*)$ by constraint (3b). This means that at least one of the three conditions in 2(ii) will automatically be satisfied for every $\omega_i \in \Omega_N(x^*, \lambda^*) \cup \Omega_M(x^*, \lambda^*)$ in DRO-TV. This shows part (d).

Finally, let us show part (e). For this part, we only need to consider $\lambda^* > 0$ or $\lambda^* = 0$ and $\Omega_4(x^*)$ is not a singleton. Otherwise, when $\lambda^* = 0$ and $\Omega_4(x^*) = \{\omega_j\}$ is a singleton, by Lemma 7(i), $\Omega_4(x^*) = \Omega_U(x^*, \lambda^*)$. Recall $\Omega_2(x^*) = \Omega_4(x^*)$ when $\lambda^* = 0$. In this case, Theorem 3(v) and Theorem 1(ii) identify ω_j as an effective scenario, while Theorem 4(i) and Theorem 2(i) identify all scenarios except ω_j as ineffective scenarios. Now let us consider the case when $\lambda^* > 0$ or the case when $\lambda^* = 0$ and $\Omega_4(x^*)$ is not a singleton. By Lemma 7, we have $\Omega_2(x^*) = \Omega_M(x^*, \lambda^*)$ when $\lambda^* > 0$ and $\Omega_2(x^*) = \Omega_4(x^*) = \Omega_M(x^*, \lambda^*)$ when $\lambda^* = 0$. Also, from Lemma 7, the transfer of probability masses from $\Omega_1(x^*)$ to $\Omega_4(x^*)$ incurs positive transportation costs. Because $c_{\omega_i\omega_j} = 1$ if $\omega_i \neq \omega_j$, this transportation cost collectively equals $q(\Omega_1(x^*))$. For the rest of the proof, let us denote $\mathcal{F} = \{\omega_j\} \subseteq \Omega_2(x^*) = \Omega_M(x^*, \lambda^*)$. We now show the conditions in Theorem 5 and Theorem 1(iii) are identical by arguing that they imply each other. Suppose the condition in Theorem 1(iii) holds. This condition is equivalent to $\rho - q(\Omega_1(x^*)) < q_{\omega_j}$. This means $\frac{q(\Omega_1(x^*))}{1-q_{\omega_j}} > \frac{\rho - q_{\omega_j}}{1-q_{\omega_j}} = \rho_{\mathcal{F}}$. Thus, not all scenarios $\omega_i \in \Omega_1(x^*)$ have costs $h_{\omega_i}(x^*)$ below $\mathrm{VaR}_{\rho_{\mathcal{F}}}[\mathbf{h}(x^*)|\mathcal{F}^c]$ in the assessment problem (cf. Proposition 5 in [39]). Therefore, by the definition of $\Omega_1(x^*)$ we have $\mathrm{VaR}_{\rho_{\mathcal{F}}}[\mathbf{h}(x^*)|\mathcal{F}^c] < \mathrm{VaR}_{\rho}[\mathbf{h}(x^*)]$. Besides, if there does not exist a scenario ω_i such that $\mathrm{VaR}_{\rho_{\mathcal{F}}}[\mathbf{h}(x^*)|\mathcal{F}^c] < h_{\omega_i}(x^*) < \mathrm{VaR}_{\rho}[\mathbf{h}(x^*)]$ with positive nominal probability $q_{\omega_i} > 0$, we have $\Psi_{|\mathcal{F}^c}(x^*, \mathrm{VaR}_{\rho}[\mathbf{h}(x^*)]) = \frac{q(\Omega_1(x^*))}{1-q_{\omega_j}} > \frac{\rho - q_{\omega_j}}{1-q_{\omega_j}} = \rho_{\mathcal{F}}$. This shows that Theorem 1(iii) implies Theorem 5. Now, suppose the conditions in Theorem 5 hold. When Theorem 5(i) holds and there is not a scenario ω_i satisfying $\mathrm{VaR}_{\rho_{\mathcal{F}}}[\mathbf{h}(x^*)|\mathcal{F}^c] < h_{\omega_i}(x^*) < \mathrm{VaR}_{\rho}[\mathbf{h}(x^*)]$ with positive nominal probability $q_{\omega_i} > 0$, then $\Psi_{|\mathcal{F}^c}(x^*, \mathrm{VaR}_{\rho}[\mathbf{h}(x^*)]) = \frac{q(\Omega_1(x^*))}{1-q_{\omega_j}}$. Therefore, $\Psi_{|\mathcal{F}^c}(x^*, \mathrm{VaR}_{\rho}[\mathbf{h}(x^*)]) > \rho_{\mathcal{F}}$ means $\frac{q(\Omega_1(x^*))}{1-q_{\omega_j}} > \frac{\rho - q_{\omega_j}}{1-q_{\omega_j}} = \rho_{\mathcal{F}}$ and hence $\rho - q(\Omega_1(x^*)) < q_{\omega_j}$, which is the condition in Theorem 1(iii). Alternatively, when Theorem 5(i) holds and there exists a scenario ω_i satisfying $\mathrm{VaR}_{\rho_{\mathcal{F}}}[\mathbf{h}(x)|\mathcal{F}^c] < h_{\omega_i}(x^*) < \mathrm{VaR}_{\rho}[\mathbf{h}(x^*)]$ with $q_{\omega_i} > 0$, not all scenarios in $\Omega_1(x^*)$ are below $\mathrm{VaR}_{\rho_{\mathcal{F}}}[\mathbf{h}(x)|\mathcal{F}^c]$ in the assessment problem. Thus we have $\rho_{\mathcal{F}} = \frac{\rho - q_{\omega_j}}{1-q_{\omega_j}} < \frac{q(\Omega_1(x^*))}{1-q_{\omega_j}}$, which also implies $\rho - q(\Omega_1(x^*)) < q_{\omega_j}$.

