

Distributionally Robust Newsvendor Problems with Variation Distance

Hamed Rahimian^{*1}, Güzin Bayraksan^{†1}, Tito Homem-de-Mello ^{‡2}

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

²School of Business, Universidad Adolfo Ibañez, Santiago, Chile

Abstract

We use distributionally robust stochastic programs (DRSP) to model a general class of newsvendor problems where the underlying demand distribution is unknown, and so the goal is to find an order quantity that minimizes the worst-case expected cost among an ambiguity set of distributions. The ambiguity set consists of those distributions that are not far—in the sense of the so-called variation distance—from a nominal distribution. The maximum distance allowed in the ambiguity set (called *level of robustness*) places the DRSP between the “classical” stochastic programming and robust optimization models, which correspond to setting the level of robustness to zero and infinity, respectively. The structure of the newsvendor problem allows us to analyze the problem from multiple perspectives: First, we derive explicit formulas and properties of the optimal solution as a function of the level of robustness. Moreover, we determine the regions of demand that are critical (in a precise sense) to optimal cost from the viewpoint of a risk-averse decision maker. Finally, we establish quantitative relationships between the distributionally robust model and the corresponding risk-neutral and classical robust optimization models, which include the price of optimism/pessimism, and the nominal/worst-case regrets, among others. Our analyses can help the decision maker better understand the role of demand uncertainty in the problem and can guide him/her to choose an appropriate level of robustness. We illustrate our results with numerical experiments on a variety of newsvendor problems with different characteristics.

^{*}rahimian.1@osu.edu

[†]bayraksan.1@osu.edu

[‡]tito.hmello@uai.cl; Corresponding author

1 Introduction

The newsvendor problem is fundamental to many operations management models in supply chain and inventory management, e.g., production of influenza vaccination (Chick et al., 2008). It has also been used in different contexts such as staffing problems (Harrison and Zeevi, 2005) and scheduling of operating rooms (Oliver et al., 2008), as well as the classical seat allocation model in revenue management (Littlewood, 1972). The newsvendor decides on how many units of a product should be produced before the uncertain demand is revealed. Because the demand is uncertain, the newsvendor must balance the costs of under- and over-production to decide on the optimal quantity.

There are multiple ways of formulating a newsvendor problem. For instance, one could consider “classical” stochastic programming (SP) or robust optimization (RO) approaches¹. In the classical SP-based newsvendor model, the decision maker (i) has complete knowledge of the underlying demand distribution, and (ii) is risk neutral—i.e., (s)he minimizes the expected cost with respect to that demand distribution. In the classical RO-based newsvendor problem, on the other hand, the decision maker (i) does not have any knowledge of the demand except for its range of possible values, and (ii) is very risk averse—i.e., (s)he minimizes the worst-case cost among all values of demand within that range of values.

Both “classical” approaches have shortcomings. For example, the expectation might not be an adequate way to capture risk. Moreover, the decision maker might have some (albeit incomplete) knowledge about the underlying demand distribution, obtained, e.g., from historical data and/or expert opinions. For instance, consider determining the number of influenza vaccines to produce before the influenza season. The manufacturer is likely to be risk averse because the lack of vaccines can cause mortalities. Also, only partial information is known about the influenza viruses before the season starts. Using the classical SP model—which ignores risk aversion and assumes the probability distribution of demand is known—may result in a suboptimal production quantity that leads to shortages, which is clearly an undesirable outcome in this setting. On the other hand, because the classical RO model ignores any knowledge gained about demand distribution, it may

¹There are, of course, many other ways of formulating the problem using stochastic programming and/or robust optimization techniques; we use the qualifier “classical” to refer to the particular formulations described in the text.

yield a suboptimal production quantity that is overly conservative, and therefore results in huge waste (since unused vaccines must be discarded). This is also problematic given the reality of limited budgets.

In cases where the decision maker is risk averse and/or there is some—but not full—knowledge about the underlying demand distribution, an alternative modeling approach is to use *distributionally robust stochastic programs* (DRSPs for short). In such models, the goal is to find a decision that minimizes the *worst-case expected cost*, where “worst-case” refers to a set of distributions called *ambiguity set*. The classical SP and RO models can be viewed as special cases of DRSP: When the ambiguity set of DRSP contains only one distribution, we obtain the classical SP model, whereas when the ambiguity set contains all demand distributions with the same support, we obtain the classical RO model. Thus, DRSP lies between the two approaches.

It is well known that under mild conditions (e.g., real-valued costs, convex ambiguity sets), DRSP is equivalent to a *risk-averse stochastic program with a coherent measure of risk* (see, e.g., [Artzner et al. 1999](#)). Our setting satisfies this equivalence relation; we will discuss this in more detail in Section 2. In these cases, there is a direct correspondence between the level of robustness—which can be viewed as the size of the ambiguity set—and the desired level of risk aversion.

One of the contributions of this paper is an analysis, in the context of the newsvendor problem, of how the decision maker can benefit from the flexibility allowed by the DRSP, parameterized by the level of robustness. We refer to the newsvendor problem modeled with DRSP as DRNV. Many existing studies on DRNV focus on obtaining the optimal order quantity and the worst-case probability distribution, ideally in closed-form expressions; we review such works in Section 1.3. In this paper we take a step further. The motivation for our study arises from the observation that a risk-averse decision maker is concerned about the possibility of having too little or too much demand, which translate respectively into excess inventory and excess backlog and hence high costs. We call such undesirable values of demand *critical regions*.

Our goal is then to quantify *exactly* what the critical regions of demand are by relating them directly to the level of risk aversion of the decision maker, which has a one-to-one correspondence with the level of robustness of DRNV. Establishing such a direct relationship yields multiple benefits. For example, it can help decision makers understand their risk attitude better so they can

determine an appropriate level of robustness for their problem. It may also encourage them to collect more accurate information surrounding critical demands since the tails of distributions are often neglected in standard statistical methods. Such analysis, of course, depends on how the ambiguity is defined. In our case, we focus on the construction of the ambiguity set via the so-called variation distance, as we shall see next.

1.1 The Setting

We study a single-period, single-product DRNV with the following characteristics. The decision maker (i) has some “belief” about the demand distribution, termed the *nominal distribution*, (ii) reckons the underlying demand distribution is close enough, in the sense of the *variation distance*, to the nominal distribution (thereby forms the ambiguity set as the set of distributions whose variation distance from the nominal distribution is bounded above by a *level-of-robustness* parameter), and (iii) minimizes the worst-case expected cost with respect to this ambiguity set. We refer to this problem as DRNV-V (see Section 2 for a formal definition and assumptions).

Using variation distance to form DRNV has several advantages. The first one is the intuitive meaning of the variation distance, which integrates/sums over absolute deviations between the probability densities/masses. Another advantage is the risk interpretation: As we shall see shortly, DRNV-V is equivalent to minimizing a convex combination of Conditional Value-at-Risk (CVaR) and worst case of the cost function under the nominal demand distribution. By adjusting the level of robustness, DRNV-V contains a range of coherent risk measures, from expectation at the least conservative to the worst-case cost at the most conservative model. This class of coherent risk measures is important for applications like influenza vaccination and water resources management, where the worst case can cause mortalities and extreme water shortages, respectively.

A third advantage of using the variation distance is tractability. DRNV-V admits a closed-form expression for the optimal order quantity, which can be analyzed under different levels of robustness. Finally, the special structure of the variation distance enables us to fully characterize the regions of demand that are most critical to the problem. As discussed above, these critical regions provide the decision makers with insightful information about their problem and its underlying uncertainty.

1.2 Contributions and Summary of Main Results

The contributions of this work and its main results are summarized as follows:

- (i) *Analysis of the optimal solution.* We derive closed-form expressions for the optimal order quantity and worst-case probability distribution to DRNV-V at different levels of robustness (Theorems 1 and 2). We show the optimal order quantity lies between the optimal solution to the classical SP model, referred to as *risk-neutral order quantity*, and the optimal solution to the classical RO model, referred to as *robust order quantity*. We also show the optimal order quantity is either monotonically non-increasing or non-decreasing in the level of robustness, depending on the distributional information and parameters of the problem. We identify a *critical level of robustness* for DRNV-V. The importance of such notion lies in that, as shown in Theorem 1, the optimal order quantity stabilizes at the robust order quantity when the decision maker's level of robustness is higher than the critical level of robustness.
- (ii) *Characterization of maximal effective demand regions.* In order to examine which demand levels are critical to optimal cost, we introduce the notion of *maximal effective subsets* of demand realizations for DRNV-V. In short, a subset is maximal effective if it is the largest set such that the removal of that set or any of its subsets causes a change in the optimal value. We exploit the structure of the variation distance and DRNV-V to analytically characterize these subsets at different levels of robustness (Theorem 3). We also provide conditions under which there exists a worst-case probability distribution to DRNV-V whose support is the maximal effective subset. Moreover, we show the maximal effective subsets are monotonically non-increasing and converge to the extremes of the support of the demand distribution as the level of robustness increases (Theorem 4).
- (iii) *Analysis of the optimal order quantities' performance in the DRNV-V, SP, and RO settings.* We investigate how the optimal DRNV-V order quantity performs in the classical SP and RO settings and how the optimal order quantities to the classical models performs in the DRNV-V setting. We propose the measures *price of optimism*, *price of pessimism*, *nominal regret*, and *worst-case regret* to analyze how valuable these order quantities and settings are. We further define the *indifferent-to-solution level of robustness* and *indifferent-to-distribution*

level of robustness (see Section 5) to balance the performance of DRNV-V relative to SP and RO.

We illustrate the above ideas via numerical examples and demonstrate how the risk-neutral order quantity plays an important role to characterize these maximal effective subsets. Finally, we discuss how our analysis can be used to provide guidelines on selecting the level of robustness and to better understand the underlying uncertainty.

1.3 Literature Review

Because we study a single-period, single-product newsvendor problem, our review focuses on this class of problems. Even within this group there are several variations; for example, in the lot-sizing problem one tries to balance the costs of inventory and backlog, whereas in the traditional newsvendor the goal is to maximize the net revenue. Such variations are expressed through the parameters defining the generic problem. Within the class of single-period, single-product model, we highlight related works on (i) DRNV formed via moment-based ambiguity sets, (ii) DRNV formed via distance-based ambiguity sets, and (iii) risk-averse newsvendor problems. We point to similarities and differences of this work to others in this group.

Scarf (1958) proposed the first DRNV where only the first two moments of a continuous demand distribution are used to form the ambiguity set. Pioneered by Scarf (1958), most studies on DRNV form the ambiguity set by all probability distributions with sufficiently close moments (typically up to second-order moments); as examples, see Gallego and Moon (1993) and Natarajan et al. (2008). Among these studies, some papers consider cost functionals other than the expectation; see, e.g., Perakis and Roels (2008) for regret-based cost functionals, and Han et al. (2014) and Yu et al. (2016) for risk functionals.

Compared to moment-based ambiguity sets, there is relatively little work on DRNV formed via distance-based ambiguity sets. These models consider probability distributions whose distances to a nominal distribution are sufficiently small. By adjusting the bound on the distance, the model can be made more, or less conservative (i.e., risk averse). Burg entropy (Wang et al., 2016), variation distance (Jiang and Guan, 2015), ζ -structure probability metrics (Zhao and Guan, 2015),

and Kolmogorov-Smirnov distance (Bertsimas et al., 2016) have been used as “distances” between probability distributions, where all models use the expectation cost functional. Our work belongs to this category and subsumes the model in Jiang and Guan (2015) as a special case (see Section 2.2 for details). It also provides a significantly more detailed study, including the critical demand regions.

Our work is also relevant to the literature on risk-averse newsvendor. Most research in this area studies the optimal solution’s behavior with respect to the level of risk. Gotoh and Takano (2007) obtain a closed-form expression of the optimal solution for the CVaR objective and provide a numerical procedure for the mean-CVaR objective. Ahmed et al. (2007) study mean-risk objective functions, where risk is either the p -th semideviation or CVaR. They discuss conditions under which the optimal solution is monotonically non-decreasing in the weight of the risk functional. Choi and Ruszczyński (2008) derive an equivalent mean-risk model for risk-averse newsvendor with a law invariant coherent measure of risk. They show the optimal solution is non-increasing in the weight of the risk functional. For a detailed overview of risk-averse analyses of the newsvendor problem, we refer to Choi et al. (2011).

We now contrast our work with those in the literature. Our paper studies the optimal order quantity of DRNV-V at different levels of robustness, similarly to the works for the risk-averse newsvendor. It also studies the worst-case probability distribution of DRNV-V, similarly to the moment-based ambiguity sets literature. However, the newsvendor model considered here is comprehensive in the sense that it encompasses virtually all parameter combinations that have appeared in the literature (see Section 2.2 for details). Consequently, we analyze conditions under which the optimal solution to DRNV-V is either non-increasing or non-decreasing depending on the problem parameters and the nominal distribution. We further relate this behavior to the risk-neutral and robust order quantities. The cost of such a comprehensive study is that we have to consider numerous cases, but we believe it is important to fully characterize the problem so the results become applicable to the various settings of the newsvendor model.

Another contribution of our work over and above the literature is the study of critical demand regions. To the best of our knowledge, this is the first work that defines the notion of maximal effective demand subsets for DRNV-V and characterizes these demand subsets in terms of the level

of robustness in the problem. The most relevant work to ours is [Rahimian et al. \(2016\)](#), which identifies effective and ineffective scenarios for convex DRSPs formed via the variation distance. However, that paper considers only random variables that can take a *finite* number of realizations, whereas here we have continuous support. The distinction is significant and goes beyond technicalities of discrete vs. continuous support: Indeed, the motivation behind the discrete support case is that of data-driven problems, where the nominal distribution is the discrete empirical distribution corresponding to observed data; so the notion of effective scenarios helps to indicate which data points are more critical. Here, the underlying distribution is continuous, which requires different notions of effectiveness and new analysis. Moreover, by exploiting the properties of the newsvendor problem, we are able—unlike [Rahimian et al. \(2016\)](#)—to fully characterize the maximal effective subsets of DRNV-V and relate these subsets to the worst-case probability distribution of DRNV-V.

Finally, we introduce the notions of price of optimism/pessimism and nominal/worst-case regrets for a general DRSP to measure the relative performance of the optimal solutions in different settings of DRSP, SP, and RO. Price of optimism has been used in the context of risk-averse optimization ([Zhang et al., 2016](#)). Nominal regret has also been used in the literature in various contexts: e.g., DRSP ([Gallego and Moon, 1993](#); [Perakis and Roels, 2008](#)), RO ([Averbakh, 2001](#)), and risk-averse optimization ([Shapiro et al., 2013](#); [Zhang et al., 2016](#)). However, to the best of our knowledge, the price of pessimism and worst-case regret are new and this is the first paper to propose levels of robustness for DRSP that balance prices and regrets with respect to the SP and RO models. For DRNV-V, we show that such levels of robustness always exist (Theorem 5).

1.4 Organization

The rest of the paper is organized as follows. The next section formally defines the problem setting and lists the assumptions and conditions used throughout the paper. It also provides examples of different variants covered by our setting. Section 3 characterizes the optimal order quantity and optimal worst-case probability distribution, and Section 4 introduces the maximal effective subsets of demand and investigates their behavior as the level of robustness increases. Section 5 introduces the concepts of the price of optimism/pessimism and nominal/worst-case regrets for a general DRSP. It also shows the levels of robustness that balance prices and regrets, which we call

indifference levels collectively, can be obtained for DRNV-V. Numerical experiments illustrate these results in Section 6, where we also present further insight on how to choose a level of robustness based on our analysis of maximal effective subsets and the indifference levels. Finally, we end the paper with a summary and future research directions in Section 7. All proofs are provided in the Appendix.

2 Problem Formulation and Assumptions

2.1 Problem Setup

Consider the classical stochastic programming model for the newsvendor problem

$$\min_{x \in \mathbb{X}} \mathbb{E}_{\mathbb{P}_0} [h(x, \xi)], \quad (\text{Risk Neutral})$$

where

$$h(x, \xi) := W(x - \xi)_+ + U(\xi - x)_+ - V\xi \quad (1)$$

represents the total net loss of the newsvendor for a fixed order quantity $x \in \mathbb{R}$ and uncertain demand realization ξ . In the above formulation, W and U can be interpreted as “overage” and “underage” costs, respectively, whereas V can be interpreted as the income resulting from the realized demand. Also, $(\cdot)_+ := \max\{0, \cdot\}$, and $\mathbb{X} := \{x : x \geq 0\}$ denotes the feasibility set for the decision variable x .

In problem (Risk Neutral), it is assumed that demand ξ follows the continuous distribution $\mathbb{P}_0$. We assume that ξ has a finite mean $\mathbb{E}_{\mathbb{P}_0} [\xi] = \mathbb{E}_{\mathbb{P}_0} [|\xi|] < \infty$. We adopt the following notation: Ω denotes the support of ξ , $\underline{\xi} := \inf\{\hat{\xi} : \hat{\xi} \in \Omega\}$ and $\bar{\xi} := \sup\{\hat{\xi} : \hat{\xi} \in \Omega\}$ with $0 \leq \underline{\xi} < \bar{\xi}$. We assume $\Omega = [\underline{\xi}, \bar{\xi}] \subset \mathbb{R}_+$ when demand is bounded and $\Omega = [\underline{\xi}, \bar{\xi}) \subseteq \mathbb{R}_+$, where $\bar{\xi} = \infty$, when demand is unbounded.

From (1), we see that the finite-mean assumption ensures that $\mathbb{E}_{\mathbb{P}_0} [h(x, \xi)] < \infty$ for any $x \in \mathbb{R}$. Let F be the cumulative distribution function (cdf) associated with $\mathbb{P}_0$: $F(t) := \mathbb{P}_0\{\xi \leq t\}$. It is well known that the optimal solution to (Risk Neutral) is obtained at the *risk-neutral order quantity* $x^{\text{neut}} := F^{-1}(Q)$, where $Q := \frac{U}{U+W}$ denotes the critical ratio of the classical SP-based

NV newsvendor problem. We will use the notation x^{neut} , Q , and F defined above throughout the paper.

Consider the distributionally robust version of the newsvendor problem, following the motivation outlined in Section 1. In such problem, the decision maker has some belief that $\mathbb{P}_0$ is the distribution of demand, but would like to allow for some perturbations of $\mathbb{P}_0$. Throughout the paper we refer to $\mathbb{P}_0$ as the nominal distribution. As discussed in Section 1, the set of possible perturbations of the nominal distribution $\mathbb{P}_0$ is called the ambiguity set. We assume that all distributions in the ambiguity set, including $\mathbb{P}_0$, are absolutely continuous with respect to the Lebesgue measure, ν . Let $p = \frac{d\mathbb{P}}{d\nu}$ denote the associated density function of $\mathbb{P}$ with respect to ν . Similarly, $p_0 = \frac{d\mathbb{P}_0}{d\nu}$ denotes the corresponding density function of $\mathbb{P}_0$. Notice that by the assumption on Ω , p_0 is strictly positive on $[\underline{\xi}, \bar{\xi}]$ or $[\underline{\xi}, \bar{\xi})$. Given this setup, recall that the variation distance between $\mathbb{P}$ and $\mathbb{P}_0$ is defined as $\int_{\Omega} |p(s) - p_0(s)| ds$.

DRNV-V, a distributionally robust version of stochastic program ([Risk Neutral](#)) via variation distance, can then be formulated as

$$\min_{x \in \mathbb{X}} \left\{ f_{\gamma}(x) := \sup_{\mathbb{P} \in \mathcal{P}_{\gamma}} \mathbb{E}_{\mathbb{P}} [h(x, \xi)] \right\}, \quad (\text{DRNV-V})$$

where

$$\mathcal{P}_{\gamma} := \left\{ p : \int_{\Omega} |p(s) - p_0(s)| ds \leq \gamma, \int_{\Omega} p(s) ds = 1, p \geq 0 \right\}. \quad (2)$$

The ambiguity set of distributions $\mathcal{P}_{\gamma}$ contains all probability distributions $\mathbb{P}$ whose variation distance to the nominal probability distribution $\mathbb{P}_0$ is limited by the level of robustness γ . For a given $x \in \mathbb{X}$, we refer to the inner problem of ([DRNV-V](#)) as the *worst-case expected problem* at x . Note that the worst-case expected problem is feasible because $\mathbb{P}_0 \in \mathcal{P}_{\gamma}$.

Consider a fixed $x \in \mathbb{X}$. By [Jiang and Guan \(2015, Theorem 1\)](#), the worst-case expected value

in (DRNV-V) can be written as the following risk measures

$$f_\gamma(x) = \begin{cases} \mathbb{E}_{\mathbb{P}_0} [h(x, \xi)], & \text{if } \gamma = 0, \\ \frac{\gamma}{2} \text{ess sup } h(x, \xi) + \left(1 - \frac{\gamma}{2}\right) \text{CVaR}_{\frac{\gamma}{2}} [h(x, \xi)], & \text{if } 0 < \gamma < 2, \\ \text{ess sup } h(x, \xi), & \text{if } \gamma \geq 2, \end{cases} \quad (3)$$

all with respect to the nominal distribution $\mathbb{P}_0$. Note that $\text{CVaR}_\beta [\cdot]$ is defined in terms of the cumulative probability $0 < \beta < 1$, i.e., we have $\text{CVaR}_\beta [h(x, \xi)] := \frac{1}{1-\beta} \int_\beta^1 \text{VaR}_\rho [h(x, \xi)] d\rho$ where $\text{VaR}_\rho [h(x, \xi)] := \inf\{u : \mathbb{P}_0 \{h(x, \xi) \leq u\} \geq \rho\}$ is the Value-at-Risk (VaR) at level ρ . Note that the variation distance in (2) has a maximum value of 2; therefore, $0 \leq \gamma \leq 2$ covers the whole spectrum.