Finally, as a special case, when $\Omega_2(x^*)$ is a singleton and $\lambda^* > 0$, we show that Theorem 3(iii) and Theorem 1(iii) are identical by arguing that they imply each other. Suppose Theorem 3(iii) holds; that is, $\{\omega_j\} = \Omega_2(x^*)$ and $p^*_{\omega_j} > 0$. Then, we must have $\rho - q(\Omega_1(x^*)) < q_{\omega_j}$. Otherwise, any solution to the worst-case expected problem, $\mathbf{p}^1$, satisfying $p^1_{\omega_j} = 0$, $p^1(\Omega_4(x^*)) = p^1(\Omega_4(x^*)) + p^*_{\omega_j}$, and $p^1_\omega = p^*_\omega$ for the remaining scenarios $\omega \in \Omega \backslash (\Omega_4(x^*) \cup \{\omega_j\})$ will be feasible and will give a better objective function value, which is a contradiction to $\mathbf{p}^*$ being the optimal worst-case distribution. Now suppose Theorem 1(iii) holds with $\{\omega_j\} = \Omega_M(x^*, \lambda^*)$ and $\rho - q(\Omega_1(x^*)) < q_{\omega_j}$. Then by [39, Proposition 4], we have $p^*_{\omega_j} > 0$. □

The proof of Proposition 13 reveals that a scenario ω_j satisfying the conditions of Theorem 3(iii) will also satisfy the conditions of Theorem 5, but the reverse is not necessarily true. In DRO-TV, Theorem 5 is invoked to identify the effectiveness of *additional* scenarios that are not captured by Theorem 3(iii), whereas in DRO-W, Theorem 1(iii) identifies *all* such scenarios. That is why condition (e) of Proposition 13 lists both Theorem 3(iii) and Theorem 5 jointly.

Multiple optimal λ^*. Let us now discuss when there are multiple optimal λ^*, i.e., $\Lambda^* = [\lambda^*_{\min}, \lambda^*_{\max}]$ with $0 \le \lambda^*_{\min} < \lambda^*_{\max}$. Then, the scenarios belonging to sets $\Omega_N(x^*, \lambda^*)$, $\Omega_U(x^*, \lambda^*)$ and $\Omega_M(x^*, \lambda^*)$ can change sets at different values of λ^*, but the scenarios belonging to $\Omega_1(x^*)$–$\Omega_4(x^*)$ remain consistent. Therefore, Lemma 7 and Proposition 13 do not hold for all $\lambda^* \in \Lambda^*$. That said, the two sets of easy-to-check conditions are still equivalent. The key to this equivalence is that, when there are multiple optimal λ^*, the specific λ^*_{TV} given in (8) is in fact the largest possible λ^* at x^*; that is, $\lambda^*_{\text{TV}} = \lambda^*_{\max}$. We formalize these below.

Proposition 14. *Suppose the transportation costs in DRO-W are defined as $c_{\omega_i\omega_j} = 1$ if $\omega_i \neq \omega_j$, and $c_{\omega_i\omega_j} = 0$ otherwise. Let x^* be an optimal solution of DRO-W (or, DRO-TV). Let $\Lambda^* = [\lambda^*_{\min}, \lambda^*_{\max}]$ with $0 \le \lambda^*_{\min} < \lambda^*_{\max}$ be the set of optimal dual solutions corresponding to (3d) (equivalently the first constraint in (7)) in the worst-case expected problem of DRO-W (or, DRO-TV), and let λ^*_{TV} be defined as in (8). Recall the partitions of Ω defined for DRO-W $\Big(\Omega_N(x^*, \lambda^*)$, $\Omega_U(x^*, \lambda^*)$, $\Omega_M(x^*, \lambda^*)\Big)$ and for DRO-TV $\Big(\Omega_1(x^*)$–$\Omega_4(x^*)\Big)$. Then,*

(a) $p^*(\Omega_2(x^*)) = 0$ and $\lambda^*_{\max} = \lambda^*_{\text{TV}} = \sup[\mathbf{h}(x^*)] - \text{VaR}_\rho[\mathbf{h}(x^*)]$;

(b) Both Theorems 1–2 and Theorems 3–5 identify the effectiveness of all scenarios.

Proof. (a) We first show that for any $\hat{\lambda} \in \Lambda^*$, we have $\hat{\lambda} \le \sup[\mathbf{h}(x^*)] - \text{VaR}_\rho[\mathbf{h}(x^*)]$. Recall by definition of $\text{VaR}_\rho[\mathbf{h}(x^*)]$, we have $\rho \le \sum_{\omega \in \Omega_1(x^*) \cup \Omega_2(x^*)} q_\omega$. For the sake of contradiction, let $\hat{\lambda} \in \Lambda^*$ but suppose $\hat{\lambda} > \sup[\mathbf{h}(x^*)] - \text{VaR}_\rho[\mathbf{h}(x^*)]$. Then, we have $\Omega_U(x^*, \hat{\lambda}) \supseteq \Omega_2(x^*) \cup \Omega_3(x^*) \cup \Omega_4(x^*)$. So, by Lemma 3 and Lemma 1, we must have $\gamma^*_{\omega_j\omega_j} = q_{\omega_j}$ for any $\omega_j \in \Omega_2(x^*) \cup \Omega_3(x^*) \cup \Omega_4(x^*)$. This means the left-hand side of constraint (3d) is at most $\sum_{\omega_j \in \Omega_1(x^*)} q_{\omega_j} < \rho$. This in turn means constraint (3d) is not binding and by complementary slackness, $\hat{\lambda} = 0$, which is a contradiction. Therefore, for any $\lambda^* \in \Lambda^*$, $\lambda^* \leq \sup[\mathbf{h}(x^*)] - \mathrm{VaR}_\rho[\mathbf{h}(x^*)]$ must hold.

We now show $\rho = \sum_{\omega \in \Omega_1(x^*) \cup \Omega_2(x^*)} q_\omega$ when there are multiple optimal λ^*, i.e., when $\Lambda^* = [\lambda^*_{\min}, \lambda^*_{\max}]$ with $\lambda^*_{\min} < \lambda^*_{\max}$. By [39, Proposition 4], this is equivalent to $p^*(\Omega_2(x^*)) = 0$. Suppose instead $\rho < \sum_{\omega \in \Omega_1(x^*) \cup \Omega_2(x^*)} q_\omega$. Then, we must have $\gamma^*_{\omega_j \omega_j} > 0$ for some $\omega_j \in \Omega_2(x^*)$. If not, the total transportation cost spent will be more than $\sum_{\omega \in \Omega_1(x^*) \cup \Omega_2(x^*)} q_\omega$, which violates constraint (3d). Note that the sets $\Omega_1(x^*)$ and $\Omega_2(x^*)$ are not affected by the particular value of λ^*. So, this must hold at all values of $\lambda^* \in \Lambda^*$. Now, pick an arbitrary $\lambda^* \in \Lambda^*$. By Lemma 1, we must have $\omega_j \in \mathcal{A}(\omega_j, \lambda^*)$ at that $\lambda^* \in \Lambda^*$. This means $\lambda^* = \sup[\mathbf{h}(x^*)] - \mathrm{VaR}_\rho[\mathbf{h}(x^*)]$. Otherwise, $h_{\omega_j}(x^*) < h_{\omega_k}(x^*) - \lambda^* c_{\omega_j \omega_k}$ for some $\omega_k \in \Omega_4(x^*)$ and so ω_j cannot be in $\mathcal{A}(\omega_j, \lambda^*)$. So, our arbitrarily chosen $\lambda^* \in \Lambda^*$ uniquely equals $\sup[\mathbf{h}(x^*)] - \mathrm{VaR}_\rho[\mathbf{h}(x^*)]$, which is a contradiction to $\lambda^*_{\min} < \lambda^*_{\max}$ and a contradiction to this holding at all $\lambda^* \in \Lambda^*$.

Since $\lambda^* \leq \sup[\mathbf{h}(x^*)] - \mathrm{VaR}_\rho[\mathbf{h}(x^*)]$, we only need to show $\sup[\mathbf{h}(x^*)] - \mathrm{VaR}_\rho[\mathbf{h}(x^*)] \in [\lambda^*_{\min}, \lambda^*_{\max}]$. Let $\bar{\lambda} = \sup[\mathbf{h}(x^*)] - \mathrm{VaR}_\rho[\mathbf{h}(x^*)]$. Then, $\Omega_N(x^*, \bar{\lambda}) = \Omega_1(x^*)$, $\Omega_M(x^*, \bar{\lambda}) = \Omega_2(x^*)$, and $\Omega_U(x^*, \bar{\lambda}) = \Omega_3(x^*) \cup \Omega_4(x^*)$. Let ω_k be any scenario in $\Omega_4(x^*)$. The following characterizes a primal and dual feasible solution to (3) at $\bar{\lambda}$ with the same objective function value; hence showing $\bar{\lambda}$ is an optimal dual solution. For all $\omega_j \in \Omega_1(x^*)$, $p^*_{\omega_j} = 0$, $\gamma^*_{\omega_j \omega_k} = q_{\omega_j}$, $\gamma^*_{\omega_j \omega} = 0$ for any $\omega \neq \omega_k$, $\gamma^*_{\omega \omega_j} = 0$ for all $\omega \in \Omega$, $\alpha_{\omega_j} = -h_{\omega_j}(x^*)$, and $\beta_{\omega_j} = h_{\omega_k}(x^*) - \bar{\lambda}$. For all $\omega_j \in \Omega_2(x^*)$, $p^*_{\omega_j} = 0$, $\gamma^*_{\omega_j \omega_k} = q_{\omega_j}$, $\gamma^*_{\omega_j \omega} = 0$ for any $\omega \neq \omega_k$, $\gamma^*_{\omega \omega_j} = 0$ for all $\omega \in \Omega$, $\alpha_{\omega_j} = -h_{\omega_j}(x^*)$, and $\beta_{\omega_j} = h_{\omega_k}(x^*) - \bar{\lambda}$. For all $\omega_j \in \Omega_3(x^*)$, $p^*_{\omega_j} = q_{\omega_j}$, $\gamma^*_{\omega_j \omega_j} = q_{\omega_j}$, $\gamma^*_{\omega_j \omega} = 0$ for any $\omega \neq \omega_j$, $\gamma^*_{\omega \omega_j} = 0$ for any $\omega \neq \omega_j$, $\alpha_{\omega_j} = -h_{\omega_j}(x^*)$, and $\beta_{\omega_j} = h_{\omega_j}(x^*)$. For all $\omega_j \in \Omega_4(x^*) \setminus \{\omega_k\}$, $p^*_{\omega_j} = q_{\omega_j}$, $\gamma^*_{\omega_j \omega_j} = q_{\omega_j}$, $\gamma^*_{\omega_j \omega} = 0$ for any $\omega \neq \omega_j$, $\gamma^*_{\omega \omega_j} = 0$ for any $\omega \neq \omega_j$, $\alpha_{\omega_j} = -h_{\omega_j}(x^*)$, and $\beta_{\omega_j} = h_{\omega_j}(x^*)$. Finally, $p^*_{\omega_k} = q_{\omega_k} + \sum_{\omega \in \Omega_1(x^*) \cup \Omega_2(x^*)} q_\omega = q_{\omega_k} + \rho$, $\gamma^*_{\omega_k \omega_k} = q_{\omega_k}$, $\gamma^*_{\omega_k \omega} = 0$ for any $\omega \neq \omega_k$, $\gamma^*_{\omega \omega_k} = 0$ for any $\omega \neq \omega_k$, $\alpha_{\omega_k} = -h_{\omega_k}(x^*)$, and $\beta_{\omega_k} = h_{\omega_k}(x^*)$. Given $\sum_{\omega \in \Omega_1(x^*) \cup \Omega_2(x^*)} q_\omega = \rho$, the objective function value of both primal and dual problems equal $\sum_{\omega_j \in \Omega_3(x^*)} h_{\omega_j}(x^*) + (\rho + q(\Omega_4(x^*)) h_{\omega_k}(x^*)$. This means $\bar{\lambda} = \lambda^*_{\max}$.