2.2 Further Assumptions, Conditions, and Examples

In addition to the assumptions stated in the problem setup, we study (DRNV-V) under the following assumption on the problem parameters:

(A1) $U > 0$ and $W > 0$.

We do not impose any restriction on the sign of V . As mentioned earlier, one can interpret W and U as overage and underage costs, respectively; so it is reasonable to assume they are positive. Assumption (A1) also ensures two important features. First, it implies that h is jointly convex in x and ξ ; so we have a convex optimization problem. Second, it ensures that the critical ratio Q satisfies $0 < Q < 1$; so we have a well-defined quantile x^{neut} as the solution to the risk-neutral counterpart.

In order to cover the maximum number of parameter configurations, we investigate (DRNV-V) under the following exclusive conditions:

(C1) $W + V > 0$, $U - V > 0$, and $\bar{\xi} < \infty$;

(C2a) $W + V > 0$, $U - V = 0$, and $\bar{\xi}$ might be infinity;

(C2b) $W + V > 0$, $U - V < 0$, and $\bar{\xi}$ might be infinity;

(C3a) $W + V = 0$, $U - V > 0$, and $\bar{\xi} < \infty$;

(C3b) $W + V < 0$, $U - V > 0$, and $\bar{\xi} < \infty$.

We refer to Conditions (C2a) and (C2b) collectively as Condition (C2). Similarly, we refer to Conditions (C3a) and (C3b) collectively as Condition (C3). Consider a fixed $x \in \mathbb{R}$. Observe that $-(W + V)$ is the slope of $h(x, \cdot)$ when $\xi < x$ (see (1)), and $U - V$ is the slope of $h(x, \cdot)$ when $\xi \geq x$. Thus, Conditions (C1)–(C3) give us further information on the shape of $h(x, \cdot)$: Under Condition (C1), h is V-shaped in ξ on $\mathbb{R}$, under Condition (C2) it is non-increasing, and under Condition (C3), it is non-decreasing in ξ . This helps us keep track of the ess sup term in (3). Figure 1 depicts the shape of $h(x, \cdot)$ under Conditions (C1)–(C3).

We also need to pay attention to the CVaR term in (3). It is evident from Figure 1 that under any of the Conditions (C1), (C2b), (C3b), the (nominal) distribution of $h(x, \xi)$ is continuous. Under either Condition (C2a) or (C3a), on the other hand, there is a probability mass at $-Vx$; thus, the (nominal) distribution of $h(x, \xi)$ is a mixed type. In particular, $\text{VaR}_{\frac{\gamma}{2}}[h(x, \xi)]$ might have a probability mass in this case. These differences between the conditions play a crucial role in characterizing the optimal solution and critical demand regions.

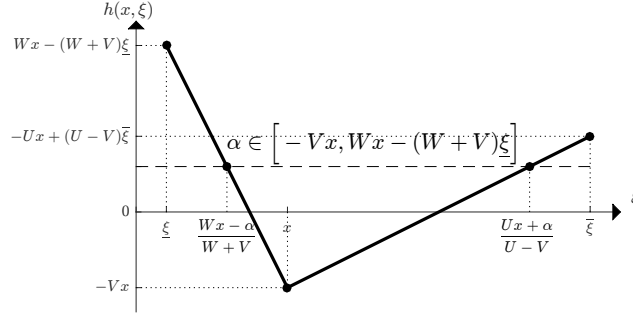
Under any of the Conditions (C1)–(C3), $|\text{ess sup } h(x, \xi)|$ is bounded. Moreover, $|\text{CVaR}_{\frac{\gamma}{2}}[h(x, \xi)]|$ is bounded because, as discussed earlier, $\mathbb{E}_{\mathbb{P}_0}[|h(x, \xi)|] < \infty$ under the assumption that ξ has a finite mean². Thus, $|f_{\gamma}(x)|$ is bounded, and as a result, (DRNV-V) has a finite optimal solution. Of course, the finite-mean assumption of ξ is automatically satisfied under (C1) and (C3) because of bounded support, but we need this assumption under Condition (C2)³.

Below, we present four examples and verify which condition their parameters belong to. All examples satisfy Assumption (A1).

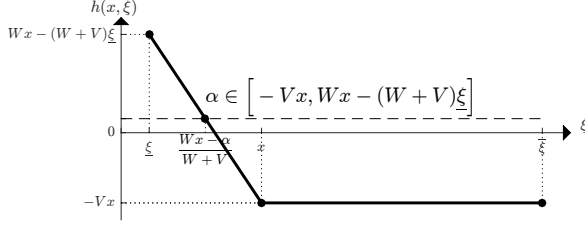
Example 1 (Lot-sizing problem) Consider a lot-sizing problem where the goal is to minimize the sum of purchase cost, inventory cost, and backlog cost, with per-unit costs respectively equal to $c > 0$, $m > 0$, and $b > c$. Then, $h(x, \xi) := cx + b(\xi - x)_+ + m(x - \xi)_+$, and one can show

² By a fundamental result in Rockafellar and Uryasev (2002), we have $\text{CVaR}_{\beta}[h(x, \xi)] \leq \mathbb{E}_{\mathbb{P}_0}[h(x, \xi) | h(x, \xi) > \text{VaR}_{\beta}[h(x, \xi)]]$. Then, $\mathbb{E}_{\mathbb{P}_0}[h(x, \xi) | h(x, \xi) > \text{VaR}_{\beta}[h(x, \xi)]] \leq \mathbb{E}_{\mathbb{P}_0}[|h(x, \xi)|] / (1 - \beta)$ for $0 < \beta < 1$ implies $|\text{CVaR}_{\frac{\gamma}{2}}[h(x, \xi)]|$ is bounded.

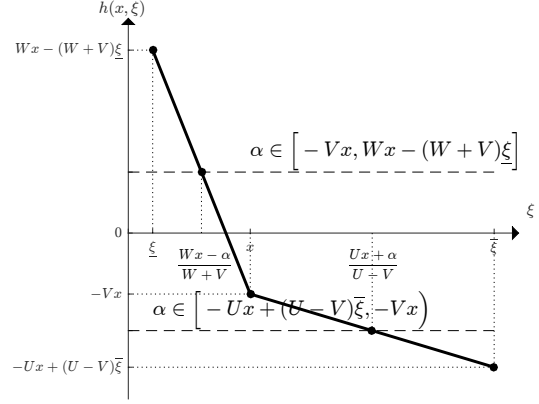
³ In fact, the finite-mean assumption of ξ is only needed for Condition (C2b). In the case of (C2a), we are guaranteed to satisfy a stronger result that $\text{ess sup } |h(x, \xi)|$ is bounded; see Figure 1b. In contrast, under Condition (C2b) when $\bar{\xi} = \infty$, even though $\text{ess sup } |h(x, \xi)| = \infty$, $|\text{ess sup } h(x, \xi)|$ is finite (see Figure 1c) and the finite-mean assumption of ξ guarantees $|\text{CVaR}_{\frac{\gamma}{2}}[h(x, \xi)]|$ is bounded.



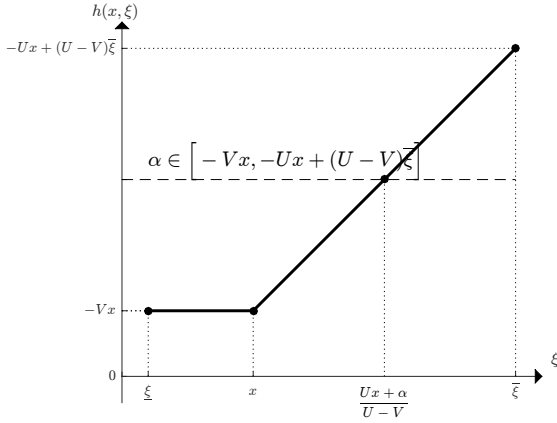
(a) Condition (C1).



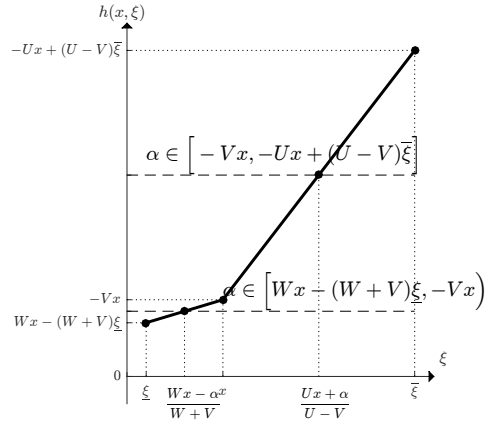
(b) Condition (C2a).



(c) Condition (C2b).



(d) Condition (C3a).



(e) Condition (C3b).

Figure 1: Cost function $h(x, \cdot)$ under Conditions (C1)–(C3) for a fixed $x \in \Omega$.

$W = c + m$, $U = b - c$, and $V = -c = U - b$. This problem satisfies Condition (C1). ■

Example 2 (Traditional newsvendor) Consider the traditional newsvendor problem where the goal is to maximize the profit. There is a per-unit purchase cost c , and per-unit revenue r and salvage value s with $s < c < r$. Then, $h(x, \xi) := cx - r \min\{x, \xi\} - s(x - \xi)_+$, and one can show $W = c - s$, $U = r - c$, and $V = r - c$. This problem satisfies Condition (C2a). ■

Example 3 (Two-stage newsvendor) Consider a newsvendor that can purchase before demand is realized at price c_1 but can also purchase after the demand is realized at a higher price c_2 . The per-unit revenue is $r > 0$, and it is assumed that $c_1 < c_2 < r$. Thus, if demand ξ is larger than the stock quantity x , the newsvendor purchases an additional $(\xi - x)$ same-day units. Then, one can show $W = c_1$, $U = c_2 - c_1$, and $V = r - c_1$. This problem satisfies Condition (C2b). ■

Example 4 (Lot-sizing with no inventory cost) Consider a lot-sizing problem where the inventory cost is negligible ($m = 0$). Then, $h(x, \xi) := cx + b(\xi - x)_+$ and one can show $W = c$, $U = b - c$, and $V = -W = U - b$. This problem satisfies Condition (C3a). ■

As mentioned before, our newsvendor model is comprehensive in the sense that it allows any parameter combinations that satisfy any of the Conditions (C1)–(C3). Condition (C1) and Condition (C2a) (e.g., Examples 1 and 2) are the most studied instances in the literature, e.g., (Wang et al., 2016; Jiang and Guan, 2015; Gotoh and Takano, 2007; Choi and Ruszczyński, 2008). Example 3 is from Zhao and Guan (2015), and it represents commodity markets (e.g., electricity), where there is a future contract plus a way to purchase in the spot market. Example 4 applies, for instance, to purchasing a computer server to store emails. When the amount of data to be stored is less than the available storage, there is no cost. An example of the remaining case of Condition (C3b) can be found in Ahmed et al. (2007). We do not present it here for brevity, but we refer the readers to Ahmed et al. (2007) for details of such a newsvendor instance.

3 Characterization of Optimal Solution

3.1 Preliminaries

Consider any $x \in \mathbb{X}$. First, because $h(x, \cdot)$ is continuous in ξ , we have $\text{ess sup } h(x, \xi) = \sup_{\xi \in \Omega} h(x, \xi)$ by [Phu and Hoffmann \(1996, Proposition 3.5\)](#). Next, for $0 < \gamma < 2$, by [Rockafellar and Uryasev \(2002, Theorem 10\)](#), $f_\gamma(x)$ in (3) is equivalent to minimizing a convex function $\mathcal{T}_\gamma(x, \alpha)$ —to be defined below—in α . Combining these and adding minimization over $x \in \mathbb{X}$, we rewrite (DRNV-V) as

$$\min_{x \in \mathbb{X}, \alpha \in \mathbb{R}} \mathcal{T}_\gamma(x, \alpha) := \frac{\gamma}{2} \sup_{\xi \in \Omega} h(x, \xi) + \left(1 - \frac{\gamma}{2}\right) \alpha + \mathbb{E}_{\mathbb{P}_0} \left[\left(h(x, \xi) - \alpha \right)_+ \right] \quad (4)$$

for $0 < \gamma < 2$. We will shortly characterize the optimal solution to (4), denoted by $(x_\gamma^*, \alpha_\gamma^*)$. The function $\mathcal{T}_\gamma(x, \alpha)$ and the optimal solution $(x_\gamma^*, \alpha_\gamma^*)$ all have subscripts γ to emphasize the dependence on the level of robustness γ . It is well known that an optimal $\alpha_\gamma^* = \text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)]$ for any $0 < \gamma < 2$. At $\gamma = 0$ and 2, we denote the optimal solution to (DRNV-V) by x_γ^* as well.

Observe from (3) that when the level of robustness is zero, i.e., $\gamma = 0$, (DRNV-V) reduces to the classical stochastic programming model, i.e., problem (Risk Neutral). On the other hand, when the decision maker is extremely conservative, i.e., $\gamma = 2$, (DRNV-V) reduces to the following classical robust optimization model

$$\min_{x \in \mathbb{X}} \sup_{\xi \in \Omega} h(x, \xi). \quad (\text{Robust})$$

Thus, by choosing an appropriate level of robustness $0 < \gamma < 2$, the decision maker becomes more conservative than problem (Risk Neutral) but he/she is less conservative than problem (Robust).

A quantity that will play an important role in characterizing the optimal solution to (DRNV-V) is the optimal order quantity to problem (Robust), which we call the *robust order quantity* and denote as x^{rob} . This quantity is

$$x^{\text{rob}} := \begin{cases} \frac{W+V}{W+U} \underline{\xi} + \frac{U-V}{W+U} \bar{\xi}, & \text{under Condition (C1),} \\ \underline{\xi}, & \text{under Condition (C2),} \\ \bar{\xi}, & \text{under Condition (C3).} \end{cases} \quad (5)$$

In Section 3.2, we show that for $0 < \gamma < 2$, the optimal order quantity to problem (DRNV-V) lies between the optimal order quantity to problems (Risk Neutral) and (Robust).

Under Condition (C1), x^{rob} is the order quantity x where the costs at $\underline{\xi}$ and $\bar{\xi}$ equalize. That is, $h(x^{\text{rob}}, \underline{\xi}) = h(x^{\text{rob}}, \bar{\xi})$. Under Condition (C1), observe that if $x > x^{\text{rob}}$, $h(x, \underline{\xi}) > h(x, \bar{\xi})$; otherwise, if $x < x^{\text{rob}}$, $h(x, \underline{\xi}) < h(x, \bar{\xi})$. Under either Condition (C2a) or (C3a), costs for all $\xi \in \Omega$ equalize at $x = x^{\text{rob}}$: $h(x^{\text{rob}}, \xi) = -Vx^{\text{rob}}$ for all $\xi \in \Omega$. For all conditions, $\sup_{\xi \in \Omega} h(x, \xi)$ is differentiable in x at every point except at the robust order quantity. The subdifferential of $\sup_{\xi \in \Omega} h(x, \xi)$ will be important in the proof of Theorem 1 below (see, e.g., Lemma 2 and proof of Theorem 1 in Appendix A).

In parallel, we define a *critical level of robustness* and denote it generally as γ^{cr} . As Theorem 1 below shows, this is the smallest level of robustness $0 \leq \gamma \leq 2$ at which the optimal order quantity to (DRNV-V) becomes the robust order quantity, $x_{\gamma}^* = x^{\text{rob}}$. For any $\gamma \geq \gamma^{\text{cr}}$, x^{rob} remains optimal. Recall that $x^{\text{neut}} = F^{-1}(Q)$, where $Q = \frac{U}{U+W}$. The critical level of robustness under various conditions is obtained as follows:

$$\gamma^{\text{cr}} := \begin{cases} 2 \left[Q - F \left(\frac{W+U}{W+V} x^{\text{rob}} - \frac{U-V}{W+V} x^{\text{neut}} \right) \right], & \text{under Condition (C1), if } x^{\text{neut}} > x^{\text{rob}}, \\ 2 \left[F \left(\frac{W+U}{U-V} x^{\text{rob}} - \frac{W+V}{U-V} x^{\text{neut}} \right) - Q \right], & \text{under Condition (C1), if } x^{\text{neut}} < x^{\text{rob}}, \\ 0, & \text{under Condition (C1), if } x^{\text{neut}} = x^{\text{rob}}, \\ 2Q, & \text{under Condition (C2),} \\ 2(1-Q), & \text{under Condition (C3).} \end{cases} \quad (6)$$

Note that $\gamma^{\text{cr}} < 2$ because $0 < Q < 1$. The relationship between x^{neut} and x^{rob} in (6) under Condition (C1) depends only on the problem parameters U, V, W , and the nominal distribution.

Theorem 1 below formally shows the optimal order quantity before and after γ^{cr} for various conditions. Remarks 1–4 discuss the implications of Theorem 1, further linking the robust order quantity to the risk-neutral order quantity. Then, Theorem 2 and remark to follow characterize the optimal worst-case probability distribution.

3.2 Optimal Solution and Discussion

Theorem 1. Consider (DRNV-V) with cost function defined in (1), x^{rob} defined in (5), and γ^{cr} defined in (6). Suppose Assumption (A1) holds.

(i) Under Condition (C1), consider the following two sets of equations:

$$x_\gamma^* = \frac{U-V}{W+U} F^{-1} \left(Q + \frac{\frac{\gamma}{2}(W-\zeta_\gamma)}{W+U} \right) + \frac{W+V}{W+U} F^{-1} \left(Q - \frac{\frac{\gamma}{2}(U+\zeta_\gamma)}{W+U} \right), \quad (7a)$$

$$\alpha_\gamma^* = \frac{W(U-V)}{W+U} F^{-1} \left(Q + \frac{\frac{\gamma}{2}(W-\zeta_\gamma)}{W+U} \right) - \frac{U(W+V)}{W+U} F^{-1} \left(Q - \frac{\frac{\gamma}{2}(U+\zeta_\gamma)}{W+U} \right), \quad (7b)$$

and

$$x_\gamma^* = x^{rob}, \quad (8a)$$

$$\frac{\gamma}{2} = F \left(\frac{Ux^{rob} + \alpha_\gamma^*}{U-V} \right) - F \left(\frac{Wx^{rob} - \alpha_\gamma^*}{W+V} \right), \quad (8b)$$

where $\zeta_\gamma = W$ or $-U$ in (7), depending on the conditions detailed below. Then, there exists a unique optimal solution to (4) as follows:

- If $x^{neut} > x^{rob}$, for $0 < \gamma < \gamma^{cr}$, $(x_\gamma^*, \alpha_\gamma^*)$ is given by (7) with $\zeta_\gamma = W$, and for $\gamma^{cr} \leq \gamma < 2$, $(x_\gamma^*, \alpha_\gamma^*)$ is given by (8).
- If $x^{neut} < x^{rob}$, for $0 < \gamma < \gamma^{cr}$, $(x_\gamma^*, \alpha_\gamma^*)$ is given by (7) with $\zeta_\gamma = -U$, and for $\gamma^{cr} \leq \gamma < 2$, $(x_\gamma^*, \alpha_\gamma^*)$ is given by (8).
- If $x^{neut} = x^{rob}$, for all $0 < \gamma < 2$, $(x_\gamma^*, \alpha_\gamma^*)$ is given by (8).

(ii) Under Condition (C2), there exists a unique optimal solution to (4) as follows:

$$\begin{cases} x_\gamma^* = F^{-1} \left(Q - \frac{\gamma}{2} \right) \text{ and } \alpha_\gamma^* = -Ux_\gamma^* + (U-V)F^{-1} \left(1 - \frac{\gamma}{2} \right), & \text{for } 0 < \gamma < \gamma^{cr}, \\ x_\gamma^* = x^{rob} \text{ and } \alpha_\gamma^* = -Ux_\gamma^* + (U-V)F^{-1} \left(1 - \frac{\gamma}{2} \right), & \text{for } \gamma^{cr} \leq \gamma < 2. \end{cases}$$

(iii) Under Condition (C3), there exists a unique optimal solution to (4) as follows:

$$\begin{cases} x_\gamma^* = F^{-1} \left(Q + \frac{\gamma}{2} \right) \text{ and } \alpha_\gamma^* = Wx_\gamma^* - (W+V)F^{-1} \left(\frac{\gamma}{2} \right), & \text{for } 0 < \gamma < \gamma^{cr}, \\ x_\gamma^* = x^{rob} \text{ and } \alpha_\gamma^* = Wx_\gamma^* - (W+V)F^{-1} \left(\frac{\gamma}{2} \right), & \text{for } \gamma^{cr} \leq \gamma < 2. \end{cases}$$

(iv) Finally, under any of the Conditions (C1)–(C3), there exists a unique optimal solution to (DRNV-V) as $x_\gamma^* = x^{\text{neut}}$ at $\gamma = 0$ and $x_\gamma^* = x^{\text{rob}}$ at $\gamma = 2$.

Remark 1. Recall the optimal order quantity $x^{\text{neut}} = F^{-1}(Q) = x_0^*$ to the classical stochastic programming model, i.e., $\gamma = 0$ in (DRNV-V), and the optimal order quantity $x^{\text{rob}} = x_2^*$ to the robust optimization model, i.e., $\gamma = 2$ in (DRNV-V). The optimal order quantity x_γ^* to (DRNV-V) for $0 < \gamma < 2$ lies between x^{neut} and x^{rob} . Remarks 2 and 3 discuss this in more detail.

Remark 2. Under Condition (C1),

- if $x^{\text{neut}} > x^{\text{rob}}$, as γ increases, the optimal order quantity x_γ^* decreases from x^{neut} according to (7a) with $\zeta_\gamma = W$, eventually reaching x^{rob} at $\gamma = \gamma^{\text{cr}}$. Then, x_γ^* stabilizes at x^{rob} for $\gamma > \gamma^{\text{cr}}$. This monotonic decreasing behavior of the optimal order quantity is implied by the $Q - \frac{\gamma}{2}$ term in (7a) with $\zeta_\gamma = W$. In this case, x^{rob} is a convex combination of $F^{-1}\left(Q - \frac{\gamma^{\text{cr}}}{2}\right)$ and $F^{-1}(Q)$.
- if $x^{\text{neut}} < x^{\text{rob}}$, as γ increases, the optimal order quantity x_γ^* increases from x^{neut} according to (7a) with $\zeta_\gamma = -U$, eventually reaching x^{rob} at $\gamma = \gamma^{\text{cr}}$. Then, x_γ^* stabilizes at x^{rob} for $\gamma > \gamma^{\text{cr}}$. This monotonic increasing behavior of the optimal order quantity is implied by the $Q + \frac{\gamma}{2}$ term in (7a) with $\zeta_\gamma = -U$. In this case, x^{rob} is a convex combination of $F^{-1}(Q)$ and $F^{-1}\left(Q + \frac{\gamma^{\text{cr}}}{2}\right)$.
- if $x^{\text{neut}} = x^{\text{rob}}$, $\gamma^{\text{cr}} = 0$. Thus, the optimal order quantity x_γ^* is already stabilized at x^{rob} , and the optimal solution remains *fixed* as γ increases.

Remark 3. We see a similar behavior under Conditions (C2) and (C3). Under Condition (C2), because $x^{\text{rob}} = \underline{\xi}$, $x^{\text{neut}} > x^{\text{rob}}$. So, as γ increases, the optimal order quantity x_γ^* decreases from x^{neut} , eventually reaching $\underline{\xi}$ at $\gamma = \gamma^{\text{cr}}$. It then stabilizes at $\underline{\xi}$. Similarly, under Condition (C3), we have $x^{\text{neut}} < x^{\text{rob}} = \bar{\xi}$. As γ increases, the optimal order quantity x_γ^* increases from x^{neut} , eventually reaching $\bar{\xi}$ at $\gamma = \gamma^{\text{cr}}$. It then stabilizes at $\bar{\xi}$. This monotonic decreasing/increasing behavior of x_γ^* can be easily seen from the corresponding formulas in Theorem 1.

Remark 4. Theorem 1 presented an *analytical* characterization of $(x_\gamma^*, \alpha_\gamma^*)$ for the following cases: under (C1) when $0 < \gamma < \gamma^{cr}$, and under either (C2) or (C3) for all $0 < \gamma < 2$. In the remaining case of Condition (C1) with $\gamma^{cr} \leq \gamma < 2$, the value of α_γ^* can be obtained by solving (8b) with a numerical (root-finding) method. In all cases, α_γ^* is monotonically increasing, as expected.

We now turn our attention to the probability distribution that attains the worst-case expected value at the optimal solution. Jiang and Guan (2015, Proposition 1) show this distribution for a general DRSP formed via the variation distance. Theorem 2 below further elaborates upon that result and specializes it to DRNV-V. In particular, we show conditions under which a (limiting) worst-case probability distribution exists. This distribution is not absolutely continuous with respect to the Lebesgue measure; so, it does not belong to the ambiguity set of (DRNV-V). However, it attains the worst-case expected value at the optimal solution in the sense that it can be written as the limit of feasible probability distributions to (DRNV-V).

As (9) below shows, under certain conditions and certain values of γ , this (limiting) distribution (i) suppresses the probability of demands below the VaR; (ii) adjusts the nominal density at VaR for (C2a) and (C3a); (iii) keeps demands whose costs are in between the VaR and the worst-case cost at the nominal-density level; and finally, (iv) distributes the suppressed probability of $\frac{\gamma}{2}$ to the worst-case demand values. The adjustment at VaR in (9), given by (10), for Conditions (C2a) and (C3a) is needed because, under any of these conditions, there is a probability atom at VaR (Figure 1). In contrast, there is no probability atom at VaR under Conditions (C1), (C2b), or (C3b). At $\gamma = 2$, this (limiting) distribution allocates a probability mass of one to only the highest-cost demands; see (11). We also show conditions under which *all* probability distributions in the ambiguity set achieve the worst-case expected value at the optimal solution.

Theorem 2. Consider (DRNV-V) with cost function defined in (1), γ^{cr} defined in (6), and x_γ^* defined in Theorem 1 as the optimal solution to (DRNV-V). Suppose Assumption (A1) holds.

- (i) Under any of the Conditions (C1), (C2b), and (C3b) and all levels of robustness $0 < \gamma < 2$, or under either Condition (C2a) or (C3a) with $0 < \gamma < \gamma^{cr}$, consider the probability distribution

$\mathbb{P}_\gamma^*$ at x_γ^* , whose (discontinuous) density function is defined as

$$p_\gamma^*(\xi) = \begin{cases} 0, & \text{if } \xi : h(x_\gamma^*, \xi) < \text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)], \\ \sigma_\gamma p_0(\xi), & \text{if } \xi : h(x_\gamma^*, \xi) = \text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)], \\ p_0(\xi), & \text{if } \xi : \text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)] < h(x_\gamma^*, \xi) < \sup_{\xi \in \Omega} h(x_\gamma^*, \xi), \\ \kappa_\xi, & \text{if } \xi : h(x_\gamma^*, \xi) = \sup_{\xi \in \Omega} h(x_\gamma^*, \xi). \end{cases} \quad (9)$$

Above, σ_γ is defined as

$$\sigma_\gamma := \begin{cases} 1, & \text{under Condition (C1), (C2b), or (C3b) with } 0 < \gamma < 2, \\ \frac{1-Q}{(1-Q)+\frac{\gamma}{2}}, & \text{under Condition (C2a) with } 0 < \gamma < \gamma^{cr}, \\ \frac{Q}{Q+\frac{\gamma}{2}}, & \text{under Condition (C3a) with } 0 < \gamma < \gamma^{cr}, \end{cases} \quad (10)$$

and κ_ξ is arbitrarily chosen to satisfy $0 \leq \kappa_\xi \leq \frac{\gamma}{2}$ and $\sum_{\{\xi \in \Omega : h(x_\gamma^*, \xi) = \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)\}} \kappa_\xi = \frac{\gamma}{2}$. Also, under any of the Conditions (C1), (C2b), and (C3b) with $\gamma = 2$, consider the probability distribution $\mathbb{P}_\gamma^*$ at x_γ^* , whose (discontinuous) density function is defined as

$$p_\gamma^*(\xi) = \begin{cases} 0, & \text{if } \xi : h(x_\gamma^*, \xi) < \sup_{\xi \in \Omega} h(x_\gamma^*, \xi), \\ \kappa_\xi, & \text{if } \xi : h(x_\gamma^*, \xi) = \sup_{\xi \in \Omega} h(x_\gamma^*, \xi), \end{cases} \quad (11)$$

where κ_ξ is arbitrarily chosen as in (9).

Then, in any of the above cases, the distribution $\mathbb{P}_\gamma^*$ defined by (9) or (11) accordingly, attains the worst-case expected value in (DRNV-V) at x_γ^* , i.e., $\sup_{\mathbb{P} \in \mathcal{P}_\gamma} \mathbb{E}_\mathbb{P} [h(x_\gamma^*, \xi)] = \mathbb{E}_{\mathbb{P}_\gamma^*} [h(x_\gamma^*, \xi)]$. Moreover, there exists a sequence of probability distributions $\mathbb{P}_n \in \mathcal{P}_\gamma$ such that

- (a) $\mathbb{P}_n$ converges to $\mathbb{P}_\gamma^*$ in distribution as $n \rightarrow \infty$, and
- (b) $\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbb{P}_n} [h(x_\gamma^*, \xi)] = \mathbb{E}_{\mathbb{P}_\gamma^*} [h(x_\gamma^*, \xi)]$.

(ii) For the remaining cases of Conditions (C2a) and (C3a) with $\gamma^{cr} \leq \gamma \leq 2$, all probability distributions in the ambiguity set $\mathcal{P}_\gamma$, including $\mathbb{P}_0$, are optimal to the worst-case expected

problem at x_γ^* .

Remark 5. When (9) attains the worst-case expected value in (DRNV-V) at x_γ^* ,

- under Condition (C1) when $\gamma^{\text{cr}} \leq \gamma < 2$, there are two $\xi \in \Omega$ that have the highest cost of $\sup_{\xi \in \Omega} h(x_\gamma^*, \xi)$. These are $\underline{\xi}$ and $\bar{\xi}$. In this case, there are *multiple* limiting worst-case probability distributions. All of them can be obtained by keeping the first three categories in (9) the same and for the last category setting $\kappa_{\underline{\xi}} = t \frac{\gamma}{2}$ and $\kappa_{\bar{\xi}} = (1 - t) \frac{\gamma}{2}$, where $0 \leq t \leq 1$.
- otherwise, there is only one $\xi \in \Omega$ with the highest cost (either $\underline{\xi}$ or $\bar{\xi}$). Then, $\kappa_\xi = \frac{\gamma}{2}$ and there is a *unique* limiting worst-case probability distribution given by (9).

A similar argument holds when (11) attains the worst-case expected value in (DRNV-V) at x_γ^* .

4 Characterization of Maximal Effective Subsets

4.1 Definitions

In this section, we discuss the notion of maximal effective subsets for (DRNV-V) and provide a precise definition for this concept. While effective sets were introduced in Rahimian et al. (2016) for finite Ω , here we generalize that definition for the setting of continuous distributions. This requires dealing with new technical issues, as we shall see shortly. The idea behind effective subsets is to examine whether the optimal value of (DRNV-V) changes when a non-empty subset of demand $\mathcal{F} \subset \Omega$ is removed from the problem. Let us first define what we mean by “removing” a subset of demand from the problem. We remove a subset $\mathcal{F}$ by restricting the ambiguity set $\mathcal{P}_\gamma$ to those probability distributions $\mathbb{P}$ for which $\mathbb{P}\{\mathcal{F}\} = 0$, i.e., $\frac{d\mathbb{P}}{d\nu} = p = 0$ on $\mathcal{F}$ Lebesgue-almost surely. This ensures that $\mathcal{F}$ is not in the support of the optimal worst-case probability distribution Lebesgue-almost surely. We call the resulting problem the *assessment problem of $\mathcal{F}$* .

More formally, the assessment problem of $\mathcal{F}$ can be formulated as

$$\min_{x \in \mathbb{X}} \left\{ f_\gamma^A(x; \mathcal{F}) := \sup_{p \in \mathcal{P}_\gamma^A(\mathcal{F})} \int_{\mathcal{F}^c} h(x, s) p(s) ds \right\}, \quad (12)$$

where

$$\mathcal{P}_\gamma^A(\mathcal{F}) := \left\{ p : \int_{\mathcal{F}^c} |p(s) - p_0(s)| ds \leq \gamma - \int_{\mathcal{F}} p_0(s) ds, \int_{\mathcal{F}^c} p(s) ds = 1, p \geq 0 \right\} \quad (13)$$

is the ambiguity set of probability distributions for the assessment problem, and $\mathcal{F}^c$ denotes the complement of set $\mathcal{F}$. We adopt the convention that the optimal value of the inner problem in (12) is equal to $-\infty$ if the set $\mathcal{P}_\gamma^A(\mathcal{F})$ is empty. The ambiguity set $\mathcal{P}_\gamma^A(\mathcal{F})$ of the assessment problem is a rearrangement of $\mathcal{P}_\gamma \cap \{p : p = 0 \text{ on } \mathcal{F} \text{ Lebesgue-almost surely}\}$. Compare (13) to (2) and observe the adjustments to the variation distance constraint and the next constraint that ensures p to be a probability density on $\mathcal{F}^c$.

We first adapt the following definition from Rahimian et al. (2016). Then, we formally define the maximal effective subsets.

Definition 1. *A subset $\mathcal{F} \subset \Omega$ is effective if the optimal value of the corresponding assessment problem (12) is strictly less than the optimal value of (DRNV-V). A subset $\mathcal{F} \subset \Omega$ is called ineffective if it is not effective.*

Observe that when Lebesgue measure of $\mathcal{F}$ is zero (i.e., $\nu(\mathcal{F}) = 0$), the problem essentially remains the same, and the optimal values of both (12) and (DRNV-V) are equal. Therefore, such $\mathcal{F}$ is ineffective. For example, a singleton set $\mathcal{F}$ (i.e., $\mathcal{F} = \{\xi\}$) is ineffective. To adequately handle subsets $\mathcal{F}$ with zero Lebesgue measure, we define the notion of *effective-in-limit*, which is a set that can be written as the limit of effective sets.

To define the notion of effective-in-limit and maximal effective subsets precisely, we first narrow our focus to intervals. Let

$$\mathfrak{S} := \{\mathcal{F} \subset \Omega : \mathcal{F} = \cup_{n=1}^N [a_n, b_n], \text{ for some } N \in \mathbb{N} \text{ and } a_n < b_n, n = 1, \dots, N\}.$$

In other words, $\mathfrak{S}$ contains all finite unions of closed intervals on Ω . Note that any set in $\mathfrak{S}$ has positive Lebesgue measure. In the following definition, we tie the effectiveness of a *singleton* subset $\mathcal{F}$ to the effectiveness of members of $\mathfrak{S}$. All definitions from this point on are new.

Definition 2. A singleton subset $\mathcal{F} \subset \Omega$ is called effective-in-limit if it can be written as a countably infinite intersection of effective subsets $\mathcal{F}_n \in \mathfrak{S}$, $n \in \mathbb{N}$, i.e., $\mathcal{F} = \bigcap_{n \in \mathbb{N}} \mathcal{F}_n$.

In this paper, we aim to characterize a subset of Ω that is maximal effective. Roughly speaking, it is the largest set such that any of its subsets is effective.

Definition 3. A subset $\mathcal{F} \subset \Omega$ is called maximal effective if it is the largest set such that (i) any subset of $\mathcal{F}$ that belongs to $\mathfrak{S}$ is effective and (ii) any singleton subset of $\mathcal{F}$ is effective-in-limit.

Note that a maximal effective subset can be a singleton set. In particular, it can be effective-in-limit (see Section 4.2).

4.2 Maximal Effective Subsets of DRNV-V

In this section, we first identify the maximal effective subsets of (DRNV-V) under different conditions. Then, we study the properties of these subsets in terms of the level of robustness. Finally, we relate maximal effective subsets to the (limiting) worst-case distribution.

We use the following notation in this section. For a given value of γ , set $\mathcal{E}_\gamma$ denotes the maximal effective subset of (DRNV-V). For any $B \subset \Omega$, we use $[\xi \in B]$ as a shorthand notation for the set $\{\xi \in \Omega : \xi \in B\}$. We are now ready to present the main results.

First, observe that when $\gamma = 0$, the maximal effective subset is given by $\mathcal{E}_0 = \Omega$ for any of the Conditions (C1)–(C3). This is because the ambiguity set (13) is empty for any subset $\mathcal{F}$ of Ω that belongs to $\mathfrak{S}$. Hence, the optimal value of the corresponding assessment problem (12) is $-\infty$ and, of course, smaller than $f_\gamma(x_\gamma^*)$. By Definition 1, all such $\mathcal{F}$ are effective. Therefore, we assume $\gamma > 0$ in Theorem 3 presented below.

Theorem 3. Consider (DRNV-V) with cost function defined in (1), γ^{cr} defined in (6), and $(x_\gamma^*, \alpha_\gamma^*)$ defined in Theorem 1 as the optimal solution to (4) for $0 < \gamma < 2$. Suppose Assumption (A1) holds. Then, for problem (DRNV-V), for all $0 < \gamma < 2$ we have:

$$(i) \text{ Under Condition (C1), } \mathcal{E}_\gamma = \left[\xi \leq \frac{Wx_\gamma^* - \alpha_\gamma^*}{W+V} \right] \cup \left[\xi \geq \frac{Ux_\gamma^* + \alpha_\gamma^*}{U-V} \right].$$

(ii) Under Condition (C2a),

- if $\gamma < \gamma^{cr}$, $\mathcal{E}_\gamma = \left[\xi \leq \frac{Wx_\gamma^* - \alpha_\gamma^*}{W+V} \right]$,
- if $\gamma \geq \gamma^{cr}$, $\mathcal{E}_\gamma = \{\underline{\xi}\}$.

(iii) Under Condition (C2b), $\mathcal{E}_\gamma = \left[\xi \leq \frac{Ux_\gamma^* + \alpha_\gamma^*}{U-V} \right]$.

(iv) Under Condition (C3a),

- if $\gamma < \gamma^{cr}$, $\mathcal{E}_\gamma = \left[\xi \geq \frac{Ux_\gamma^* + \alpha_\gamma^*}{U-V} \right]$.
- if $\gamma \geq \gamma^{cr}$, $\mathcal{E}_\gamma = \{\bar{\xi}\}$.

(v) Under Condition (C3b), $\mathcal{E}_\gamma = \left[\xi \geq \frac{Wx_\gamma^* - \alpha_\gamma^*}{W+V} \right]$.

At $\gamma = 2$, under Condition (C1), $\mathcal{E}_\gamma = \{\underline{\xi}, \bar{\xi}\}$; under Condition (C2), $\mathcal{E}_\gamma = \{\underline{\xi}\}$; and under Condition (C3), $\mathcal{E}_\gamma = \{\bar{\xi}\}$.

Remark 6. As discussed in Section 2.2, under Conditions (C1), (C2b), or (C3b), the (nominal) distribution of $h(x_\gamma^*, \xi)$ is continuous. For these cases, $\mathcal{E}_\gamma$ is equivalent to $\left\{ \xi \in \Omega : h(x_\gamma^*, \xi) \geq \text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)] \right\}$ for $0 < \gamma < 2$ and $\mathcal{E}_\gamma$ is equivalent to $\left\{ \xi \in \Omega : h(x_\gamma^*, \xi) = \sup_{\xi \in \Omega} h(x_\gamma^*, \xi) \right\}$ for $\gamma = 2$. However, under either Condition (C2a) or (C3a) with $0 < \gamma < \gamma^{cr}$, there is a probability mass at $\text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)]$ and $\left\{ \xi \in \Omega : h(x_\gamma^*, \xi) \geq \text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)] \right\} = \Omega$. In these cases, $\mathcal{E}_\gamma$ is equivalent to the closure of $\left\{ \xi \in \Omega : h(x_\gamma^*, \xi) > \text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)] \right\}$; further details can be found in the proof of Theorem 3. Furthermore, once $\gamma \geq \gamma^{cr}$, under either Condition (C2a) or (C3a), all costs equalize: $h(x^{\text{rob}}, \xi) = -Vx^{\text{rob}}$ for all $\xi \in \Omega$. This creates the split in parts (ii) and (iv) of Theorem 3.

Remark 7. Remark 4 discussed cases where an analytical characterization of $(x_\gamma^*, \alpha_\gamma^*)$ is available in Theorem 1. Using those analytical results, we can further refine $\mathcal{E}_\gamma$ for $0 < \gamma < 2$ as follows.

(i) Under Condition (C1),

- if $x^{\text{neut}} > x^{\text{rob}}$ and $0 < \gamma < \gamma^{cr}$, then $\mathcal{E}_\gamma = [\xi \leq F^{-1}(Q - \frac{\gamma}{2})] \cup [\xi \geq F^{-1}(Q)]$.
- if $x^{\text{neut}} < x^{\text{rob}}$ and $0 < \gamma < \gamma^{cr}$, then $\mathcal{E}_\gamma = [\xi \leq F^{-1}(Q)] \cup [\xi \geq F^{-1}(Q + \frac{\gamma}{2})]$.

(ii) Under Condition (C2a), if $0 < \gamma < \gamma^{cr}$, then $\mathcal{E}_\gamma = [\xi \leq F^{-1}(Q - \frac{\gamma}{2})]$.

(iii) Under Condition (C2b), $\mathcal{E}_\gamma = [\xi \leq F^{-1}(1 - \frac{\gamma}{2})]$.