(b) By (a), using a version of Lemma 7 and Proposition 13 at $\lambda^*_{\max} = \sup[\mathbf{h}(x^*)] - \mathrm{VaR}_\rho[\mathbf{h}(x^*)]$, we can show the equivalence of Theorems 1–2 and 3–5 at $\lambda^*_{\max}$. Note that not all conditions will be activated at $\lambda^*_{\max}$. By the results in Section 4.4, the effectiveness of any scenario identified by Theorems 1–2 at $\lambda^*_{\max}$ will be the same at all $\lambda^* \in \Lambda^*$. Corollary 1 then concludes the proof. □

Proposition 14 shows that when there are multiple optimal λ^* in DRO-TV, the effectiveness of all scenarios are identified and $p^*(\Omega_2(x^*)) = 0$, further refining the results in [39] in this special case. Specifically, under multiple optimal λ^* and assuming $q_{\omega_j} \leq \rho$ for all ω_j, Theorem 4(i)–(ii) identify the scenarios in $\Omega_1(x^*)$ and $\Omega_2(x^*)$ as ineffective and Theorem 3(ii) and (iv) identify the scenarios in $\Omega_3(x^*)$ and $\Omega_4(x^*)$ as effective. These conditions correspond to Theorem 2(i)–(ii) and Theorem 1(ii) at $\lambda^*_{\max} = \lambda^*_{\mathrm{TV}}$ in DRO-W. At different $\lambda^* \in \Lambda^*$, alternate conditions in Theorems 1–2 can be activated, but in all cases, the effectiveness of all scenarios will be identified in the multiple optima case.

5.3 Results in DRO-TV that do not hold in DRO-W

Under DRO-TV, Theorem 5 of [39] shows that when we observe an effective scenario with zero worst-case probability (i.e., $p^*_\omega = 0$) or an ineffective scenario with positive worst-case probability (i.e., $p^*_\omega > 0$), there exist multiple optimal solutions to the worst-case expected problem at x^* and that there will be at least two scenarios ω_i and $\omega_j \neq \omega_i$ having the same cost, i.e., $h_{\omega_i}(x^*) = h_{\omega_j}(x^*)$. While the existence of multiple optima is generally true (shown in Proposition 3 in Section 3), having at least two scenarios with the same cost is not. In the following, we provide a counterexample showing this part of Theorem 5 of [39] does not hold for DRO-W.

Example 1 (Part of Theorem 5 in [39] does not hold for DRO-W)**.** *Consider the worst-case expected problem of DRO-W with* $|\Omega| = 3$ *scenarios, where the random parameter corresponding to each scenario* ω_i *takes the values* $\xi_{\omega_i} = i$ *for* $i = 1, 2, 3$*. Suppose the transportation cost is given by* $c_{\omega_i\omega_j} = |\xi_{\omega_i} - \xi_{\omega_j}|$*, the nominal probabilities by* $\mathbf{q} = (0.05\ 0.05\ 0.9)$*, and the transportation budget is set to* $\rho = 0.05$*. Further, suppose the cost functions are defined as* $h_{\omega_i}(x) = \xi_{\omega_i}$ *for all* $\omega_i \in \Omega$ *and* $\mathcal{X} = \{x : 0 \leq x \leq 1\}$*. Then, an optimal worst-case probability distribution is given by* $\mathbf{p}^{(*,0)} = (0.025\ 0.05\ 0.925)$*, and all solutions* $x \in \mathcal{X}$ *are optimal. Scenario* ω_2 *has a positive probability in* $\mathbf{p}^{(*,0)}$*. However, by setting* $p_{\omega_2} = 0$ *and solving the assessment problem for* ω_2*, one can show that scenario* ω_2 *is ineffective. In fact, there exists an alternative worst-case distribution* $\mathbf{p}^{(*,1)} = (0.05\ 0\ 0.95)$*. However, we do not have any scenarios with the same costs* $h_{\omega_i}(x^*)$*.*

6 Numerical Results

We now present our computational experiments investigating the performance of easy-to-check conditions, comparing DRO-W with DRO-TV, and seeking further insights. We begin by describing the test problems and our computational setup.

6.1 Experimental Setup

To conduct computational experiments, we chose two problems from the literature, denoted 20TERM and APL1P. 20TERM, described in [32], models a motor freight carrier's operations. The first-stage determines the positions of a fleet of vehicles at the beginning of a day, and the second-stage decides the movement of the fleet through a network to satisfy point-to-point demands for shipments, with penalties for unsatisfied demands. The second-stage decisions also need to end the day with a fleet configuration matching the first-stage decisions. The problem has 40 stochastic parameters and 1.1×10^{12} scenarios. Because 20TERM has prohibitively many scenarios to be solved exactly, we sampled 1000 scenarios via Latin Hypercube sampling method to generate a test case. ALP1P, described in [24] and also used in [39], is a power capacity expansion problem. The first-stage allocates capacities that must be installed on the generators, while the second stage decides on the amount of additional capacities purchased to fulfill the unmet demands. The problem has 5 stochastic parameters that signify demands and generator reliabilities and 1280 scenarios in total. We chose APL1P to be able to compare the results with [39].

In our experiments, we utilized the L_2-norm to quantify the transportation costs, i.e., $c_{\omega_i\omega_j} = ||\xi_{\omega_i} - \xi_{\omega_j}||_2$ in DRO-W. It is important to emphasize that in the context of APL1P, where both demand and generator availability are subject to uncertainty, we adopted a "scaled" L_2-norm. Because demand values are three orders of magnitude larger than generator availabilities for ALP1P, we scaled the generator availabilities by 1000. This ensured that both the demand and generator availability play an equally significant role in determining the transportation cost.

Both problems were solved on a 64-bit PC with Intel Xeon 2.60 GHz processor and 128GB RAM. A variant of Benders' decomposition was used to solve the DROs; we refer to Section A-2 of the Appendix in [39] for details. We implemented the decomposition algorithm via C++ (Visual Studio 2017) and CPLEX 12.8, using SUTIL [15] to read in the problems and generate samples. The optimality threshold in the algorithm was set to 10^{-9}. We also used the same tolerance, 10^{-9}, for our conditions to check the effectiveness of scenarios. For instance, any value less then 10^{-9} is treated as zero. Also, if two values have an absolute difference less than 10^{-9}, they are treated as being equal.

6.2 Performance of Easy-to-Check Conditions

We first tested the performance of the proposed conditions in identifying the effectiveness of scenarios in DRO-W as the transportation budget ρ varies. The smallest value of ρ is set to zero, which essentially turns DRO-W into a regular stochastic program that minimizes the expected cost with respect to the nominal distribution. The largest value of ρ for each problem is obtained so that at least 98% of scenarios become effective. Then, for each problem, 20 different values of ρ are used (plus $\rho = 0$). Tables 2 and 3 present the number of scenarios in each category $\Omega_{\mathfrak{c}}(x^*, \lambda^*)$ for $\mathfrak{c} = N, M, U$ as well as the number of effective and ineffective scenarios being identified by our conditions for APL1P and 20TERM instances, respectively. In all cases reported in Tables 2 and 3, we did not observe any multiple optimal λ^* at the numerically obtained optimal solutions x^*.

The results reveal the following. First, the conditions successfully identify the effectiveness of a large fraction of scenarios. For APL1P, only a small portion of the total number of scenarios remain undetermined, and this portion can be higher for medium values of the transportation budget ρ. At small values of ρ (e.g., $\rho = 60, 120$), scenarios in $\Omega_M(x^*, \lambda^*)$ are classified as effective by Theorem 1(iii). At $\rho = 540$ in Table 2, scenarios in $\Omega_M(x^*, \lambda^*)$ are classified as ineffective by Theorem 2(ii). Classifying the scenarios in $\Omega_M(x^*, \lambda^*)$ appears to be more difficult for medium values of ρ for APL1P. For 20TERM, there are no undetermined scenarios except when $\rho = 45$. At this value of ρ, the problem is equivalent to minimizing the highest cost, and there are only two scenarios that have the highest cost. The easy-to-check conditions cannot differentiate between these two highest-cost scenarios without having to solve assessment problems. Otherwise, the conditions work very well for 20TERM. We will further examine the differences between these two problems in Section 6.4.

Second, as the transportation budget (or, level of robustness) increases, the total number of scenarios in $\Omega_U(x^*, \lambda^*)$ (consequently, the number of effective scenarios) decreases and the total number of scenarios in $\Omega_N(x^*, \lambda^*)$ (consequently, the number of ineffective scenarios) increases. The problem becomes more robust and focuses more on the expensive scenarios as the budget ρ increases. In addition, an optimal dual variable λ^* obtained by the numerical method decreases. This dual value quantifies the incremental improvement in the worst-case expectation for an additional unit of budget (see (3d)), which tends to go down when there is a large budget. Consequently, a more expensive scenario ω_j becomes more likely to achieve $\omega_j \in \arg\max_\omega \{h_\omega(x^*) - \lambda^* c_{\omega_i \omega}\}$ for any ω_i. An extreme case of this observation is when $\lambda^* = 0$, only the most expensive scenarios can be in the set $\mathcal{A}(\omega_i, \lambda^*) = \arg\max_\omega \{h_\omega(x^*) - \lambda^* c_{\omega_i \omega}\}$ for any ω_i (last row of Table 3).

Table 2: Number of scenarios in each category $\Omega_\mathfrak{c}(x^*, \lambda^*)$, $\mathfrak{c} = N, M, U$ as well as number of ineffective, effective, and undetermined scenarios in DRO-W using APL1P.

	# of scenarios			# of scenarios		
ρ	$\Omega_U(x^*, \lambda^*)$	$\Omega_N(x^*, \lambda^*)$	$\Omega_M(x^*, \lambda^*)$	Effective	Ineffective	Undetermined
0	1276	0	4	1280	0	0
60	1221	58	1	1222	58	0
120	1156	122	2	1158	122	0
160	1116	164	0	1116	164	0
200	1082	190	8	1082	190	8
240	1062	215	3	1063	215	3
280	1022	256	2	1022	256	2
320	964	314	2	966	314	0
360	832	444	4	834	444	2
380	797	480	3	798	480	2
400	758	580	2	758	580	2
420	714	553	13	714	553	13
440	664	604	12	664	604	12
460	568	712	0	568	712	0
480	511	767	2	511	767	2
500	400	878	2	400	878	2
520	354	926	0	354	926	0
540	211	975	94	211	1069	0
560	182	1098	0	182	1098	0
580	81	1194	5	81	1194	5
600	20	1260	0	20	1260	0

6.3 Effectiveness versus Cost of Scenarios

We now analyze and compare the effectiveness of scenarios in DRO-W and DRO-TV by focusing on the scenario costs. In DRO-TV, we use $\rho = 0.8$ for both problems and in DRO-W, we use $\rho = 510$ for APL1P and $\rho = 35$ for 20TERM. These values of ρ are chosen so that approximately 20% of scenarios are effective in each case.

Figure 2 shows the effectiveness of scenarios categorized by our conditions as effective (1 in the y-axis), ineffective (0 in the y-axis) or undetermined (2 in the y-axis) with respect to their sorted costs (x-axis). Figures 2(a) and 2(b) depict the results for DRO-TV, while Figures 2(c) and 2(d) illustrate them for DRO-W. The left graphs ((a) and (c)) show the results for APL1P, and the

Table 3: Number of ineffective, effective, and undetermined scenarios in DRO-W using 20TERM.

	# of scenarios			# of scenarios		
ρ	$\Omega_U(x^*,\lambda^*)$	$\Omega_N(x^*,\lambda^*)$	$\Omega_M(x^*,\lambda^*)$	Effective	Ineffective	Undetermined
0	999	0	1	1000	0	0
2	954	45	1	955	45	0
4	909	90	1	910	90	0
6	867	132	1	868	132	0
8	824	175	1	825	175	0
10	780	219	1	781	219	0
12	735	264	1	736	264	0
14	692	307	1	693	307	0
16	647	352	1	648	352	0
18	602	397	1	603	397	0
20	558	441	1	559	441	0
22.5	502	497	1	503	497	0
25	446	553	1	447	553	0
27.5	390	609	1	391	609	0
30	334	665	1	335	665	0
32.5	276	723	1	277	723	0
35	219	780	1	220	780	0
37.5	157	841	2	159	841	0
40	91	907	1	92	907	0
42.5	27	973	0	27	973	0
45	0	998	2	0	998	2

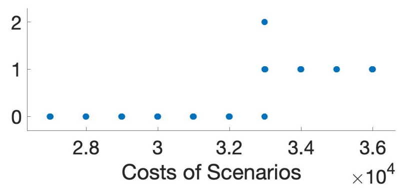
right graphs ((b) and (d)) for 20TERM. Note that the effectiveness of the undetermined scenarios can be checked by solving the assessment problems. However, we label them as undetermined in Figure 2 to illustrate the performance of our conditions.