(iv) Under Condition (C3a), if $0 < \gamma < \gamma^{cr}$, then $\mathcal{E}_\gamma = [\xi \geq F^{-1}(Q + \frac{\gamma}{2})]$.

(v) Under Condition (C3b), $\mathcal{E}_\gamma = [\xi \geq F^{-1}(\frac{\gamma}{2})]$.

Under Condition (C1) with $\gamma^{\text{cr}} \leq \gamma < 2$, we do not have analytical solution for α_γ^* . Thus, we cannot further refine $\mathcal{E}_\gamma$ like above.

Our next result shows that the sequence of maximal effective subsets $\mathcal{E}_\gamma$ are nested and monotonically non-increasing as $\gamma \rightarrow 2$. The nestedness of $\mathcal{E}_\gamma$ implies if a subset of realizations switches from effective to ineffective at some level of robustness, it will keep its status the same and will not change anymore as the level of robustness increases. [Rahimian et al. \(2016\)](#) show that such monotonicity and nestedness of effective subsets are not true in general (cf. Section 6.4 of that paper). However, we are able to obtain these properties for (DRNV-V).

Theorem 4. *Consider (DRNV-V) with cost function defined in (1). Suppose Assumption (A1) holds. Then,*

(i) $\mathcal{E}_{\gamma_1} \supseteq \mathcal{E}_{\gamma_2}$ for any $0 \leq \gamma_1 < \gamma_2 \leq 2$.

(ii) As $\gamma \rightarrow 2$, under Condition (C1), $\mathcal{E}_\gamma$ converges to $\{\underline{\xi}, \bar{\xi}\}$; under Condition (C2), $\mathcal{E}_\gamma$ converges to $\{\underline{\xi}\}$; and under Condition (C3), $\mathcal{E}_\gamma$ converges to $\{\bar{\xi}\}$.

Remark 8. By Theorem 4, the Lebesgue measure of $\mathcal{E}_\gamma$ (i.e., $\nu(\mathcal{E}_\gamma)$) is non-increasing as $\gamma \rightarrow 2$, but the rate at which $\nu(\mathcal{E}_\gamma)$ decreases depends on the problem parameters and the inverse of cdf of F . The probability of $\mathcal{E}_\gamma$ with respect to the nominal distribution, however, shows a trend. Observe that $\mathbb{P}_0\{\mathcal{E}_\gamma\} = 1 - \frac{\gamma}{2}$ under any of the Conditions (C1), (C2b), or (C3b). Hence, the probability of $\mathcal{E}_\gamma$ with respect to the nominal probability distribution shrinks linearly in γ in these cases. Likewise, under Condition (C2a) or (C3a), the probability of $\mathcal{E}_\gamma$ with respect to the nominal probability distribution shrinks linearly in γ until γ^{cr} , and then it remains at 0. In particular, under Condition (C2a), if $\gamma < \gamma^{\text{cr}}$, $\mathbb{P}_0\{\mathcal{E}_\gamma\} = Q - \frac{\gamma}{2}$. Otherwise, if $\gamma \geq \gamma^{\text{cr}}$, $\mathbb{P}_0\{\mathcal{E}_\gamma\} = 0$. Similarly, under Condition (C3a), if $\gamma < \gamma^{\text{cr}}$, $\mathbb{P}_0\{\mathcal{E}_\gamma\} = 1 - (Q + \frac{\gamma}{2})$. Otherwise, if $\gamma \geq \gamma^{\text{cr}}$, $\mathbb{P}_0\{\mathcal{E}_\gamma\} = 0$.

Our final result of this section makes a connection between the maximal effective subset and the worst-case probability distribution of (DRNV-V).

Corollary 1. Consider (DRNV-V) with cost function defined in (1) and γ^{cr} defined in (6). Suppose Assumption (A1) holds. Under any of the Conditions (C1), (C2b), and (C3b) and all levels of robustness $0 < \gamma \leq 2$, or under either Condition (C2a) or (C3a) with $\gamma^{cr} \leq \gamma \leq 2$, there exists a limiting worst-case probability distribution whose support is the maximal effective subset of (DRNV-V).

Remark 9. The result of Corollary 1 does not hold under either Condition (C2a) or (C3a) with $0 < \gamma < \gamma^{cr}$. As discussed in Remark 6, in these cases $\mathcal{E}_\gamma$ is equivalent to the closure of $\left\{ \xi \in \Omega : h(x_\gamma^*, \xi) > \text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)] \right\}$, whereas the support of the limiting worst-case probability distribution is Ω , according to Theorem 2.

5 Price of Optimism/Pessimism and Regret

Recall DRSP acts in between the classical SP and the classical RO models. We further elaborated in Remarks 1–3 that the distributionally robust optimal order quantity x_γ^* to (DRNV-V) lies between the risk-neutral order quantity x^{neut} to the classical SP-based newsvendor problem (i.e., $\gamma = 0$ in (DRNV-V)), and the robust order quantity x^{rob} to the classical RO-based newsvendor problem (i.e., $\gamma = 2$ in (DRNV-V)). In this section, we propose measures to evaluate the performance of x_γ^* in the classical SP and the classical RO models and vice versa. These measures may help decision makers understand how valuable these order quantities are and choose an appropriate level of robustness.

Recall function $f_\gamma(x)$ for a fixed $x \in \mathbb{X}$ and γ in (DRNV-V). For $0 \leq \gamma \leq 2$, the following two sets of inequalities hold:

$$f_\gamma(x^{\text{neut}}) \geq f_\gamma(x_\gamma^*) \geq f_0(x_\gamma^*) \geq f_0(x^{\text{neut}}), \quad (14)$$

$$f_\gamma(x_\gamma^*) \leq f_\gamma(x^{\text{rob}}) \leq f_2(x^{\text{rob}}) \leq f_2(x_\gamma^*). \quad (15)$$

In (14), $f_\gamma(x_\gamma^*) \geq f_0(x_\gamma^*)$ because for any $0 < \gamma \leq 2$, the worst-case expected problem at $x \in \mathbb{X}$ is a relaxation of the worst-case expected problem at x for $\gamma = 0$. At $\gamma = 0$, all the quantities in (14) are equal. Similarly in (15), $f_2(x^{\text{rob}}) \geq f_\gamma(x^{\text{rob}})$ because, for $\gamma = 2$, the worst-case expected problem at $x \in \mathbb{X}$ is a relaxation of the worst-case expected problem at x for any $0 \leq \gamma < 2$. At

$\gamma = 2$ all the quantities in (15) are equal. The other inequalities in (14) and (15) are justified by a suboptimality argument.

To conduct our analysis, we define the following measures for $0 \leq \gamma \leq 2$:

$$\text{PO}_\gamma := f_\gamma(x^{\text{neut}}) - f_\gamma(x_\gamma^*), \quad \text{PP}_\gamma := f_\gamma(x^{\text{rob}}) - f_\gamma(x_\gamma^*), \quad (16)$$

$$\text{NR}_\gamma := f_0(x_\gamma^*) - f_0(x^{\text{neut}}), \quad \text{WR}_\gamma := f_2(x_\gamma^*) - f_2(x^{\text{rob}}). \quad (17)$$

The first equation in (16) measures the *Price of Optimism (PO)*—what we lose by believing that the true distribution is $\mathbb{P}_0$ (and therefore using the risk-neutral order quantity) when (DRNV-V) accurately represents the ambiguity in the distribution for any level of robustness $\gamma > 0$. Similarly, the second equation in (16) measures the *Price of Pessimism (PP)*—what we lose by being overly conservative when (DRNV-V) accurately represents the ambiguity in the distribution for any level of robustness $\gamma < 2$. The price of optimism and pessimism help decision makers understand how valuable the risk-neutral and robust order quantities are in the distributionally robust setting.

The first equation in (17) measures the *Nominal Regret (NR)*—what we lose compared to the classical SP model when the nominal distribution is indeed the true underlying distribution, but we use (DRNV-V) with $\gamma > 0$. Finally, the second equation in (17) measures the *Worst-case Regret (WR)*—what we lose by using (DRNV-V) with $\gamma < 2$ when in reality the true distribution puts a probability mass of one on the worst-case cost. Nominal regret has been referred to as the “expected value of additional information” (Gallego and Moon, 1993; Perakis and Roels, 2008). It can also be interpreted as the largest value that decision maker would be willing to pay for the knowledge of the underlying distribution, when the underlying distribution is the nominal distribution.

Observe that

$$\text{PO}_\gamma - \text{PP}_\gamma = f_\gamma(x^{\text{neut}}) - f_\gamma(x^{\text{rob}}). \quad (18)$$

So, the difference between the price of optimism and pessimism measures the difference in quality between the optimistic and pessimistic order quantities when (DRNV-V) is an accurate model. That is, when the price of optimism and pessimism equalize, both the risk-neutral and robust order quantities yield the same cost in (DRNV-V). We call the *smallest* level of robustness at which this

equivalence occurs as the *indifferent-to-solution level of robustness* and denote it as γ^S .

Similarly,

$$NR_\gamma - WR_\gamma = f_0(x_\gamma^*) - f_0(x^{\text{neut}}) - \left(f_2(x_\gamma^*) - f_2(x^{\text{rob}}) \right). \quad (19)$$

So, the difference between the nominal and worst-case regret measures the difference between losses in the optimistic and pessimistic scenarios due to an ill-calibrated (DRNV-V). That is, when the nominal and worst-case regret equalize, the costs of being unnecessarily ambiguous and of not being ambiguous enough are the same. We call the *smallest* level of robustness at which this equivalence occurs as the *indifferent-to-distribution level of robustness* and denote it as γ^D .

In the following theorem, we show that the indifference levels of robustness are well defined for (DRNV-V). We further elaborate on these notions in Section 6.4 and comment on how they can be used to choose an appropriate level of robustness.

Theorem 5. *Consider (DRNV-V) with cost function defined in (1) and the price of optimism, price of pessimism, nominal and worst-case regrets defined in (16)–(17). Let $\gamma^S = \min\{\gamma \in [0, 2] : PO_\gamma - PP_\gamma = 0\}$ be the indifferent-to-solution level of robustness and $\gamma^D = \min\{\gamma \in [0, 2] : NR_\gamma - WR_\gamma = 0\}$ be the indifferent-to-distribution level of robustness. Suppose Assumption (A1) holds. Then, γ^S and γ^D are well defined for problem (DRNV-V).*

6 Numerical Experiments

We now present numerical experiments for several instances of DRNV-V to illustrate the results established in the previous sections. Section 6.1 explains our experimental setup. Then, we examine the maximal effective subsets at different levels of robustness in Section 6.2, and we analyze the price of optimism/pessimism and regrets in Section 6.3. Finally, we discuss how our analysis can be used to guide a decision maker on choosing an appropriate level of robustness in Section 6.4.

6.1 Experimental Setup

We generated five test problems that satisfy Assumption (A1), presented in Table 1. Each problem satisfies one of the Conditions (C1)–(C3b). As indicated in Table 1, we used the Beta and

Table 1: Test problems.

Problem #	Parameters				Condition it satisfies	Distribution(s) tested for
	W	U	V	Q		
1	3	1	0.5	$\frac{1}{4}$	(C1)	Beta(4, 4, 2, 5), Beta(1, 5, 2, 5), Beta(1, 2.15, 2, 5)
2	0.5	1	1	$\frac{2}{3}$	(C2a)	Exp(0.5)
3	3	1	2	$\frac{1}{4}$	(C2b)	Exp(0.5)
4	1.2	0.4	-1.2	$\frac{1}{4}$	(C3a)	Beta(2, 5, 2, 5)
5	7.5	0.5	-10	$\frac{1}{16}$	(C3b)	Beta(2, 5, 2, 5)

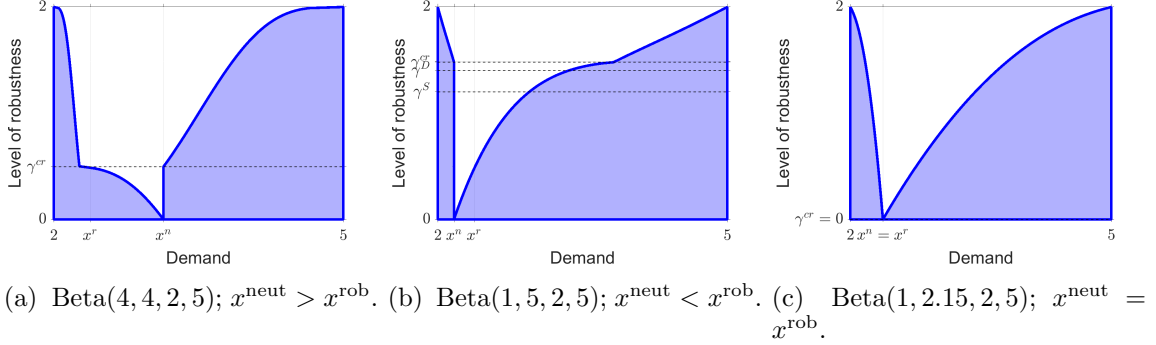


Figure 2: Maximal effective subset for Problem 1 and under Condition (C1).

Exponential distributions to model demand; so we have cases of distributions with bounded and unbounded support. In our notation, $\text{Beta}(\alpha, \beta, a, b)$ denotes a generalized beta distribution with the shape parameters α and β that is supported between a and b ($b > a$). Furthermore, $\text{Exp}(\beta)$ denotes an Exponential distribution with mean β .

We set the level-of-robustness parameter, γ , ranging from 0 to 2 in increments of 0.01. For numerical calculations to solve the root-finding problem of (8b) and to find the value of integral expressions, we used the Chebfun library (Driscoll et al., 2014).

6.2 Maximal Effective Subsets

In this section we discuss the maximal effective subsets $\mathcal{E}_\gamma$ for the test problems listed in Table 1 at different levels of robustness γ . The subsets $\mathcal{E}_\gamma$ are determined by Theorem 3. In Figures 2–4, the subsets $\mathcal{E}_\gamma$ form a shaded area. Also, x^n and x^r denote x^{neut} and x^{rob} , respectively, for short.

Figure 2 depicts the maximal effective subsets under Condition (C1) for $0 \leq \gamma \leq 2$. The three subfigures correspond respectively to the cases where $x^{\text{neut}} > x^{\text{rob}}$, $x^{\text{neut}} < x^{\text{rob}}$, and $x^{\text{neut}} = x^{\text{rob}}$.

Because all parameter values W, U, V and the range of the Beta distributions are the same for all three cases, x^{rob} remains the same across three subfigures. However, $x^{\text{neut}} = F^{-1}(Q)$ differs due to the change in the distributions. We can see that in all cases the maximal effective subsets (and hence maximal ineffective subsets) are nested; that is, if a subset of realizations is maximal effective at a level of robustness, then it is also maximal effective at all lower levels of robustness (cf. Theorem 4(i)). Moreover, the maximal effective subsets “shrink” as γ increases so that at the highest level of robustness (i.e., $\gamma = 2$), the maximal effective subset consists only of $\{\underline{\xi}, \bar{\xi}\}$, which are the extreme points of Ω (cf. Theorem 4(ii)).

Notice that the region of the maximal ineffective subset in all three graphs in Figure 2 has the shape of a “tornado” pointed at x^{neut} . That is, the region of maximal effective subset moves away from x^{neut} as γ increases. This is predicted by the theory (see Remark 7), but it can be interpreted as follows. Under Condition (C1) both the net demand cost $U - V$ and demand profit $W + V$ per unit are positive; so demand that is too high or too low yields larger costs. Thus, the goal of the newsvendor is to find a trade-off between the costs of under- and over-production so as to balance the costs of demand that is too high or too low. Of course, “too high” and “too low” are relative terms, which depend on the decision maker’s risk attitude. More specifically, the values of these thresholds correspond (via the choice of γ) to the ranges of demand realizations—or, equivalently, the range of costs—the decision maker cares about. A more conservative decision maker prefers the thresholds for “too high” and “too low” demand to be closer to the highest and lowest demands, respectively. Similarly, the less conservative the decision maker is, the farther the thresholds are from the highest and lowest demands, until of course they coincide—which is precisely where the risk-neutral order quantity is attained. Thus, in that sense, it is no surprise that the graph of maximal ineffective subsets is always pointed at x^{neut} in Figure 2.

Figure 2 also shows that, under (DRNV-V), for low levels of robustness $0 < \gamma < \gamma^{\text{cr}}$ the decision maker is only risk averse regarding one side of demand. This is determined by the relationship between x^{neut} and x^{rob} . For example, in Figure 2a (i.e., when $x^{\text{neut}} > x^{\text{rob}}$), we could say that the decision maker is increasingly concerned about demand realizations that are low, whereas the opposite occurs in Figure 2b. There is, however, a critical level of robustness, given by $\gamma = \gamma^{\text{cr}}$, after which the decision maker becomes risk averse to both high and low demand, and the optimal

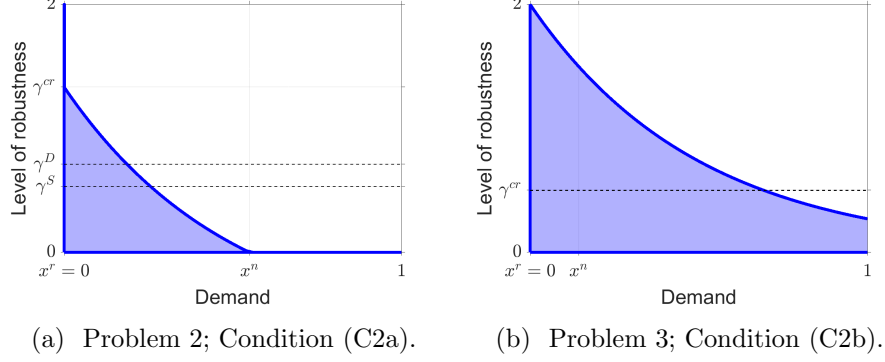


Figure 3: Maximal effective subset under Condition (C2) with Exp(0.5).

order quantity stabilizes at x^{rob} (cf. Theorem 1). Note that if $x^{\text{neut}} = x^{\text{rob}}$ then there is no such breakpoint, i.e., the decision maker is always risk averse to both high and low demand since $\gamma^{\text{cr}} = 0$ (see Figure 2c).

Figure 3 illustrates the maximal effective subsets under Condition (C2). In this case, the newsvendor is concerned about the low values of demand because the high values of demand do not increase the costs (Figures 1b and 1c). For both Problems 2 and 3, when $\gamma \geq \gamma^{\text{cr}}$, $x_\gamma^* = x^{\text{rob}} = \underline{x} = 0$, but their α_γ^* differ. Under (C2a), all costs equalize at $h(x^{\text{rob}}, \xi) = -Vx^{\text{rob}}$ for each $\xi \in \Omega$; so, α_γ^* remains constant. In contrast, under (C2b), the costs $h(x^{\text{rob}}, \xi)$ differ for each $\xi \in \Omega$, and α_γ^* steadily increases as $\gamma \rightarrow 2$. This results in the maximal effective subset being equal to $\{\underline{x}\}$ under (C2a) after $\gamma \geq \gamma^{\text{cr}}$, whereas the maximal effective subsets under (C2b) smoothly converge to that same singleton set as $\gamma \rightarrow 2$ (cf. Theorem 3(ii)–(iii) and Remark 7(iii)).

We see the same trends in Figure 4 under Condition (C3), but in the opposite direction. Under Condition (C3), the high values of demand become more important, and the newsvendor is indifferent to low values of demand because they do not increase the costs.

6.3 Price of Optimism/Pessimism and Regrets

Figure 5 shows the price of optimism, price of pessimism, and regrets for select problems from Table 1. As discussed in Section 5, the price of optimism and nominal regret approaches to zero as $\gamma \rightarrow 0$, whereas the price of pessimism and worst-case regret approaches to zero as $\gamma \rightarrow 2$. Figure 5 also reveals that the price of optimism and nominal regret are non-decreasing in γ , whereas the

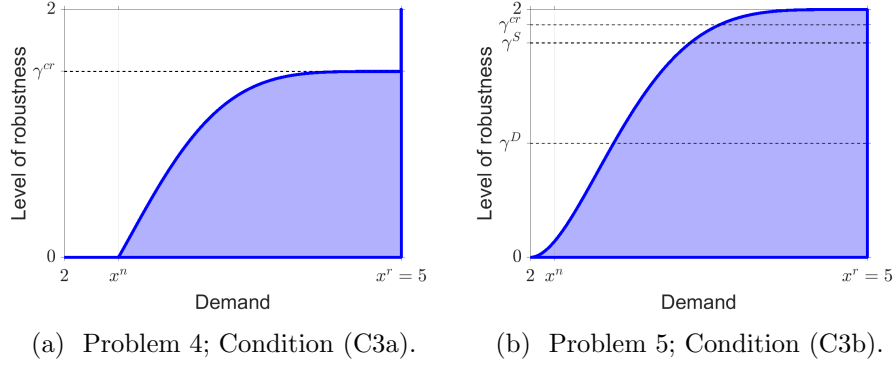


Figure 4: Maximal effective subset under Condition (C3) with Beta(2, 5, 2, 5).