In DRO-TV, effectiveness of scenarios is directly related to their costs. All scenarios with costs below $\text{VaR}_\rho[\mathbf{h}(x^*)]$ are ineffective and almost all scenarios with costs above this threshold are effective. Scenarios whose costs equal $\text{VaR}_\rho[\mathbf{h}(x^*)]$ can be either effective or ineffective, and some of them cannot be categorized by our conditions; see Figures 2(a) and 2(b). The same is true for scenarios whose costs are equal to $\sup[\mathbf{h}(x^*)]$ under some cases (when $\lambda^*_{\text{TV}} = 0$ and $\Omega_4(x^*)$ is *not* a singleton; see Theorem 3(v)). While such a case is not present in Figure 2, see the last row of Table 3 and note that at $\rho = 45$, DRO-W is equivalent to DRO-TV with $\rho \geq 1$.

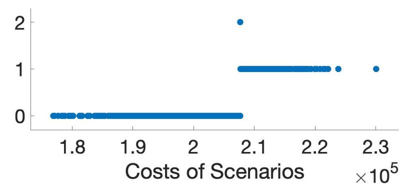
DRO-W has a different behavior than DRO-TV. Cost alone cannot determine the effectiveness of a scenario in DRO-W. However, it can provide some indication. For example, in Figure 2(d), scenarios of 20TERM with low costs tend to be ineffective and scenarios with high costs tend to be effective. However, scenarios whose costs are in the middle (2–2.1×10^5) can either be effective or ineffective. This is more pronounced in APL1P, where only the very low-cost scenarios are ineffective and only the very high-cost scenarios are effective. Figure 2(c) shows there is a relatively large region where the cost of a scenario cannot indicate its effectiveness. We will examine this in more detail in the next section.

We end this section with a remark that in Figure 2(a), costs of scenarios are concentrated around specific points at this optimal solution x^*. However, this does not mean the costs within a small neighborhood are the same. Their costs still exhibit some variation, with the differences between them not significant enough to be visually seen in the plot. For example, around the cost of

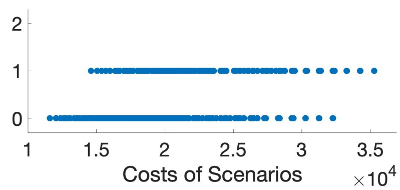
3.3×10^4, there are scenarios labeled as effective, ineffective, and undetermined. These scenarios do not necessarily have the same cost. For instance, the ones labeled as effective are more expensive than the ones labeled undetermined (among which the costs are the same since they are all at $\text{VaR}_\rho[\mathbf{h}(x^*)]$).



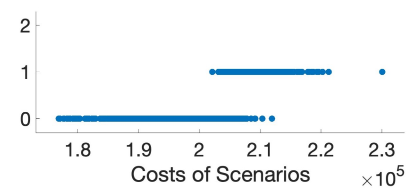
(a) APL1P, DRO-TV, $\rho = 0.8$.



(b) 20TERM, DRO-TV, $\rho = 0.8$.



(c) APL1P, DRO-W, $\rho = 510$.



(d) 20TERM, DRO-W, $\rho = 35$.

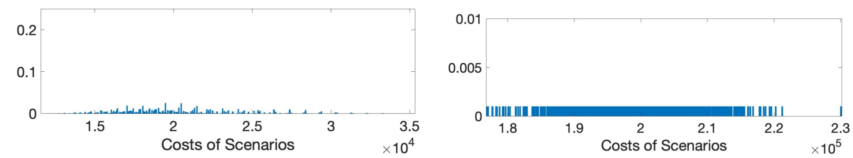
Figure 2: Sorted costs of scenarios (x-axis) versus categorized effectiveness of scenarios (y-axis). The y-axis in each plot stands for 0: Ineffective, 1: Effective, 2: Undetermined. Left figures show results for APL1P and right figures for 1000-sample 20TERM.

6.4 Characteristics of Effective Scenarios and Worst-Case Distributions

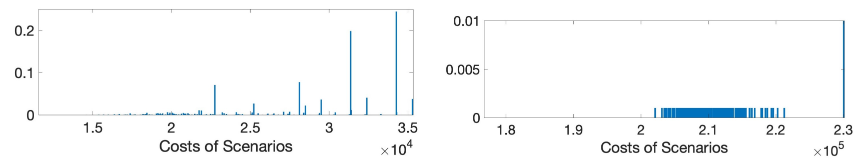
We now dive deeper into DRO-W's varying behavior in APL1P versus 20TERM. We keep the same experimental setup as in Section 6.3 and first investigate the worst-case distributions. Figures 3(a) and 3(b) (top row) depict the nominal distributions, and Figures 3(c) and 3(d) (middle row) illustrate the worst-case distributions for the two problems. Observe that the y-axis of Figure 3(d) is scaled to show the worst-case probability of other scenarios, but the worst-case probability of the highest-cost scenario is 0.8. The last row of Figure 3 repeats the last row of Figure 2 for an easy comparison. Left figures show the results for APL1P, the right figures for 20TERM, and the x-axes of all graphs provide the sorted costs of scenarios.

By directing our attention to the worst-case distribution (middle row), we notice that in APL1P, a small number of scenarios experience a substantial inflow of probability mass. In contrast, in 20TERM, only the scenario with the highest cost encounters a probability mass inflow. When examining effective and ineffective scenarios, we note that in APL1P, the correlation between a scenario's effectiveness and its cost is considerably weaker compared to 20TERM.

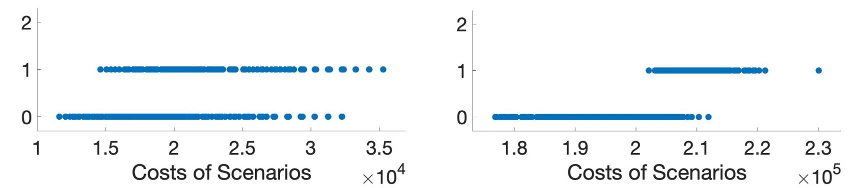
To gain insight into this behavior, let us examine the transportation costs between distinct scenarios $c_{\omega_i \omega_j}$ whenever $\omega_i \neq \omega_j$ (recall $c_{\omega_i \omega_i}$ is always zero for all ω_i). Figure 4 depicts the distribution of these transportation costs, which reveals that the variation in these costs to be considerably lower



(a) APL1P, nominal distribution. (b) 20TERM, nominal distribution.

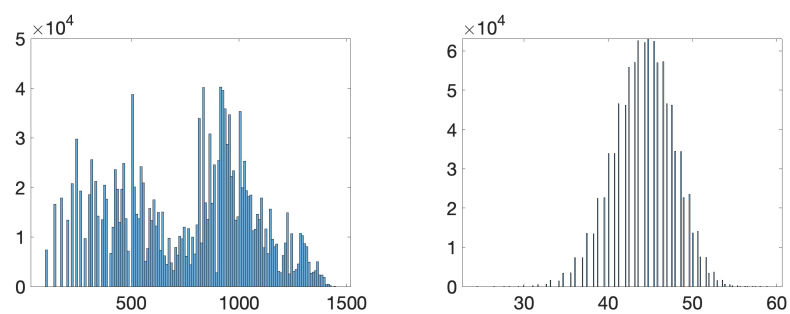


(c) APL1P, worst-case distribution. (d) 20TERM, worst-case distribution (prob. of highest-cost scen. is 0.8.)



(e) APL1P, effectiveness of scenarios. (f) 20TERM, effectiveness of scenarios.

Figure 3: Sorted cost of scenarios (x-axis) vs. top row: nominal distributions, middle row: worst-case distributions, and bottom row: effective/ineffective scenarios in DRO-W (y-axis). Left figures show results for APL1P and right figures for 1000-sample 20TERM.



(a) APL1P (b) 20TERM

Figure 4: Distribution of transportation costs between distinct scenarios in (a) APL1P and (b) 1000-sample 20TERM.

in 20TERM compared to APL1P. In fact, the average transportation cost between distinct pairs of scenarios for 20TERM is 44.40 while the standard deviation is 3.56. On the other hand, APL1P has an average transportation cost of 756.71 with a 312.2 standard deviation—a much higher value relative to its mean. As a point of comparison, in DRO-TV, transportation costs between distinct scenarios is always 1, and hence have a mean of 1 with 0 standard deviation.

In 20TERM, since the transportation costs between scenarios are similar, the worst-case expected problem of DRO-W tends to transport the probability mass to the highest-cost scenarios. Also, the ineffective scenarios are more likely to be the ones with lower costs. This results in DRO-W behaving similarly to DRO-TV, where mainly the highest-cost scenarios will have inflow probability mass (see [39] for details); also compare Figures 2(b) and 2(d). On the other hand, in APL1P, where the variance of the distances between distinct scenarios is more significant, multiple scenarios have relatively large probability mass inflows. Moreover, those scenarios do not necessarily have the highest costs. This is because the worst-case expected problem of DRO-W tends to transport the probability mass to a high-cost scenario *nearby* (i.e., using less transportation cost).

These results imply that the specific transportation cost used and how that transportation cost is distributed among different scenarios have an impact on the worst-case distribution and hence the effectiveness of scenarios in DRO-W.

7 Conclusion and Future Research

This paper studied effective scenarios in DROs formed via distance-based ambiguity sets, and it significantly expanded the results in [39] in several directions. First, for general distance-based ambiguity sets on a finite support, it streamlined and generalized the analysis by showing that the worst-case probability of a scenario being uniquely zero or always positive is sufficient to determine whether that scenario is ineffective or effective, respectively. Then, it provided easy-to-check sufficient conditions for identifying the effectiveness of scenarios in DRO-W. The earlier results for DRO-TV became a special case of the conditions presented in this paper. Furthermore, in several cases (e.g., when the dual variable corresponding to the transportation budget has multiple optima), the paper established that the proposed easy-to-check conditions identify the effectiveness of *all* scenarios in DRO-W. Numerical results reveal that these conditions work well and can identify a large portion of scenarios in the tested problems. In addition, the experiments imply that when the cost of transporting probability mass to a different scenario is relatively uniform (i.e., when the transportation costs between distinct scenarios have relatively lower variance), DRO-W appears to behave similar to DRO-TV. Otherwise, there could be a large range of scenarios whose costs cannot provide a good indication of their effectiveness.

This paper focused on fundamental properties of effective scenarios for general distance-based ambiguity sets and also easy-to-check conditions for DRO-W in a post-optimality fashion. This analysis forms the basis of their use for computational purposes, which merit future work. These include (i) problem-specific scenario reduction for DRO-W [e.g., 2], (ii) computation of efficient bounds [e.g., 6, 14], (iii) estimation of effective scenarios using machine learning (e.g., classification) methods, and (iv) devising algorithms to solve DROs that focus on effective scenarios for computational speedups [e.g., 52]. Future research also includes utilizing the general results presented in Section 3 to study effective scenarios in DROs formed with other ambiguity sets. Finally, extending such general results to continuous distributions, including more general settings of DRO-W, and for multistage DROs [e.g., 38] also constitute valuable future research directions.

Acknowledgments. This research is supported in part by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under Contract No. DE-AC02-06CH11357 and Grant DE-SC0023361.

References

[1] Agarwal, P. K., Har-Peled, S., Varadarajan, K. R., et al. (2005). Geometric approximation via coresets. *Combinatorial and computational geometry*, 52(1):1–30.

[2] Arpón, S., Homem-de Mello, T., and Pagnoncelli, B. (2018). Scenario reduction for stochastic programs with conditional value-at-risk. *Mathematical Programming*, 170(1):327–356.

[3] Arrigo, A., Ordoudis, C., Kazempour, J., De Grève, Z., Toubeau, J.-F., and Vallée, F. (2022). Wasserstein distributionally robust chance-constrained optimization for energy and reserve dispatch: An exact and physically-bounded formulation. *European Journal of Operational Research*, 296(1):304–322.

[4] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

[5] Bayraksan, G. and Love, D. K. (2015). Data-driven stochastic programming using phi-divergences. In *The Operations Research Revolution*, Tutorials in Operations Research, pages 1–19. INFORMS.

[6] Bayraksan, G., Maggioni, F., Faccini, D., and Yang, M. (2024). Bounds for multistage mixed-integer distributionally robust optimization. *SIAM Journal on Optimization*, 34(1):682–717.