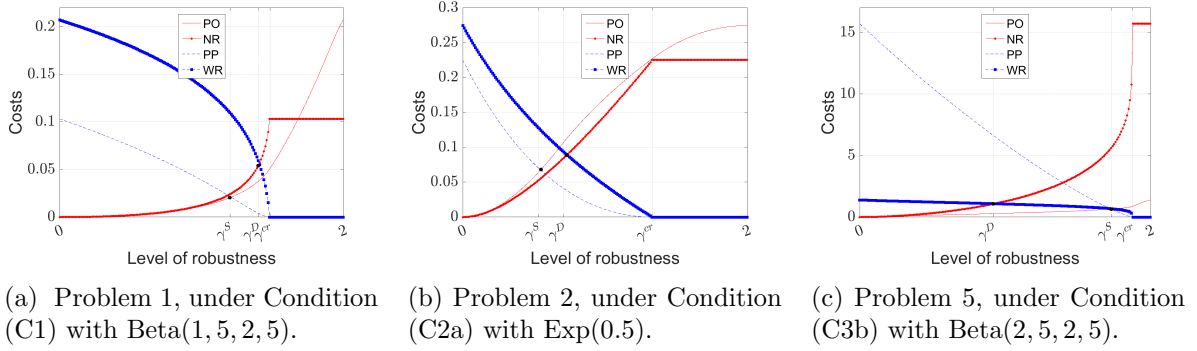


Figure 5: Price of optimism, price pessimism, nominal regret, and worst-case regret.

price of pessimism and worst-case regret are non-increasing in γ for these problems.

Under any of the Conditions (C1)–(C3), for $\gamma \geq \gamma^{\text{cr}}$, the price of pessimism and worst-case regret are zero because x_γ^* is stabilized at $x^{\text{rob}} = x_2^*$ (Figure 5). This implies that even the most conservative decision maker has no reason to choose a level of robustness higher than γ^{cr} . On the other hand, the nominal regret is a constant in these cases because $f_0(x_\gamma^*)$ is a constant. In particular, under Condition (C1) if $x^{\text{neut}} = x^{\text{rob}}$, the price optimism, the price pessimism, and regrets are all zero for $0 \leq \gamma \leq 2$ because $x_\gamma^* = x^{\text{neut}} = x^{\text{rob}}$. Note that $\gamma^{\text{S}} < \gamma^{\text{D}}$ in Figures 5a and 5b, and $\gamma^{\text{S}} > \gamma^{\text{D}}$ in Figure 5c, and both indifference levels of robustness $\gamma^{\text{S}}, \gamma^{\text{D}}$ are always less than γ^{cr} for all examples depicted in Figure 5.

6.4 Discussion

In general, the decision maker may not know which level of robustness γ to choose. We discuss how the analysis on the maximal effective subsets in Section 6.2 and on the price of optimism/pessimism, and regrets in Section 5 can help the decision maker gain insight about (DRNV-V) and guide on choosing an appropriate level of robustness. We illustrate these below with three examples:

- **Example 5 (Condition C1).** Problem 1, where demand follows Beta(1, 5, 2, 5).
- **Example 6 (Condition C2a).** Problem 2, where demand follows Exp(0.5).
- **Example 7 (Condition C3b).** Problem 5, where demand follows Beta(2, 5, 2, 5).

Choice of γ using Maximal Effective Subsets. The decision maker can associate the level of robustness γ with the “size” of the corresponding maximal effective subset $\mathcal{E}_\gamma$. If the decision maker wants to protect against $q_0\%$ of highest costs, then according to Remark 8, under Conditions (C1), (C2b), or (C3b), this can be interpreted as $\mathbb{P}_0 \left\{ h(x_\gamma^*, \xi) \geq \text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)] \right\} = q_0$ for $0 < \gamma < 2$; thus $\gamma = 2(1 - q_0)$. Moreover, under Conditions (C2a) and (C3a), this can be interpreted as $\mathbb{P}_0 \left\{ h(x_\gamma^*, \xi) > \text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)] \right\} = q_0$ for $0 < \gamma < \gamma^{\text{cr}}$. Thus, $\gamma = 2(Q - q_0)$ under Condition (C2a) and $\gamma = 2(1 - q_0 - Q)$ under Condition (C3a) with $0 < \gamma < \gamma^{\text{cr}}$. Note that if the decision maker wants to protect against a lower $q_0\%$ of highest costs, he/she chooses a higher level of robustness. This interpretation is in agreement with the decision maker’s risk attitude.

Suppose the decision maker wants to protect against 60% of highest costs. Then, we have

- Example 5: $\gamma = 0.8$ and $\mathcal{E}_\gamma \sim [\xi \leq 2.17] \cup [\xi \geq 2.57]$;
- Example 6: $\gamma = 0.13$ and $\mathcal{E}_\gamma \sim [\xi \leq 5.33]$;
- Example 7: $\gamma = 0.8$ and $\mathcal{E}_\gamma \sim [\xi \geq 2.67]$.

Choice of γ using Price of Optimism/Pessimism and Regrets. As discussed, the most conservative decision maker will choose γ^{cr} . But, if decision maker is less conservative, he/she can choose γ between γ^{S} and γ^{D} , depending on the risk attitude. If the decision maker wants to be indifferent regarding the error from using either the robust or the risk-neutral order quantities, then he/she chooses γ^{S} . If the decision maker wants to be indifferent regarding the error from using an ill-calibrated (DRNV-V) in either the optimistic and pessimistic scenarios, then he/she chooses γ^{D} .

Table 2 summarizes the results for Examples 5–7. Please see Figures 2b, 3a, and 4b for an illustration of the levels of robustness γ^S, γ^D and the corresponding maximal effective subsets $\mathcal{E}_\gamma$. In Example 5, we have $x^{\text{neut}} < x^{\text{rob}}$. Therefore, for this example, for all levels of robustness smaller than or equal to γ^{cr} —including γ^S and γ^D —the decision maker is equally concerned about demand realizations that are low (see Figure 2b). Table 2 shows this, where for $\gamma^{\text{cr}}, \gamma^S$, and γ^D , the corresponding $\mathcal{E}_\gamma$ are all the same at the left tail, however, the right tail changes according to γ . Moreover, for Example 6, $\mathcal{E}_{\gamma^{\text{cr}}} = \{\xi\}$ (see Theorem 3(ii)). So, $q_0\%$ is not well defined for this example because all costs are the same at γ^{cr} .

Table 2: The level of robustness γ , maximal effective subset $\mathcal{E}_\gamma$, and protection against $q_0\%$ of highest costs with the choice of the critical level of robustness γ^{cr} , indifferent-to-solution level of robustness γ^S , and indifferent-to-distribution level of robustness γ^D for Examples 5–7.

Example #	γ^{cr}			γ^S			γ^D		
	γ^{cr}	$\mathcal{E}_{\gamma^{\text{cr}}}$	$q_0\%$	γ^S	$\mathcal{E}_{\gamma^S}$	$q_0\%$	γ^D	$\mathcal{E}_{\gamma^D}$	$q_0\%$
5	1.48	$[\xi \leq 2.17] \cup [\xi \geq 3.82]$	26.00	1.21	$[\xi \leq 2.17] \cup [\xi \geq 2.96]$	39.50	1.41	$[\xi \leq 2.17] \cup [\xi \geq 3.39]$	29.50
6	1.33	$\{0\}$	—	0.55	$[\xi \leq 0.25]$	39.17	0.73	$[\xi \leq 0.18]$	30.17
7	1.88	$[\xi \geq 3.69]$	6.00	1.73	$[\xi \geq 3.42]$	7.25	0.92	$[\xi \geq 2.74]$	47.75

7 Conclusions and Future Research

In this paper, we studied a single-period, single-product distributionally robust newsvendor problem under uncertain demand and unknown demand distribution. We formed an ambiguity set of demand distributions by using the variation distance and protected against the expected cost with respect to the worst demand distribution in the ambiguity set. By exploiting the structure of the ambiguity set formed using the variation distance, we analytically characterized the optimal solution and the maximal effective subsets at different levels of robustness. We also defined the price of optimism/pessimism and nominal/worst-case regrets for a general DRSP to evaluate the value of the solution and distribution.

A summary of conclusions from this study are as follows:

- The optimal order quantity to DRNV-V is monotone in the level of robustness, lying between the optimal order quantities $x^{\text{neut}}, x^{\text{rob}}$ to the classical SP and RO models—moving in the direction of x^{neut} to x^{rob} , either monotonically non-increasing or non-decreasing according to

problem parameters and the nominal distribution.

- The optimal order quantity is stabilized at the robust order quantity x^{rob} when the decision maker's level of risk is higher than a critical level of robustness.
- The maximal effective subsets are nested and converge to the extremes of the support of demand.
- The maximal effective subsets and the price of optimism/pessimism and regrets provide insights on how to choose the level of robustness.
- The most conservative decision maker chooses the critical level of robustness. Less conservative decision makers can choose a level of robustness between the indifferent-to-solution level of robustness and the indifferent-to-distribution level of robustness, and by considering the percentage of highest costs using the maximal effective subsets.

Our notion of the maximal effective subset and the main results in this paper can lay the foundation for studying other problem settings in the context of DRNV. One direction for future research is studying multi-period and/or multi-product DRNVs; see, e.g., [Gallego and Moon \(1993\)](#); [Hanasusanto et al. \(2015\)](#); [Natarajan et al. \(2008\)](#); [Ardestani-Jaafari and Delage \(2016\)](#) for studies on single-period, multi-product DRNVs and [Ahmed et al. \(2007\)](#); [Shapiro \(2012\)](#); [Xin et al. \(2013\)](#); [Xin and Goldberg \(2015\)](#) for studies on multi-period, single-product DRNVs. Another avenue for future work is studying other cost functionals and ambiguity sets, including regret-based cost functionals and moment-based ambiguity sets, possibly in conjunction with more general ϕ -divergence based ambiguity sets ([Bayraksan and Love, 2015](#)). In all these settings, defining the notion of effectiveness of realizations and characterizing maximal effective subsets could provide the decision makers with useful insight about the underlying uncertainty.

Acknowledgements

First author gratefully acknowledges the support provided by a Presidential Fellowship from the Graduate School at the Ohio State University. The second author gratefully acknowledges the support of the National Science Foundation through grants CMMI-1345626 and CMMI-1563504.

References

- Ahmed, S., U. Çakmak, and A. Shapiro (2007). Coherent risk measures in inventory problems. *European Journal of Operational Research* 182(1), 226–238.
- Ardestani-Jaafari, A. and E. Delage (2016). Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research* 64(2), 474–494.
- Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance* 9, 203–227.
- Averbakh, I. (2001). On the complexity of a class of combinatorial optimization problems with uncertainty. *Mathematical Programming* 90(2), 263–272.
- Bayraksan, G. and D. K. Love (2015). Data-driven stochastic programming using phi-divergences. In *Tutorials in Operations Research*, pp. 1–19. INFORMS.
- Bertsimas, D., V. Gupta, and N. Kallus (2016). Robust sample average approximation. **arXiv preprint arXiv:1408.4445v3 [math.OC]**.
- Billingsley, P. (1999). *Convergence of Probability Measures* (2 ed.). John Wiley & Sons.
- Chick, S. E., H. Mamani, and D. Simchi-Levi (2008). Supply chain coordination and influenza vaccination. *Operations Research* 56(6), 1493–1506.
- Choi, S. and A. Ruszczyński (2008). A risk-averse newsvendor with law invariant coherent measures of risk. *Operations Research Letters* 36(1), 77–82.
- Choi, S., A. Ruszczyński, and Y. Zhao (2011). A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. *Operations Research* 59(2), 346–364.
- Driscoll, T. A., N. Hale, and L. N. Trefethen (2014). *Chebfun Guide*. Oxford: Pafnuty Publications.
- Gallego, G. and I. Moon (1993). The distribution free newsboy problem: review and extensions. *Journal of the Operational Research Society* 44(8), 825–834.
- Gotoh, J.-y. and Y. Takano (2007). Newsvendor solutions via conditional value-at-risk minimization. *European Journal of Operational Research* 179(1), 80–96.
- Han, Q., D. Du, and L. F. Zuluaga (2014). Technical note-a risk-and ambiguity-averse extension of the max-min newsvendor order formula. *Operations Research* 62(3), 535–542.
- Hanasusanto, G. A., D. Kuhn, S. W. Wallace, and S. Zymler (2015). Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming* 152(1-2), 1–32.
- Harrison, J. M. and A. Zeevi (2005). A method for staffing large call centers based on stochastic fluid models. *Manufacturing & Service Operations Management* 7, 20–36.
- Hiriart-Urruty, J. B. and C. Lemaréchal (2001). *Fundamentals of Convex Analysis*. Springer Berlin Heidelberg.
- Jiang, R. and Y. Guan (2015). Risk-averse two-stage stochastic program with distributional ambiguity. Technical report, Available on Optimization Online http://www.optimization-online.org/DB_HTML/2015/05/4908.html/.
- Krokhmal, P., M. Zabarankin, and S. Uryasev (2011). Modeling and optimization of risk. *Surveys in Operations Research and Management Science* 16(2), 49–66.
- Littlewood, K. (1972). Forecasting and control of passenger bookings. In *AGIFORS 12th Annual Symposium Proceedings*, Nathanya, Israel.

- Natarajan, K., M. Sim, and J. Uichanco (2008). Asymmetry and ambiguity in newsvendor models. Technical report, Available on Optimization Online http://www.optimization-online.org/DB_HTML/2008/07/2042.html/.
- Olivares, M., C. Terwiesch, and L. Cassorla (2008). Structural estimation of the newsvendor model: An application to reserving operating room time. *Management Science* 54(1), 41–55.
- Perakis, G. and G. Roels (2008). Regret in the newsvendor model with partial information. *Operations Research* 56(1), 188–203.
- Phu, H. and A. Hoffmann (1996). Essential supremum and supremum of summable functions: Summable functions. *Numerical Functional Analysis and Optimization* 17(1-2), 161–180.
- Rahimian, H., G. Bayraksan, and T. Homem-de Mello (2016). Identifying effective scenarios in distributionally robust stochastic programs with variation distance. Technical report, Available on Optimization Online http://www.optimization-online.org/DB_HTML/2016/03/5351.html/.
- Rockafellar, R. T. and S. Uryasev (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* 26(7), 1443–1471.
- Scarf, H. (1958). A min-max solution of an inventory problem. In H. Scarf, K. Arrow, and S. Karlin (Eds.), *Studies in the mathematical theory of inventory and production*, Volume 10, pp. 201–209. Stanford University Press Stanford, CA.
- Shapiro, A. (2012). Minimax and risk averse multistage stochastic programming. *European Journal of Operational Research* 219(3), 719–726.
- Shapiro, A., W. Tekaya, J. P. da Costa, and M. P. Soares (2013). Risk neutral and risk averse stochastic dual dynamic programming method. *European Journal of Operational Research* 224(2), 375–391.
- Shapiro, A. and Y. Wardi (1994). Nondifferentiability of the steady-state function in discrete event dynamic systems. *IEEE transactions on Automatic Control* 39(8), 1707–1711.
- Wang, Z., P. W. Glynn, and Y. Ye (2016). Likelihood robust optimization for data-driven problems. *Computational Management Science* 13(2), 241–261.
- Xin, L. and D. A. Goldberg (2015). Distributionally robust inventory control when demand is a martingale. *arXiv preprint arXiv:1511.09437 [math.PR]*.
- Xin, L., D. A. Goldberg, and A. Shapiro (2013). Time (in) consistency of multistage distributionally robust inventory models with moment constraints. *arXiv preprint arXiv:1304.3074 [math.OC]*.
- Yu, H., J. Zhai, and G.-Y. Chen (2016). Robust optimization for the loss-averse newsvendor problem. *Journal of Optimization Theory and Applications*, 1–25.
- Zhang, W., H. Rahimian, and G. Bayraksan (2016). Decomposition algorithms for risk-averse multistage stochastic programs with application to water allocation under uncertainty. *INFORMS Journal on Computing* 28(3), 385–404.
- Zhao, C. and Y. Guan (2015). Data-driven risk-averse two-stage stochastic program with ζ -structure probability metrics. Technical report, Available on Optimization Online http://www.optimization-online.org/DB_HTML/2015/07/5014.html/.

A Appendix A

A.1 Proofs of Section 3 (Characterization of Optimal Solution)

A.1.1 Notation and Preliminary Results for Theorem 1

Let us begin with some notation. Throughout this section, for $0 < \gamma < 2$, we work with the objective function $\mathcal{T}_\gamma(x, \alpha)$ given in (4). In particular, we refer to the $\mathbb{E}_{\mathbb{P}_0} \left[\left(h(x, \xi) - \alpha \right)_+ \right]$ term of $\mathcal{T}_\gamma(x, \alpha)$ as the *mean excess loss* and denote it as $\mathfrak{E}(x, \alpha)$. Furthermore, for $\eta \in \mathbb{R}$, we use $\Psi(x, \eta)$ and $\Psi(x, \eta^-)$ to denote $\mathbb{P}_0 \{h(x, \xi) \leq \eta\}$ and $\mathbb{P}_0 \{h(x, \xi) < \eta\}$, respectively. Recall that the cdf of ξ with respect to the nominal distribution is denoted as $F(t) = \mathbb{P}_0 \{\xi \leq t\}$.

The road map for the proof of Theorem 1 for $0 < \gamma < 2$ is as follows. First, we solve an unconstrained version of problem (4): $\min_{x \in \mathbb{R}, \alpha \in \mathbb{R}} \mathcal{T}_\gamma(x, \alpha)$, replacing $x \in \mathbb{X}$ with $x \in \mathbb{R}$. To find a stationary point of this unconstrained convex problem, we need to evaluate the partial subdifferentials of $\mathcal{T}_\gamma(x, \alpha)$ with respect to α and x

$$\partial_\alpha \mathcal{T}_\gamma(x, \alpha) = \left(1 - \frac{\gamma}{2}\right) + \partial_\alpha \mathfrak{E}(x, \alpha), \quad (20a)$$

$$\partial_x \mathcal{T}_\gamma(x, \alpha) = \frac{\gamma}{2} \partial_x \sup_{\xi \in \Omega} h(x, \xi) + \partial_x \mathfrak{E}(x, \alpha). \quad (20b)$$

We obtain a zero point of above system by considering $x \in \Omega$ (see Lemma 3) and $\inf_{\xi \in \Omega} h(x, \xi) \leq \alpha \leq \sup_{\xi \in \Omega} h(x, \xi)$ (see Lemma 4). Once we obtain a stationary point of the unconstrained problem for each case, we show that this solution is feasible and hence optimal for different values of γ .

The key to calculating (20) is the partial (sub)differentials of mean excess loss $\partial_\alpha \mathfrak{E}(x, \alpha)$ and $\partial_x \mathfrak{E}(x, \alpha)$ and the (sub)differential of highest cost $\partial_x \sup_{\xi \in \Omega} h(x, \xi)$. We first discuss these in Lemmas 1 and 2 below.

Lemma 1. *Suppose Assumption (A1) holds. Under any of the Conditions (C1)–(C3), for a fixed $x \in \mathbb{R}$ and $\alpha \in \mathbb{R}$, the partial (sub)differentials of $\mathfrak{E}(x, \alpha)$ with respect to x and α are obtained as*

$$\partial_\alpha \mathfrak{E}(x, \alpha) = [\Psi(x, \alpha^-) - 1, \Psi(x, \alpha) - 1], \quad (21a)$$

$$\begin{aligned}\partial_x \mathfrak{E}(x, \alpha) = & -U[\mathbb{P}_0 \{h(x, \xi) \geq \alpha \cap \xi > x\}, \mathbb{P}_0 \{h(x, \xi) > \alpha \cap \xi > x\}] + \\ & W[\mathbb{P}_0 \{h(x, \xi) > \alpha \cap \xi \leq x\}, \mathbb{P}_0 \{h(x, \xi) \geq \alpha \cap \xi \leq x\}].\end{aligned}\tag{21b}$$

Proof. For a fixed $x \in \mathbb{R}$, we can interchange the order of differentiation and expectation, i.e., $\partial_\alpha \mathfrak{E}(x, \alpha) = \mathbb{E}_{\mathbb{P}_0} \left[\partial_\alpha \left(h(x, \xi) - \alpha \right)_+ \right]$, by [Shapiro and Wardi \(1994, Proposition 2.1\)](#) because $(h(x, \xi) - \alpha)_+$ is convex in α . We can equivalently write this equation as

$$\partial_\alpha \mathfrak{E}(x, \alpha) = [-1, 0] \mathbb{P}_0 \{h(x, \xi) = \alpha\} - \mathbb{P}_0 \{h(x, \xi) > \alpha\} = [\Psi(x, \alpha^-) - 1, \Psi(x, \alpha) - 1].$$

With a similar argument, for a fixed $\alpha \in \mathbb{R}$, we have

$$\begin{aligned}\partial_x \mathfrak{E}(x, \alpha) = & -U \mathbb{P}_0 \{h(x, \xi) > \alpha \cap \xi > x\} + [-U, 0] \mathbb{P}_0 \{h(x, \xi) = \alpha \cap \xi > x\} + \\ & W \mathbb{P}_0 \{h(x, \xi) > \alpha \cap \xi \leq x\} + [0, W] \mathbb{P}_0 \{h(x, \xi) = \alpha \cap \xi \leq x\},\end{aligned}$$

which results in (21b). ■

For a given $x \in \mathbb{R}$, we can rewrite (20a) as $\partial_\alpha \mathcal{T}_\gamma(x, \alpha) = [\Psi(x, \alpha^-) - \frac{\gamma}{2}, \Psi(x, \alpha) - \frac{\gamma}{2}]$ by using [Lemma 1](#). So, $0 \in \partial_\alpha \mathcal{T}_\gamma(x, \alpha)$, or equivalently, $\frac{\gamma}{2} \in [\Psi(x, \alpha^-), \Psi(x, \alpha)]$ implies the corresponding optimal α is $\text{VaR}_{\frac{\gamma}{2}}[h(x, \xi)]$ for $0 < \gamma < 2$. Also note that $\mathcal{T}_\gamma(x, \alpha)$ is subdifferentiable with respect to x because $\sup_{\xi \in \Omega} h(x, \xi)$ is subdifferentiable with respect to x . In particular, as mentioned in [Section 3.1](#), $\sup_{\xi \in \Omega} h(x, \xi)$ is differentiable at any point except for at x^{rob} . We discuss $\partial_x \sup_{\xi \in \Omega} h(x, \xi)$ for a fixed $x \in \mathbb{R}$ next.