[7] Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357.

[8] Bertsimas, D. and Tsitsiklis, J. N. (1997). *Introduction to Linear Optimization*. Athena Scientific.

[9] Blanchet, J. and Murthy, K. (2019). Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600.

[10] Blanchet, J., Murthy, K., and Zhang, F. (2022). Optimal transport-based distributionally robust optimization: Structural properties and iterative schemes. *Mathematics of Operations Research*, 47(2):1500–1529.

[11] Calafiore, G. and Campi, M. C. (2005). Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46.

[12] Calafiore, G. C. (2007). Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18(3):853–877.

[13] Chen, Z., Kuhn, D., and Wiesemann, W. (2022). Data-driven chance constrained programs over Wasserstein balls. *Operations Research*, 72(1):410–424.

[14] Cheramin, M., Cheng, J., Jiang, R., and Pan, K. (2022). Computationally efficient approximations for distributionally robust optimization under moment and Wasserstein ambiguity. *INFORMS J. Comput.*

[15] Czyzyk, J., Linderoth, J., and Shen, J. (2008). *SUTIL 0.1 (A Stochastic Programming Utility Library).*

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

[17] Dupačová, J., Gröwe-Kuska, N., and Römisch, W. (2003). Scenario reduction in stochastic programming: an approach using probability metrics. *Mathematical Programming*, 95:493–511.

[18] Duque, D. and Morton, D. P. (2020). Distributionally robust stochastic dual dynamic programming. *SIAM Journal on Optimization*, 30(4):2841–2865.

[19] Erdoğan, E. and Iyengar, G. (2006). Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1):37–61.

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

[21] Gao, R. and Kleywegt, A. (2023). Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research*, 48(2):603–655.

[22] Hiriart-Urruty, J.-B. and Lemaréchal, C. (2004). *Fundamentals of convex analysis.* Springer.

[23] Hu, Z. and Hong, L. J. (2013). Kullback-Leibler divergence constrained distributionally robust optimization. *Available at Optimization Online*, pages 1695–1724. https://optimization-online.org/?p=12225.

[24] Infanger, G. (1992). Monte Carlo (importance) sampling within a Benders decomposition algorithm for stochastic linear programs. *Annals of Operations Research*, 39(1):69–95.

[25] Jiang, R. and Guan, Y. (2018). Risk-averse two-stage stochastic program with distributional ambiguity. *Operations Research*, 66(5):1390–1405.

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

[27] Kuhn, D., Esfahani, P. M., Nguyen, V. A., and Shafieezadeh-Abadeh, S. (2019). Wasserstein distributionally robust optimization: Theory and applications in machine learning. In *Operations research & management science in the age of analytics*, Tutorials in Operations Research, pages 130–166. INFORMS.

[28] Kuhn, D., Shafiee, S., and Wiesemann, W. (2024). Distributionally robust optimization. *arXiv preprint arXiv:2411.02549*.

[29] Lee, C. and Mehrotra, S. (2015). A distributionally-robust approach for finding support vector machines. *Available from Optimization Online.*

[30] Lee, J. and Raginsky, M. (2018). Minimax statistical learning with Wasserstein distances. *Advances in Neural Information Processing Systems*, 31.

[31] Luo, F. and Mehrotra, S. (2019). Decomposition algorithm for distributionally robust optimization using Wasserstein metric with an application to a class of regression models. *European Journal of Operational Research*, 278(1):20–35.

[32] Mak, W.-K., Morton, D. P., and Wood, R. K. (1999). Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1-2):47–56.

[33] Mehrotra, S. and Zhang, H. (2014). Models and algorithms for distributionally robust least squares problems. *Mathematical Programming*, 146(1):123–141.

[34] Peajcariaac, J. E. and Tong, Y. L. (1992). *Convex functions, partial orderings, and statistical applications*. Academic Press.

[35] Pflug, G. C. (2001). Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming*, 89:251–271.

[36] Pflug, G. C. and Wozabal, D. (2007). Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442.

[37] Philpott, A. B., de Matos, V. L., and Kapelevich, L. (2018). Distributionally robust SDDP. *Computational Management Science*, 15(3):431–454.

[38] Rahimian, H., Bayraksan, G., and De-Mello, T. H. (2022). Effective scenarios in multistage distributionally robust optimization with a focus on total variation distance. *SIAM Journal on Optimization*, 32(3):1698–1727.

[39] Rahimian, H., Bayraksan, G., and Homem-de Mello, T. (2018). Identifying effective scenarios in distributionally robust stochastic programs with total variation distance. *Mathematical Programming*, 173:393–430.

[40] Rahimian, H., Bayraksan, G., and Homem-de Mello, T. (2019). Controlling risk and demand ambiguity in newsvendor models. *European Journal of Operational Research*, 279(3):854–868.

[41] Rahimian, H. and Mehrotra, S. (2022). Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization*, 3:1–85.

[42] Rockafellar, R. T. (1997). *Convex analysis*, volume 11. Princeton University Press.

[43] Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471.

[44] Shafieezadeh Abadeh, S., Mohajerin Esfahani, P. M., and Kuhn, D. (2015). Distributionally robust logistic regression. *Advances in Neural Information Processing Systems*, 28.

[45] Shapiro, A. (2021). Tutorial on risk neutral, distributionally robust and risk averse multistage stochastic programming. *European Journal of Operational Research*, 288(1):1–13.

[46] Sierksma, G. and Zwols, Y. (2015). *Linear and integer optimization: theory and practice, third edition.* CRC Press.

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

[48] Wozabal, D. (2012). A framework for optimization under ambiguity. *Annals of Operations Research*, 193(1):21–47.

[49] Xie, W. (2020). Tractable reformulations of two-stage distributionally robust linear programs over the type-∞ Wasserstein ball. *Operations Research Letters*, 48(4):513–523.

[50] Zhao, C. and Guan, Y. (2015). Data-driven stochastic unit commitment for integrating wind generation. *IEEE Transactions on Power Systems*, 31(4):2587–2596.

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

[52] Zhou, C. and Bayraksan, G. (2024). Accelerating Benders decomposition using effective scenarios for two-stage distributionally robust optimization with total variation distance. *Working paper.*