Lemma 2. *Suppose Assumption (A1) holds. Recall x^{rob} defined in (5). Consider a fixed $x \in \mathbb{R}$. Under any of the Conditions (C1)–(C3),*

- if $x > x^{\text{rob}}$, then $\partial_x \sup_{\xi \in \Omega} h(x, \xi) = \{W\}$,
- $x = x^{\text{rob}}$, then $\partial_x \sup_{\xi \in \Omega} h(x, \xi) = [-U, W]$,
- if $x < x^{\text{rob}}$, then $\partial_x \sup_{\xi \in \Omega} h(x, \xi) = \{-U\}$.

Proof. Under any of the Conditions (C1), (C2a), and (C3a), observe that $\sup_{\xi \in \Omega} h(x, \xi) = W(x - \underline{\xi}) - V\underline{\xi}$ if $x > x^{\text{rob}}$; $\sup_{\xi \in \Omega} h(x, \xi) = W(x - \underline{\xi}) - V\underline{\xi} = U(\bar{\xi} - x) - V\bar{\xi}$ if $x = x^{\text{rob}}$; and $\sup_{\xi \in \Omega} h(x, \xi) = U(\bar{\xi} - x) - V\bar{\xi}$ if $x < x^{\text{rob}}$. Under (C2b), observe that $\sup_{\xi \in \Omega} h(x, \xi) = W(x - \underline{\xi}) - V\underline{\xi}$ if $x > x^{\text{rob}}$;

$\sup_{\xi \in \Omega} h(x, \xi) = W(x - \underline{\xi}) - V\underline{\xi} = U(\underline{\xi} - x) - V\underline{\xi}$ if $x = x^{\text{rob}}$; and $\sup_{\xi \in \Omega} h(x, \xi) = U(\underline{\xi} - x) - V\underline{\xi}$ if $x < x^{\text{rob}}$. Under (C3b), $\sup_{\xi \in \Omega} h(x, \xi)$ is similar to that under (C2b) except for $\underline{\xi}$ is replaced by $\bar{\xi}$. Under any of the Conditions (C1)–(C3), observe that if $x > x^{\text{rob}}$, $\partial_x \sup_{\xi \in \Omega} h(x, \xi) = \{W\}$ and if $x < x^{\text{rob}}$, $\partial_x \sup_{\xi \in \Omega} h(x, \xi) = \{-U\}$. The only non-differentiability point is at $x = x^{\text{rob}}$, where $\partial_x \sup_{\xi \in \Omega} h(x, \xi) = [-U, W]$. ■

Next, we show that x_γ^* lies in the support of ξ , i.e., $x_\gamma^* \in \Omega$, and $\inf_{\xi \in \Omega} h(x_\gamma^*, \xi) \leq \alpha_\gamma^* \leq \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)$. The proof of Lemma 3 below is provided in Appendix B.

Lemma 3. *Suppose Assumption (A1) holds. For $0 \leq \gamma \leq 2$, the optimal x_γ^* to problem (DRNV-V) lies in the support of ξ , i.e., $x_\gamma^* \in \Omega$.*

Lemma 4. *Suppose Assumption (A1) holds. For $0 < \gamma < 2$, the optimal α_γ^* to problem (4) satisfies $\inf_{\xi \in \Omega} h(x_\gamma^*, \xi) \leq \alpha_\gamma^* \leq \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)$.*

Proof. Choose $0 < \gamma < 2$. First, suppose $\alpha < \inf_{\xi \in \Omega} h(x_\gamma^*, \xi)$. This implies $\Psi(x_\gamma^*, \alpha) = \Psi(x_\gamma^*, \alpha^-) = 0$, and hence $\frac{\partial}{\partial \alpha} \mathcal{T}_\gamma(x_\gamma^*, \alpha) = -\frac{\gamma}{2}$ by Lemma 1 and (20a). Now, suppose $\alpha > \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)$. This implies $\Psi(x_\gamma^*, \alpha) = \Psi(x_\gamma^*, \alpha^-) = 1$, and hence $\frac{\partial}{\partial \alpha} \mathcal{T}_\gamma(x_\gamma^*, \alpha) = 1 - \frac{\gamma}{2}$ by Lemma 1 and (20a). Thus, we find a zero point to (20) only if $\gamma = 0$ in the first case and only if $\gamma = 2$ in the second case, which are contradictions to our hypothesis. ■

Observe that Lemma 4 also holds for any fixed $x \in \mathbb{R}$. As discussed before, for a fixed $x \in \mathbb{R}$, the corresponding optimal α is $\text{VaR}_{\frac{\gamma}{2}}[h(x, \xi)]$, which lies between $\inf_{\xi \in \Omega} h(x, \xi)$ and $\sup_{\xi \in \Omega} h(x, \xi)$.

Before we present the proof of Theorem 1, let us elaborate on (20) under different conditions. For the cost function $h(x, \xi)$ defined in (1), only under either Condition (C2a) or (C3a), there is a probability atom at $\alpha = \inf_{\xi \in \Omega} h(x, \xi) = -Vx$. Otherwise, there is no probability atom at $\alpha \in \mathbb{R}$. In this latter case of no probability atom at α , $\mathfrak{E}(x, \alpha)$ is continuously differentiable in both α and x , and by using Lemma 1, we can rewrite (20) as

$$\partial_\alpha \mathcal{T}_\gamma(x, \alpha) = \left\{ \Psi(x, \alpha) - \frac{\gamma}{2} \right\}, \quad (22a)$$

$$\partial_x \mathcal{T}_\gamma(x, \alpha) = \frac{\gamma}{2} \partial_x \sup_{\xi \in \Omega} h(x, \xi) - U \mathbb{P}_0 \{h(x, \xi) > \alpha \cap \xi > x\} + W \mathbb{P}_0 \{h(x, \xi) > \alpha \cap \xi \leq x\}. \quad (22b)$$

If there is a probability atom at $\alpha \in \mathbb{R}$, we replace $\alpha = -Vx$ in $\mathcal{T}_\gamma(x, \alpha)$ to obtain a univariate function in x and directly calculate its (sub)derivative (cf. proof of Theorem 1, under (C2a)).

A.1.2 Proof of Theorem 1

Proof of Theorem 1. First, choose $0 < \gamma < 2$. By Lemma 3, $x_\gamma^* \in \Omega$ and by Lemma 4, $\inf_{\xi \in \Omega} h(x_\gamma^*, \xi) \leq \alpha_\gamma^* \leq \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)$. So, we fix an $x \in \Omega$ and $\inf_{\xi \in \Omega} h(x, \xi) \leq \alpha \leq \sup_{\xi \in \Omega} h(x, \xi)$. Figure 1 depicts the details under each condition.

- (i) Condition (C1). Observe that $\inf_{\xi \in \Omega} h(x, \xi) = -Vx$ (see Figure 1a). Also, $\sup_{\xi \in \Omega} h(x, \xi)$ is achieved either at $h(x, \underline{\xi})$, $h(x, \bar{\xi})$, or both. Suppose $\sup_{\xi \in \Omega} h(x, \xi) = h(x, \underline{\xi})$ and choose $\alpha \in [-Vx, Wx - (W + V)\underline{\xi}]$. If $\sup_{\xi \in \Omega} h(x, \xi) = h(x, \bar{\xi})$, a similar argument follows.

In this setting, it can be shown that α intersects with $h(x, \xi)$ at $\frac{Wx - \alpha}{W + V}$ when $\xi \leq x$ and at $\frac{Ux + \alpha}{U - V}$ when $\xi > x$. So, $\Psi(x, \alpha) = F\left(\frac{Ux + \alpha}{U - V}\right) - F\left(\frac{Wx - \alpha}{W + V}\right)$. As discussed, under Condition (C1), $\mathfrak{E}(x, \alpha)$ is continuously differentiable in both α and x , and by using Lemma 1, we have

$$\partial_\alpha \mathfrak{E}(x, \alpha) = F\left(\frac{Ux + \alpha}{U - V}\right) - F\left(\frac{Wx - \alpha}{W + V}\right) - 1, \quad (23a)$$

$$\partial_x \mathfrak{E}(x, \alpha) = UF\left(\frac{Ux + \alpha}{U - V}\right) + WF\left(\frac{Wx - \alpha}{W + V}\right) - U. \quad (23b)$$

Note that $\frac{Ux + \alpha}{U - V}$ might be greater than $\bar{\xi}$. In this case, $F\left(\frac{Ux + \alpha}{U - V}\right) = 1$. Now, by (20) and setting $\frac{\partial}{\partial \alpha} \mathcal{T}_\gamma(x, \alpha) = 0$ and $0 \in \partial_x \mathcal{T}_\gamma(x, \alpha)$, we obtain

$$F\left(\frac{Ux + \alpha}{U - V}\right) - F\left(\frac{Wx - \alpha}{W + V}\right) = \frac{\gamma}{2}, \quad (24)$$

$$U - UF\left(\frac{Ux + \alpha}{U - V}\right) - WF\left(\frac{Wx - \alpha}{W + V}\right) \in \frac{\gamma}{2} \partial_x \sup_{\xi \in \Omega} h(x, \xi). \quad (25)$$

By substituting (24) in (25), the stationary conditions are written as

$$U - \frac{\gamma}{2}U - (W + U)F\left(\frac{Wx - \alpha}{W + V}\right) \in \frac{\gamma}{2} \partial_x \sup_{\xi \in \Omega} h(x, \xi), \quad (26a)$$

$$U + \frac{\gamma}{2}W - (W + U)F\left(\frac{Ux + \alpha}{U - V}\right) \in \frac{\gamma}{2} \partial_x \sup_{\xi \in \Omega} h(x, \xi), \quad (26b)$$

in addition to (24). Now, we can use Lemma 2 and rewrite the above system (24), (26).

When $x > x^{\text{rob}}$, it simplifies to

$$F\left(\frac{Wx - \alpha}{W + V}\right) = Q - \frac{\gamma}{2} \text{ and } F\left(\frac{Ux + \alpha}{U - V}\right) = Q. \quad (27)$$

When $x < x^{\text{rob}}$, this system simplifies to

$$F\left(\frac{Wx - \alpha}{W + V}\right) = Q \text{ and } F\left(\frac{Ux + \alpha}{U - V}\right) = Q + \frac{\gamma}{2}. \quad (28)$$

When $x = x^{\text{rob}}$, it simplifies to

$$(24), \quad Q - \frac{\gamma}{2} \leq F\left(\frac{Wx - \alpha}{W + V}\right) \leq Q, \text{ and } Q \leq F\left(\frac{Ux + \alpha}{U - V}\right) \leq Q + \frac{\gamma}{2}. \quad (29)$$

Suppose $x^{\text{neut}} > x^{\text{rob}}$ and choose $0 < \gamma < \gamma^{\text{cr}}$. Because $\gamma^{\text{cr}} < 2Q$ in this case, (27) is valid and we can find the following point as the solution to (27)

$$x_{\gamma}^* = \frac{U - V}{W + U} F^{-1}(Q) + \frac{W + V}{W + U} F^{-1}\left(Q - \frac{\gamma}{2}\right), \quad (30a)$$

$$\alpha_{\gamma}^* = \frac{W(U - V)}{W + U} F^{-1}(Q) - \frac{U(W + V)}{W + U} F^{-1}\left(Q - \frac{\gamma}{2}\right). \quad (30b)$$

Now, because $Q - \frac{\gamma}{2} > Q - \frac{\gamma^{\text{cr}}}{2}$, we can conclude

$$\begin{aligned} x_{\gamma}^* &= \frac{U - V}{W + U} F^{-1}(Q) + \frac{W + V}{W + U} F^{-1}\left(Q - \frac{\gamma}{2}\right) \\ &> \frac{U - V}{W + U} F^{-1}(Q) + \frac{W + V}{W + U} \left(\frac{W + U}{W + V} x^{\text{rob}} - \frac{U - V}{W + V} x^{\text{neut}}\right) \\ &= x^{\text{rob}}, \end{aligned}$$

and hence (30) gives a stationary point. At $\gamma = \gamma^{\text{cr}}$, (30a) gives $x_{\gamma}^* = x^{\text{rob}}$, and (30) is a stationary point at $\gamma = \gamma^{\text{cr}}$ because it satisfies (29). Now, choose $\gamma^{\text{cr}} < \gamma < 2$. We verify $(x_{\gamma}^*, \alpha_{\gamma}^*)$ given by (8) satisfies (29); hence, it is optimal. First, note that $x_{\gamma}^* = x^{\text{rob}}$ by (8a), and (24) is the same as (8b). Recall at $x_{\gamma}^* = x^{\text{rob}}$, $\alpha_{\gamma}^* = \text{VaR}_{\frac{\gamma}{2}}[h(x^{\text{rob}}, \xi)]$. So, α_{γ}^* is increasing in

γ . This, combined with the fact that $(x^{\text{rob}}, \alpha_{\gamma^{\text{cr}}}^*)$ satisfies (27), imply $(x^{\text{rob}}, \alpha_{\gamma}^*)$ also satisfies $Q \leq F\left(\frac{Ux^{\text{rob}} + \alpha_{\gamma}^*}{U - V}\right)$ and $F\left(\frac{Wx^{\text{rob}} - \alpha_{\gamma}^*}{W + V}\right) \leq Q$ for $\gamma^{\text{cr}} < \gamma < 2$. To complete the proof, we only need to show $Q - \frac{\gamma}{2} \leq F\left(\frac{Wx^{\text{rob}} - \alpha_{\gamma}^*}{W + V}\right)$ and $F\left(\frac{Ux^{\text{rob}} + \alpha_{\gamma}^*}{U - V}\right) \leq Q + \frac{\gamma}{2}$ for $\gamma^{\text{cr}} < \gamma < 2$. Suppose by contradiction $F\left(\frac{Wx^{\text{rob}} - \alpha_{\gamma}^*}{W + V}\right) < Q - \frac{\gamma}{2}$. This implies $F\left(\frac{Ux^{\text{rob}} + \alpha_{\gamma}^*}{U - V}\right) - F\left(\frac{Wx^{\text{rob}} - \alpha_{\gamma}^*}{W + V}\right) > F\left(\frac{Ux^{\text{rob}} + \alpha_{\gamma}^*}{U - V}\right) - Q + \frac{\gamma}{2} \geq \frac{\gamma}{2}$, which is contradiction to (24) at $(x_{\gamma}^*, \alpha_{\gamma}^*)$. A similar contradiction is reached with the other term. As a result, $Q - \frac{\gamma}{2} \leq F\left(\frac{Wx^{\text{rob}} - \alpha_{\gamma}^*}{W + V}\right)$ and $F\left(\frac{Ux^{\text{rob}} + \alpha_{\gamma}^*}{U - V}\right) \leq Q + \frac{\gamma}{2}$.

Now, suppose $x^{\text{neut}} < x^{\text{rob}}$. Following similar arguments as above, using (28) instead of (27), and noting $\gamma^{\text{cr}} < 2(1 - Q)$ in this case, we obtain $(x_{\gamma}^*, \alpha_{\gamma}^*)$ given by (7) with $\zeta_{\gamma} = -U$ for $0 < \gamma < \gamma^{\text{cr}}$ and $(x_{\gamma}^*, \alpha_{\gamma}^*)$ given by (8) for $\gamma^{\text{cr}} < \gamma < 2$. Finally, suppose $x^{\text{neut}} = x^{\text{rob}}$. It is easy to verify that $Q \leq F\left(\frac{Ux^{\text{rob}} + \alpha_{\gamma}^*}{U - V}\right)$ and $F\left(\frac{Wx^{\text{rob}} - \alpha_{\gamma}^*}{W + V}\right) \leq Q$ for $\gamma^{\text{cr}} < \gamma < 2$. The rest of the proof follows a similar argument as that for the case $x^{\text{neut}} > x^{\text{rob}}$ and $\gamma^{\text{cr}} < \gamma < 2$.

To prove $(x_{\gamma}^*, \alpha_{\gamma}^*)$ is the unique optimal solution to (4), we show $\mathcal{T}_{\gamma}(x, \alpha)$ is strictly convex on $\Omega \times \mathbb{R}$. Let us first consider the $\mathfrak{E}(x, \alpha)$ term for $(x, \alpha) \in \Omega \times \mathbb{R}$. Observe from (23) that $\mathfrak{E}(x, \alpha)$ is twice differentiable and its Hessian is positive definite. So, it is a strictly convex function. On the other hand, $\frac{\gamma}{2} \sup_{\xi \in \Omega} h(x, \xi)$ is a convex function on Ω by Lemma 2 and $(1 - \frac{\gamma}{2})\alpha$ is linear in α . Consequently, $\mathcal{T}_{\gamma}(x, \alpha)$ is a strictly convex function, implying there is a unique optimal solution $(x_{\gamma}^*, \alpha_{\gamma}^*)$ to (4) for $0 < \gamma < 2$.

- (ii) Condition (C2). Observe that $\sup_{\xi \in \Omega} h(x, \xi) = W(x - \underline{\xi}) - V\underline{\xi}$ if $x > \underline{\xi}$, and $\sup_{\xi \in \Omega} h(x, \xi) = -Vx$ if $x = \underline{\xi}$. Figure 1 illustrates different cases below.

Condition (C2b). If $\bar{\xi} < \infty$, $\inf_{\xi \in \Omega} h(x, \xi) = -Ux + (U - V)\bar{\xi}$ and otherwise $-\infty$. We can break α into two separate cases: (1) $\alpha \in [-Ux + (U - V)\bar{\xi}, -Vx]$, where if $\bar{\xi} = \infty$, we choose α in an open interval, and (2) $\alpha \in [-Vx, Wx - (W + V)\underline{\xi}]$. First, choose $\alpha \in [-Ux + (U - V)\bar{\xi}, -Vx]$. When $x > \underline{\xi}$, we have $\frac{\partial}{\partial \alpha} \mathcal{T}_{\gamma}(x, \alpha) = 1 - F\left(\frac{Ux + \alpha}{U - V}\right) - \frac{\gamma}{2}$ and

$$\frac{\partial}{\partial x} \mathcal{T}_{\gamma}(x, \alpha) = \frac{\gamma}{2}W + (W + U)F(x) - UF\left(\frac{Ux + \alpha}{U - V}\right)$$

by (22) and Lemma 2. Setting $\frac{\partial}{\partial \alpha} \mathcal{T}_{\gamma}(x, \alpha) = 0$ and $\frac{\partial}{\partial x} \mathcal{T}_{\gamma}(x, \alpha) = 0$, we conclude $F(x) = Q - \frac{\gamma}{2}$. Consequently, if $\gamma < 2Q = \gamma^{\text{cr}}$, we obtain $x_{\gamma}^* = F^{-1}(Q - \frac{\gamma}{2}) > \underline{\xi}$ and $\alpha_{\gamma}^* = -Ux_{\gamma}^* + (U -$

$$V)F^{-1}\left(1 - \frac{\gamma}{2}\right).$$

Now, suppose $x = \underline{\xi}$. By (22), we have $\frac{\partial}{\partial \alpha} \mathcal{T}_\gamma(x, \alpha) = 1 - F\left(\frac{Ux+\alpha}{U-V}\right) - \frac{\gamma}{2}$ and $\partial_x \mathcal{T}_\gamma(x, \alpha) = \frac{\gamma}{2} \partial_x \sup_{\xi \in \Omega} h(x, \xi) - UF\left(\frac{Ux+\alpha}{U-V}\right)$. By setting $\frac{\partial}{\partial \alpha} \mathcal{T}_\gamma(x, \alpha) = 0$ and $0 \in \partial_x \mathcal{T}_\gamma(x, \alpha)$, we obtain $-Q \leq 0 \leq -Q + \frac{\gamma}{2}$. Hence, for $\gamma \geq 2Q$, we have $x_\gamma^* = \underline{\xi}$ and $\alpha_\gamma^* = -Ux_\gamma^* + (U-V)F^{-1}\left(1 - \frac{\gamma}{2}\right)$.

Because we already found an optimal solution, we no longer need to examine the case $\alpha \in [-Vx, Wx - (W+V)\underline{\xi}]$. Similar to Condition (C1), there is no probability atom at $\alpha \in \mathbb{R}$ for the cost function $h(x, \xi)$, and $\mathfrak{E}(x, \alpha)$ is twice differentiable in both x and α . So, with a similar argument, it can be verified that $\mathcal{T}_\gamma(x, \alpha)$ is strictly convex on the interior of $\Omega \times \mathbb{R}$, implying there is a unique optimal solution $(x_\gamma^*, \alpha_\gamma^*)$ to (4).

Condition (C2a). In this case, $\inf_{\xi \in \Omega} h(x, \xi) = -Vx$ and $\sup_{\xi \in \Omega} h(x, \xi) = Wx - (W+V)\underline{\xi}$. So, we choose $\alpha \in [-Vx, Wx - (W+V)\underline{\xi}]$. Below, we first show that $x > \underline{\xi}$ and $\alpha > -Vx$ is not optimal. Then, we examine the optimal solution when $x > \underline{\xi}$ and when $x = \underline{\xi}$.

First, suppose $x > \underline{\xi}$ and $\alpha > -Vx$. By (22) and Lemma 2, we have $\frac{\partial}{\partial \alpha} \mathcal{T}_\gamma(x, \alpha) = -F\left(\frac{Wx-\alpha}{W+V}\right) + 1 - \frac{\gamma}{2}$ and $\frac{\partial}{\partial x} \mathcal{T}_\gamma(x, \alpha) = W\left(F\left(\frac{Wx-\alpha}{W+V}\right) + \frac{\gamma}{2}\right)$. But, because $W > 0$ by Assumption (A1), we cannot find a stationary point in this case. Now, suppose $x > \underline{\xi}$ and $\alpha = -Vx$. By the definition of VaR, we must have $F(x) \leq 1 - \frac{\gamma}{2}$ in this case. Replacing $\alpha = -Vx$ in $\mathcal{T}_\gamma(x, \alpha)$, we obtain $\mathcal{T}_\gamma(x)$ —a univariate function in x . This function is differentiable with respect to x :

$$\frac{\partial}{\partial x} \mathcal{T}_\gamma(x) = \frac{\gamma}{2}W - \left(1 - \frac{\gamma}{2}\right)V + (W+V)F(x) = (W+U)\left(\frac{\gamma}{2} + F(x)\right) - U.$$

By setting $\frac{\partial}{\partial x} \mathcal{T}_\gamma(x) = 0$, we conclude $F(x) = Q - \frac{\gamma}{2}$, which satisfies the condition $F(x) \leq 1 - \frac{\gamma}{2}$. Consequently, if $\gamma < 2Q = \gamma^{\text{cr}}$, we have $x_\gamma^* = F^{-1}(Q - \frac{\gamma}{2}) > \underline{\xi}$ and $\alpha_\gamma^* = -Vx_\gamma^*$. Observe that $\mathcal{T}_\gamma(x)$ is a strictly convex univariate function in x on the interior of Ω because $(W+U) > 0$ (which shows up in its second derivative) by Assumption (A1). Hence, there exists a unique optimal solution on Ω .

Now, suppose $x = \underline{\xi}$. Then, for all $\xi \in \Omega$, we have $h(x, \xi) = -Vx$. This implies $\alpha = -Vx$ and $\mathfrak{E}(x, \alpha) = 0$. Substituting $\alpha = -Vx$, we obtain $\mathcal{T}_\gamma(x) = \frac{\gamma}{2} \sup_{\xi \in \Omega} h(x, \xi) - \left(1 - \frac{\gamma}{2}\right)Vx$.

By setting $0 \in \partial_x \mathcal{T}_\gamma(x)$, we obtain $-Q \leq 0 \leq -Q + \frac{\gamma}{2}$. Hence, for $\gamma \geq 2Q$, we have $x_\gamma^* = \underline{\xi}$ and $\alpha_\gamma^* = -Vx_\gamma^*$, which is a unique optimal solution on Ω .

- (iii) Condition (C3). The proof is similar to that under (C2). We skip the proof for brevity.
- (iv) At $\gamma = 0$, (DRNV-V) reduces to (Risk Neutral). It is well known that $x^{\text{neut}} = F^{-1}(Q)$ is optimal to (Risk Neutral). Optimality of $x_0^* = F^{-1}(Q)$ can be verified from (3). Consider $f_\gamma(x)$ defined in (3) for $\gamma = 0$. Observe that $f_\gamma''(x) > 0$ for any $x \in \Omega$; hence there is a unique optimal solution to (DRNV-V) at $\gamma = 0$.

At $\gamma = 2$, (DRNV-V) reduces to (Robust), where $x_2^* = x^{\text{rob}}$. By Lemma 2, $f_2(x) = \sup_{\xi \in \Omega} h(x, \xi)$ decreases as x increases to x^{rob} , and then increases after x^{rob} . Furthermore, $0 \in \partial_x f_2(x^{\text{rob}})$. Hence, there exists a unique optimal solution to (DRNV-V) at $\gamma = 2$. ■

A.1.3 Proof of Theorem 2

Proof of Theorem 2. For ease of exposition, let us define the following sets:

$$\Pi_\gamma := \{\xi \in \Omega : \xi \leq x_\gamma^*\}, \quad (31)$$

$$\Lambda_\gamma := \left\{ \xi \in \Omega : h(x_\gamma^*, \xi) = \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)] \right\}, \quad (32)$$

$$\Xi_\gamma := \left\{ \xi \in \Omega : h(x_\gamma^*, \xi) > \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)] \right\}, \quad (33)$$

$$M_\gamma := \left\{ \xi \in \Omega : h(x_\gamma^*, \xi) = \sup_{\xi \in \Omega} h(x_\gamma^*, \xi) \right\}, \quad (34)$$

$$\Gamma_n := \left\{ \xi \in \Omega : h(x_\gamma^*, \xi) \geq \sup_{\xi \in \Omega} h(x_\gamma^*, \xi) - \frac{1}{n} \right\}, \quad n \in \mathbb{N}, \quad (35)$$

where we suppressed the dependence on γ from Γ_n for simplicity.

- (i) Consider any of the Conditions (C1), (C2b), and (C3b) with $0 < \gamma < 2$, or Conditions (C2a) or (C3a) with $0 < \gamma < \gamma^{\text{cr}}$. As discussed in Remark 5, in these cases, M_γ contains only one or two elements ($\{\underline{\xi}, \bar{\xi}\}$, $\{\underline{\xi}\}$, or $\{\bar{\xi}\}$). We first show p_γ^* defined in (9) forms a probability

distribution $\mathbb{P}_\gamma^*$. Clearly, $p_\gamma^*(\xi) \geq 0$ for all $\xi \in \Omega$, and we have

$$\int_{\Omega} p_\gamma^*(\xi) d\xi = \int_{\Lambda_\gamma} \sigma_\gamma p_0(\xi) d\xi + \int_{\Xi_\gamma} p_0(\xi) d\xi + \sum_{M_\gamma} \kappa_\gamma = \sigma_\gamma \mathbb{P}_0 \{ \Lambda_\gamma \} + \mathbb{P}_0 \{ \Xi_\gamma \} + \frac{\gamma}{2} = 1,$$

where the last equality is true because under (C1), (C2b), and (C3b) with $0 < \gamma < 2$, $\mathbb{P}_0 \{ \Lambda_\gamma \} = 0$ and $\mathbb{P}_0 \{ \Xi_\gamma \} = 1 - \frac{\gamma}{2}$, and under either (C2a) or (C3a) with $0 < \gamma < \gamma^{\text{cr}}$, $\sigma_\gamma \mathbb{P}_0 \{ \Lambda_\gamma \} = \mathbb{P}_0 \{ \Xi_\gamma^c \} - \frac{\gamma}{2}$ (see the proof of Theorem 1 for more details). Next, we show $\mathbb{P}_\gamma^*$ associated with density p_γ^* defined in (9) attains the worst-case expected value in (DRNV-V) at x_γ^* . We have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_\gamma^*} [h(x_\gamma^*, \xi)] &= \int_{\Omega} h(x_\gamma^*, \xi) p_\gamma^*(\xi) d\xi = \int_{\Lambda_\gamma} h(x_\gamma^*, \xi) \sigma_\gamma p_0(\xi) d\xi + \int_{\Xi_\gamma} h(x_\gamma^*, \xi) p_0(\xi) d\xi + \sum_{M_\gamma} \kappa_\gamma \sup_{\xi \in \Omega} h(x_\gamma^*, \xi) \\ &= \left[\mathbb{P}_0 \{ \Xi_\gamma^c \} - \frac{\gamma}{2} \right] \text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)] + \int_{\Xi_\gamma} h(x_\gamma^*, \xi) p_0(\xi) d\xi + \frac{\gamma}{2} \sup_{\xi \in \Omega} h(x_\gamma^*, \xi), \\ &= \left(1 - \frac{\gamma}{2} \right) \text{CVaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)] + \frac{\gamma}{2} \sup_{\xi \in \Omega} h(x_\gamma^*, \xi), \end{aligned}$$

where the last equality is due to Rockafellar and Uryasev (2002, Proposition 6) and is equal to $\sup_{\mathbb{P} \in \mathcal{P}_\gamma} \mathbb{E}_{\mathbb{P}} [h(x_\gamma^*, \xi)]$ by (3).

Now, we show there exists a sequence of probability distributions $\mathbb{P}_n \in \mathcal{P}_\gamma$ such that (1) $\mathbb{P}_n \Rightarrow \mathbb{P}_\gamma^*$ as $n \rightarrow \infty$, where $\Rightarrow$ denotes the convergence in distribution, and (2) $\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbb{P}_n} [h(x_\gamma^*, \xi)] = \mathbb{E}_{\mathbb{P}_\gamma^*} [h(x_\gamma^*, \xi)]$. We construct a sequence of probability distributions $\mathbb{P}_n \in \mathcal{P}_\gamma$, $n \in \mathbb{N}$, whose density functions $p_n := p^1 + p^2 + p_n^3$, $n \in \mathbb{N}$, are as follows

$$p^1(\xi) = \begin{cases} p_0(\xi), & \text{if } \xi \in \Xi_\gamma, \\ 0, & \text{otherwise (o.w.)}, \end{cases} \quad p^2(\xi) = \begin{cases} \sigma_\gamma p_0(\xi), & \text{if } \xi \in \Lambda_\gamma, \\ 0, & \text{o.w.}, \end{cases} \quad p_n^3(\xi) = \begin{cases} \frac{\delta_\xi}{\nu(\Gamma_n \cap \Pi_\gamma)}, & \text{if } \xi \in \Gamma_n \cap \Pi_\gamma, \\ \frac{\delta_\xi}{\nu(\Gamma_n \cap \Pi_\gamma^c)}, & \text{if } \xi \in \Gamma_n \cap \Pi_\gamma^c, \\ 0, & \text{o.w.}, \end{cases}$$

where $\delta_\xi = \kappa_\xi$ on M_γ and $\delta_\xi = 0$ otherwise. Similar to Jiang and Guan (2015, Proposition 1), it is easy to show $\mathbb{P}_n \in \mathcal{P}_\gamma$, $n \in \mathbb{N}$.

Let F_γ^* and F_n , $n \in \mathbb{N}$, be the cdf associated with $\mathbb{P}_\gamma^*$ and $\mathbb{P}_n$, $n \in \mathbb{N}$. Also, let p^1 , p^2 , and

p_n^3 be the density functions associated with the absolutely continuous measures $\mathbb{P}^1$, $\mathbb{P}^2$, and $\mathbb{P}_n^3$, $n \in \mathbb{N}$, respectively. Note that $\mathbb{P}_\gamma^*$ is discontinuous on M_γ . In order to show $\mathbb{P}_n \Rightarrow \mathbb{P}_\gamma^*$ as $n \rightarrow \infty$, we need to show $F_n(b) \rightarrow F_\gamma^*(b)$ as $n \rightarrow \infty$ for all $b \in \Omega \setminus M_\gamma$. Let B denote the event $[\xi \leq b]$, for $b \in \Omega \setminus M_\gamma$. Then, we have

$$\begin{aligned}
F_n(b) &= \mathbb{P}_n \{B\} = \mathbb{P}^1 \{B\} + \mathbb{P}^2 \{B\} + \mathbb{P}_n^3 \{B\} \\
&= \mathbb{P}^1 \{B\} + \mathbb{P}^2 \{B\} + \mathbb{P}_n^3 \{B \cap \Pi_\gamma\} + \mathbb{P}_n^3 \{B \cap \Pi_\gamma^c\} \\
&= \mathbb{P}_\gamma^* \{B\} - \delta_\xi + \mathbb{P}_n^3 \{B \cap \Gamma_n \cap \Pi_\gamma\} + \mathbb{P}_n^3 \{B \cap \Gamma_n \cap \Pi_\gamma^c\} \\
&= F_\gamma^*(b) - \delta_\xi + \mathbb{P}_n^3 \{B \cap \Gamma_n \cap \Pi_\gamma\} + \mathbb{P}_n^3 \{B \cap \Gamma_n \cap \Pi_\gamma^c\}. \tag{36}
\end{aligned}$$

For a fixed $b \in \Omega \setminus M_\gamma$, it is easy to verify that there exists $N_1 := N(b, \text{Condition})$ such that for all $n \geq N_1$, $B \cap \Gamma_n \cap \Pi_\gamma^c = \emptyset$. Therefore, the last term in (36) becomes zero. Also, there exists $N_2 := N(b, \text{Condition})$ such that for all $n \geq N_2$, (i) if $M_\gamma = \{\xi, \bar{\xi}\}$, $B \cap \Gamma_n \cap \Pi_\gamma = \Gamma_n \cap \Pi_\gamma$, so $\mathbb{P}_n^3 \{B \cap \Gamma_n \cap \Pi_\gamma\} = \delta_\xi \frac{\nu(B \cap \Gamma_n \cap \Pi_\gamma)}{\nu(\Gamma_n \cap \Pi_\gamma)} = \delta_\xi$; (ii) if $M_\gamma = \{\xi\}$, $B \cap \Gamma_n \cap \Pi_\gamma = \Gamma_n \cap \Pi_\gamma = \Gamma_n$ and again $\mathbb{P}_n^3 \{B \cap \Gamma_n \cap \Pi_\gamma\} = \delta_\xi$; and (iii) if $M_\gamma = \{\bar{\xi}\}$, $B \cap \Gamma_n \cap \Pi_\gamma = \emptyset$, so $\mathbb{P}_n^3 \{B \cap \Gamma_n \cap \Pi_\gamma\} = 0$ and δ_ξ are both zero. As a result, for all $n \geq N = \max\{N_1, N_2\}$, $F_n(b) = F_\gamma^*(b)$. Thus, $\lim_{n \rightarrow \infty} F_n(b) = F_\gamma^*(b)$ for $b \in \Omega \setminus M_\gamma$ and this concludes $\mathbb{P}_n \Rightarrow \mathbb{P}_\gamma^*$ as $n \rightarrow \infty$.

Now, we show $\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbb{P}_n} [h(x_\gamma^*, \xi)] = \mathbb{E}_{\mathbb{P}_\gamma^*} [h(x_\gamma^*, \xi)]$. Define random variables Y_γ^* and Y_n , $n \in \mathbb{N}$ with distribution $\mathbb{P}_\gamma^*$ and $\mathbb{P}_n$, respectively. Then, by the continuous mapping theorem, $h(x_\gamma^*, Y_n) \Rightarrow h(x_\gamma^*, Y_\gamma^*)$ as $n \rightarrow \infty$. By Billingsley (1999, Theorem 3.5), if $h(x_\gamma^*, Y_n) \Rightarrow h(x_\gamma^*, Y_\gamma^*)$ as $n \rightarrow \infty$ and $h(x_\gamma^*, Y_n)$, $n \in \mathbb{N}$ are uniformly integrable, then $\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbb{P}_n} [h(x_\gamma^*, \xi)] = \mathbb{E}_{\mathbb{P}_\gamma^*} [h(x_\gamma^*, \xi)]$. To prove uniform integrability of $h(x_\gamma^*, Y_n)$, $n \in \mathbb{N}$, we show $\sup_{n \in \mathbb{N}} \mathbb{E}_{\mathbb{P}_n} [|h(x_\gamma^*, \xi)|^{1+\delta}] < \infty$ for some $\delta > 0$. Choose $\delta > 0$. For $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_n} [|h(x_\gamma^*, \xi)|^{1+\delta}] &= \int_{\Xi_\gamma} |h(x_\gamma^*, \xi)|^{1+\delta} p^1(\xi) d\xi + \int_{A_\gamma} |h(x_\gamma^*, \xi)|^{1+\delta} p^2(\xi) d\xi + \int_{\Gamma_n} |h(x_\gamma^*, \xi)|^{1+\delta} p_n^3(\xi) d\xi \\
&\leq \mathbb{P}^1 \{\Xi_\gamma\} \sup_{\xi \in \Xi_\gamma} |h(x_\gamma^*, \xi)|^{1+\delta} + \mathbb{P}^2 \{A_\gamma\} |\text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)]|^{1+\delta} + \mathbb{P}_n^3 \{\Gamma_n\} \sup_{\xi \in \Gamma_n} |h(x_\gamma^*, \xi)|^{1+\delta}.
\end{aligned}$$

Note that $\text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)] > -\infty$ because $\gamma > 0$. Above, the first term is bounded because

$|h(x_\gamma^*, \xi)| \leq \max\{|\sup_{\xi \in \Omega} h(x_\gamma^*, \xi)|, |\text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]|\} < \infty$ on Ξ_γ and the second term is bounded because $|\text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]| < \infty$. To see why the third term is bounded, first observe $\mathbb{P}_n^3\{\Gamma_n\} = \frac{\gamma}{2} < \infty$ and $|h(x_\gamma^*, \xi)| \leq \max\{|\sup_{\xi \in \Omega} h(x_\gamma^*, \xi) - \frac{1}{n}|, |\sup_{\xi \in \Omega} h(x_\gamma^*, \xi)|\}$ on Γ_n . If the right-hand side of the above inequality is $|\sup_{\xi \in \Omega} h(x_\gamma^*, \xi)|$ for all $n \in \mathbb{N}$ (a sufficient condition is $\sup_{\xi \in \Omega} h(x_\gamma^*, \xi) \geq \frac{1}{2}$), then $\sup_{n \in \mathbb{N}} \sup_{\xi \in \Gamma_n} |h(x_\gamma^*, \xi)|^{1+\delta} \leq |\sup_{\xi \in \Omega} h(x_\gamma^*, \xi)|^{1+\delta} < \infty$. Otherwise, the right-hand side is non-increasing in n ; hence $\sup_{n \in \mathbb{N}} \sup_{\xi \in \Gamma_n} |h(x_\gamma^*, \xi)|^{1+\delta} \leq |\sup_{\xi \in \Omega} h(x_\gamma^*, \xi) - 1|^{1+\delta} < \infty$. Consequently, $\sup_{n \in \mathbb{N}} \mathbb{E}_{\mathbb{P}_n}[|h(x_\gamma^*, \xi)|^{1+\delta}] < \infty$ and $h(x_\gamma^*, Y_n)$, $n \in \mathbb{N}$ are uniformly integrable. This completes the proof.

Now, consider Conditions (C1), (C2b), and (C3b) with $\gamma = 2$. With a similar argument as above, we can show p_γ^* defined in (11) forms a probability distribution $\mathbb{P}_\gamma^*$ and attains the worst-case expected value in (DRNV-V) at x_γ^* . We can also construct a sequence of probability distributions $\mathbb{P}_n \in \mathcal{P}_\gamma$, $n \in \mathbb{N}$, with density function $p_n = p_n^3$, $n \in \mathbb{N}$. Similar to above we can show that (i) $\mathbb{P}_n \Rightarrow \mathbb{P}_\gamma^*$ as $n \rightarrow \infty$ and (ii) $\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbb{P}_n}[h(x_\gamma^*, \xi)] = \mathbb{E}_{\mathbb{P}_\gamma^*}[h(x_\gamma^*, \xi)]$.

- (ii) Under Condition (C2a) and (C3a) with $\gamma \geq \gamma^{\text{cr}}$, $h(x_\gamma^*, \xi) = -Vx_\gamma^*$ for all $\xi \in \Omega$. So, all probability distributions in the ambiguity set $\mathcal{P}_\gamma$, including $\mathbb{P}_0$, are optimal to (DRNV-V). ■

A.2 Proofs of Section 4 (Characterization of Maximal Effective Subsets)

Our aim in this section is to prove Theorems 3 and 4, and Corollary 1. Before presenting the proof of Theorem 3, we state some intermediate results in Sections A.2.1–A.2.3.

For a subset $\mathcal{F} \subset \Omega$, as we shall see shortly in Proposition 1, either the ambiguity set $\mathcal{P}_\gamma^A(\mathcal{F})$ of the assessment problem is empty, or that problem is feasible. Let $\mathcal{F} \subset \Omega$ be such that $\mathcal{P}_\gamma^A(\mathcal{F})$ is non-empty, and let $x_\gamma^*(\mathcal{F})$ denote an optimal solution to (12). For such $\mathcal{F}$, we use Definition 1 in two ways: (1) with the knowledge of $x_\gamma^*(\mathcal{F})$ and (2) without the knowledge of $x_\gamma^*(\mathcal{F})$. In the first proof-technique, we evaluate $f_\gamma^A(x_\gamma^*(\mathcal{F}); \mathcal{F})$ and $f_\gamma(x_\gamma^*)$. If these two quantities are equal, we conclude $\mathcal{F}$ is ineffective (see for instance, Proposition 3). Otherwise, $\mathcal{F}$ is effective. In the second proof-technique, we evaluate $f_\gamma^A(x_\gamma^*; \mathcal{F})$ and $f_\gamma(x_\gamma^*)$. If $f_\gamma^A(x_\gamma^*; \mathcal{F}) < f_\gamma(x_\gamma^*)$, we conclude $\mathcal{F}$ is effective because $f_\gamma^A(x_\gamma^*(\mathcal{F}); \mathcal{F}) \leq f_\gamma^A(x_\gamma^*; \mathcal{F})$ (see for instance, Propositions 4 and 6). Otherwise, through convex analysis arguments on $f_\gamma^A(x; \mathcal{F})$ we show x_γ^* is also optimal to the assessment problem of

$\mathcal{F}$; hence $\mathcal{F}$ is ineffective (see Proposition 5).

A.2.1 Notation and Preliminary Results for the Assessment Problem

Before we proceed, let us define some notation. Let $\text{ess sup}_{\mathcal{F}^c} h(x, \xi) := \inf\{a \in \mathbb{R} : \nu(\{\xi \in \mathcal{F}^c : h(x, \xi) > a\}) = 0\}$. We use $\underline{\xi}_{|\mathcal{F}^c}$ and $\bar{\xi}_{|\mathcal{F}^c}$ to denote $\inf\{\hat{\xi} : \hat{\xi} \in \mathcal{F}^c\}$ and $\sup\{\hat{\xi} : \hat{\xi} \in \mathcal{F}^c\}$, respectively. We also use $p_{0|\mathcal{F}^c}$ and $F(\cdot|\mathcal{F}^c)$ to denote the conditional versions of p_0 and F on $\mathcal{F}^c$, respectively. Finally as before, for any $B \subset \mathbb{R}$ and a fixed $x \in \mathbb{X}$, we use $[h(x, \xi) \in B]$ as shorthand notation for the set $\{\xi \in \Omega : h(x, \xi) \in B\}$.

Both proof-techniques discussed above rely on the assessment problem. So, we begin by presenting an analogous formulation to the inner maximization problem in (12) in Proposition 1.

Proposition 1. *Consider a fixed $x \in \mathbb{X}$ and a subset $\mathcal{F} \subset \Omega$. Then, the Lagrangian dual formulation of the inner maximization problem in (12) can be written as*

$$f_\gamma^A(x; \mathcal{F}) = \begin{cases} \frac{\gamma}{2} \text{ess sup}_{\mathcal{F}^c} h(x, \xi) + \left(1 - \frac{\gamma}{2}\right) \mathbb{E}_{\mathbb{P}_0} [h(x, \xi)|\mathcal{F}^c], & \text{if } \gamma = 2\mathbb{P}_0\{\mathcal{F}\}, \\ \frac{\gamma}{2} \text{ess sup}_{\mathcal{F}^c} h(x, \xi) + \left(1 - \frac{\gamma}{2}\right) \text{CVaR}_{\gamma'} [h(x, \xi)|\mathcal{F}^c], & \text{if } 2\mathbb{P}_0\{\mathcal{F}\} < \gamma < 2, \\ \text{ess sup}_{\mathcal{F}^c} h(x, \xi), & \text{if } \gamma \geq 2, \end{cases} \quad (37)$$

where $\mathbb{E}_{\mathbb{P}_0} [h(x, \xi)|\mathcal{F}^c]$ is the (conditional) expectation of $h(x, \xi)$ and $\text{CVaR}_{\gamma'} [h(x, \xi)|\mathcal{F}^c]$ is the (conditional) CVaR of $h(x, \xi)$ at level $\gamma' := \frac{\frac{\gamma}{2} - \mathbb{P}_0\{\mathcal{F}\}}{1 - \mathbb{P}_0\{\mathcal{F}\}}$ with respect to the conditional distribution $p_{0|\mathcal{F}^c}$. If $0 < \gamma < 2\mathbb{P}_0\{\mathcal{F}\}$, then the inner maximization problem in (12) is infeasible.

Observe that $f_\gamma^A(x; \mathcal{F})$ is equivalent to $f_\gamma(x)$ at any $x \in \mathbb{X}$ if $\mathcal{F}$ has a zero Lebesgue measure. Proposition 1 can be proved using a similar method as Jiang and Guan (2015, Theorem 1); hence the proof is relegated to Appendix B. The only two differences are: first, the support Ω should be reduced to $\mathcal{F}^c$ and second, the nominal probability density function p_0 should be adjusted conditionally on $\mathcal{F}^c$. These modifications result in changes in the essential supremum and CVaR terms relative to (3). That is, $\text{ess sup} h(x, \xi)$ changes to $\text{ess sup}_{\mathcal{F}^c} h(x, \xi)$. Also, the CVaR of $h(x, \cdot)$ at level $\frac{\gamma}{2}$ changes to the conditional CVaR of $h(x, \cdot)$ at an adjusted level γ' .

Now, we derive an equivalent formulation to (12). Consider a fixed $x \in \mathbb{X}$ and a subset $\mathcal{F} \subset \Omega$.

As argued before, $\text{ess sup}_{\mathcal{F}^c} h(x, \xi) = \sup_{\xi \in \mathcal{F}^c} h(x, \xi)$ for the cost function defined in (1). Next, for $2\mathbb{P}_0\{\mathcal{F}\} < \gamma < 2$, we can equivalently write the second term in the right-hand side of (37) as

$$\left(1 - \frac{\gamma}{2}\right) \text{CVaR}_{\gamma'}[h(x, \xi)|\mathcal{F}^c] = \min_{\alpha \in \mathbb{R}} \left\{ \left(1 - \frac{\gamma}{2}\right) \alpha + \left(1 - \mathbb{P}_0\{\mathcal{F}\}\right) \int_{\mathcal{F}^c} \left(h(x, \xi) - \alpha\right)_+ p_{0|\mathcal{F}^c}(\xi) d\xi \right\} \quad (38)$$

by an application of [Rockafellar and Uryasev \(2002, Theorem 10\)](#) on the reduced space $\mathcal{F}^c$ and using the fact that $1 - \frac{\gamma}{2} = (1 - \gamma')(1 - \mathbb{P}_0\{\mathcal{F}\})$. Hence, we can write $f_\gamma^\Lambda(x; \mathcal{F})$ in (37) as minimizing a convex function $\mathcal{T}_\gamma(x, \alpha; \mathcal{F})$ in α . Combining these and adding minimization over $x \in \mathbb{X}$, we rewrite (12) as

$$\min_{x \in \mathbb{X}, \alpha \in \mathbb{R}} \mathcal{T}_\gamma(x, \alpha; \mathcal{F}) := \frac{\gamma}{2} \sup_{\xi \in \mathcal{F}^c} h(x, \xi) + \left(1 - \frac{\gamma}{2}\right) \alpha + \left(1 - \mathbb{P}_0\{\mathcal{F}\}\right) \int_{\mathcal{F}^c} \left(h(x, \xi) - \alpha\right)_+ p_{0|\mathcal{F}^c}(\xi) d\xi. \quad (39)$$

for $2\mathbb{P}_0\{\mathcal{F}\} < \gamma < 2$.

A.2.2 Effective/Ineffective Subsets under Conditions (C2a) and (C3a)

Under (C2a) and (C3a), we first present an optimal solution to (12) in Proposition 2. Then, we characterize some ineffective subsets in Proposition 3 using the first proof-technique. Finally, we characterize an effective subset in Proposition 4 using the second proof-technique.

Let $(x_\gamma^*(\mathcal{F}), \alpha_\gamma^*(\mathcal{F}))$ denote an optimal solution to (39) for $2\mathbb{P}_0\{\mathcal{F}\} < \gamma < 2$, and let $x_\gamma^*(\mathcal{F})$ denote an optimal solution to the assessment problem (12) at $\gamma = 2\mathbb{P}_0\{\mathcal{F}\}$ and $\gamma = 2$. Notice that α is only used for $2\mathbb{P}_0\{\mathcal{F}\} < \gamma < 2$ in (39).

Proposition 2. *Consider (12) with cost function defined in (1) and γ^{cr} defined in (6). Suppose Assumption (A1) holds. Consider a subset $\mathcal{F} \subset \Omega$ with $\mathbb{P}_0\{\mathcal{F}\} > 0$.*

(i) *Under Condition (C2a) with $2\mathbb{P}_0\{\mathcal{F}\} < \gamma^{cr}$, the set of optimal solutions to (12) is characterized by:*

$$\begin{cases} F(x_\gamma^*(\mathcal{F})|\mathcal{F}^c) = \frac{Q - \frac{\gamma}{2}}{1 - \mathbb{P}_0\{\mathcal{F}\}}, & \text{for } 2\mathbb{P}_0\{\mathcal{F}\} \leq \gamma < \gamma^{cr}, \\ x_\gamma^*(\mathcal{F}) = \underline{\xi}_{|\mathcal{F}^c}, & \text{for } \gamma^{cr} \leq \gamma \leq 2. \end{cases}$$

Otherwise, if $2\mathbb{P}_0\{\mathcal{F}\} \geq \gamma^{cr}$, then $x_\gamma^(\mathcal{F}) = \underline{\xi}_{|\mathcal{F}^c}$ for $2\mathbb{P}_0\{\mathcal{F}\} \leq \gamma \leq 2$.*

(ii) Under Condition (C3a) with $2\mathbb{P}_0\{\mathcal{F}\} < \gamma^{cr}$, the set of optimal solutions to (12) is characterized by:

$$\begin{cases} F(x_\gamma^*(\mathcal{F})|\mathcal{F}^c) = \frac{Q + \frac{\gamma}{2} - \mathbb{P}_0\{\mathcal{F}\}}{1 - \mathbb{P}_0\{\mathcal{F}\}}, & \text{for } 2\mathbb{P}_0\{\mathcal{F}\} \leq \gamma < \gamma^{cr}, \\ x_\gamma^*(\mathcal{F}) = \bar{\xi}_{|\mathcal{F}^c}, & \text{for } \gamma^{cr} \leq \gamma \leq 2. \end{cases}$$

Otherwise, if $2\mathbb{P}_0\{\mathcal{F}\} \geq \gamma^{cr}$, then $x_\gamma^*(\mathcal{F}) = \bar{\xi}_{|\mathcal{F}^c}$ for $2\mathbb{P}_0\{\mathcal{F}\} \leq \gamma \leq 2$.

For $2\mathbb{P}_0\{\mathcal{F}\} < \gamma < 2$, there exists an optimal solution to (39) that satisfies $\alpha_\gamma^*(\mathcal{F}) = -Vx_\gamma^*(\mathcal{F})$.

With a similar approach as Theorem 1 we can prove the above proposition. We skip the proof for brevity.

Proposition 3. Consider (DRNV-V) and (12) with cost function defined in (1) and γ^{cr} defined in (6). Suppose Assumption (A1) holds. Then, under any of the following settings,

1. Condition (C2a) or (C3a) with $0 < \gamma < \gamma^{cr}$ and $\mathcal{F}$ is a subset of $\left[h(x_\gamma^*, \xi) \leq \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]\right]$ such that $0 < \mathbb{P}_0\{\mathcal{F}\} < \frac{\gamma}{2}$,
2. Condition (C2a) with $\gamma^{cr} \leq \gamma \leq 2$ and $\mathcal{F}$ is a subset of Ω such that $\xi_{|\mathcal{F}^c} = \underline{\xi}$ and $0 < \mathbb{P}_0\{\mathcal{F}\} < \frac{\gamma}{2}$,
3. Condition (C3a) with $\gamma^{cr} \leq \gamma \leq 2$ and $\mathcal{F}$ is a subset of Ω such that $\bar{\xi}_{|\mathcal{F}^c} = \bar{\xi}$ and $0 < \mathbb{P}_0\{\mathcal{F}\} < \frac{\gamma}{2}$,

there exists an optimal solution to (12) that satisfies $x_\gamma^*(\mathcal{F}) = x_\gamma^*$ and $f_\gamma^A(x_\gamma^*(\mathcal{F}); \mathcal{F}) = f_\gamma(x_\gamma^*)$. Consequently, the corresponding $\mathcal{F}$ is ineffective for (DRNV-V).

Proof. We present the proof under (C2a). The proof under (C3a) is similar and, we skip it for brevity. Under (C2a), we first show that there exists an optimal solution to (12) such that $x_\gamma^*(\mathcal{F}) = x_\gamma^*$. Then, we show that $f_\gamma^A(x_\gamma^*(\mathcal{F}); \mathcal{F}) = f_\gamma(x_\gamma^*)$ and conclude $\mathcal{F}$ is ineffective.

- Suppose $0 < \gamma < \gamma^{cr}$ and $\mathcal{F}$ is a subset of $\mathcal{G} := \left[h(x_\gamma^*, \xi) \leq \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]\right]$ with $0 < \mathbb{P}_0\{\mathcal{F}\} < \frac{\gamma}{2}$. Note that when $[\xi \geq x_\gamma^*]$, $h(x_\gamma^*, \xi) = -Vx$. Otherwise, when $[\xi < x_\gamma^*]$, $h(x_\gamma^*, \xi) = Wx_\gamma^* - (W + V)\xi > -Vx_\gamma^*$. Thus, by replacing the expression of $h(x_\gamma^*, \xi)$ on $[\xi \geq x_\gamma^*]$ and $[\xi < x_\gamma^*]$, and setting $\alpha_\gamma^* = \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)] = -Vx_\gamma^*$ from Theorem 1, we can equivalently write $\mathcal{G} = [\xi \geq x_\gamma^*]$. Also, observe that $\mathbb{P}_0\{\mathcal{G}\} > \frac{\gamma}{2}$ in this case (cf. the proof of Theorem 1).

These imply $\underline{x}_{|\mathcal{F}^c} = \underline{x}$ and $\bar{x}_{|\mathcal{F}^c} > x_\gamma^*$. On the other hand, $F(x_\gamma^*|\mathcal{F}^c) = \frac{F(x_\gamma^*)}{1-\mathbb{P}_0\{\mathcal{F}\}} = \frac{Q-\frac{\gamma}{2}}{1-\mathbb{P}_0\{\mathcal{F}\}} = F(x_\gamma^*(\mathcal{F})|\mathcal{F}^c)$ by Theorem 1 and Proposition 2. Therefore, by Proposition 2, x_γ^* is an optimal solution to (12).

Now, we show that $f_\gamma^A(x_\gamma^*(\mathcal{F}); \mathcal{F}) = f_\gamma(x_\gamma^*)$. First, because x_γ^* is an optimal solution to (12), by Proposition 2 and Theorem 1, there exists an optimal solution such that $(x^*(\mathcal{F}), \alpha_\gamma^*(\mathcal{F})) = (x_\gamma^*, \alpha_\gamma^*)$. Thus, by (38) we have

$$\begin{aligned} \left(1 - \frac{\gamma}{2}\right) \text{CVaR}_{\gamma'} [h(x_\gamma^*(\mathcal{F}), \xi) | \mathcal{F}^c] &= \left(1 - \frac{\gamma}{2}\right) \alpha_\gamma^*(\mathcal{F}) + \int_{\mathcal{F}^c} \left(h(x_\gamma^*(\mathcal{F}), \xi) - \alpha_\gamma^*(\mathcal{F})\right)_+ p_0(\xi) d\xi \\ &= \left(1 - \frac{\gamma}{2}\right) \alpha_\gamma^* + \int_{\mathcal{F}^c} \left(h(x_\gamma^*, \xi) - \alpha_\gamma^*\right)_+ p_0(\xi) d\xi \\ &= \left(1 - \frac{\gamma}{2}\right) \alpha_\gamma^* + \int_{\Omega} \left(h(x_\gamma^*, \xi) - \alpha_\gamma^*\right)_+ p_0(\xi) d\xi \\ &= \left(1 - \frac{\gamma}{2}\right) \text{CVaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)], \end{aligned} \quad (40)$$

where (40) is true by the definition of $\mathcal{F}$ and the fact that $\alpha_\gamma^* = \text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)]$. On the other hand, $\sup_{\xi \in \mathcal{F}^c} h(x_\gamma^*(\mathcal{F}), \xi) = \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)$. Consequently, $f_\gamma^A(x_\gamma^*(\mathcal{F}); \mathcal{F}) = f_\gamma(x_\gamma^*)$. Finally, an application of Definition 1 concludes $\mathcal{F}$ is ineffective.

- If $\gamma^{\text{cr}} \leq \gamma \leq 2$ and $\mathcal{F}$ is any subset of Ω such that $\underline{x}_{|\mathcal{F}^c} = \underline{x}$ and $0 < \mathbb{P}_0\{\mathcal{F}\} < \frac{\gamma}{2}$, then there exists an optimal solution such that $x_\gamma^*(\mathcal{F}) = \underline{x}$ by Proposition 2. Also, we have $x_\gamma^* = \underline{x}$ by Theorem 1. On the other hand, $h(x_\gamma^*(\mathcal{F}), \xi) = -V\underline{x}$ for all $\xi \in \mathcal{F}^c$ and $h(x_\gamma^*, \xi) = -V\underline{x}$ for all $\xi \in \Omega$. So, it is easy to verify $f_\gamma^A(x_\gamma^*(\mathcal{F}); \mathcal{F}) = f_\gamma(x_\gamma^*)$; thus, $\mathcal{F}$ is ineffective. ■

Proposition 4. Consider (DRNV-V) with cost function defined in (1). Suppose Assumption (A1) holds. Under either Condition (C2a) or (C3a) with $0 < \gamma < \gamma^{\text{cr}}$ and $\mathcal{F} = \left[h(x_\gamma^*, \xi) > \text{VaR}_{\frac{\gamma}{2}} [h(x_\gamma^*, \xi)]\right]$, we have

- (i) $\mathcal{F}$ and any of its subsets that belong to $\mathfrak{S}$ are effective for (DRNV-V),
- (ii) any singleton subset of $\mathcal{F}$ is effective-in-limit for (DRNV-V).

Proof. Choose $0 < \gamma < \gamma^{\text{cr}}$ and let $\mathcal{G}$ be a subset of $\mathcal{F}$ that belongs to $\mathfrak{S}$. First, suppose $0 < \mathbb{P}_0\{\mathcal{G}\} \leq \min\{\mathbb{P}_0\{\mathcal{F}\}, \frac{\gamma}{2}\}$; so, the inner maximization problem in the assessment problem of $\mathcal{G}$ is feasible by Proposition 1. Notice that the assessment problem of $\mathcal{G}$ is a relaxation

of (DRNV-V). This, combined with the suboptimality of x_γ^* to the assessment problem, imply $f_\gamma^A(x_\gamma^*(\mathcal{G}); \mathcal{G}) \leq f_\gamma^A(x_\gamma^*; \mathcal{G}) \leq f_\gamma(x_\gamma^*)$. Consequently, $f_\gamma(x_\gamma^*) - f_\gamma^A(x_\gamma^*; \mathcal{G})$ gives a lower bound on $f_\gamma(x_\gamma^*) - f_\gamma^A(x_\gamma^*(\mathcal{G}); \mathcal{G})$. If this lower bound is positive, then $\mathcal{G}$ is effective. To check this lower bound for the subset $\mathcal{G}$, consider a fixed $x \in \mathbb{X}$. Let λ and μ denote Lagrange multipliers for the first and second constraints in (2), respectively. By the proof of Jiang and Guan (2015, Theorem 1), the Lagrangian dual formulation of the inner maximization problem in (DRNV-V) at x can be written as $\min_{(\mu, \lambda) \in \mathcal{D}(x)} \mathcal{L}_\gamma(\mu, \lambda; x) := \mu - \lambda(1 - \gamma) + \int_\Omega \left(h(x, \xi) - \mu + \lambda \right)_+ p_0(\xi) d\xi$, where $\mathcal{D}(x) := \{(\mu, \lambda) : \lambda \geq 0, \mu + \lambda \geq \sup_{\xi \in \Omega} h(x, \xi)\}$. Similarly, let λ and μ denote Lagrange multipliers for the first and second constraints in (13), respectively. Then, by the proof of Proposition 1, the Lagrangian dual formulation of the inner maximization problem in (12) at x can be written as $\min_{(\mu, \lambda) \in \mathcal{D}^A(x; \mathcal{G})} \mathcal{L}_\gamma^A(\mu, \lambda; x, \mathcal{G}) := \mu - \lambda(1 - \gamma) + \int_{\mathcal{G}^c} \left(h(x, \xi) - \mu + \lambda \right)_+ p_0(\xi) d\xi$, where $\mathcal{D}^A(x; \mathcal{G}) := \{(\mu, \lambda) : \lambda \geq 0, \mu + \lambda \geq \sup_{\xi \in \mathcal{G}^c} h(x, \xi)\}$. Note that $\mathcal{D}(x) \subseteq \mathcal{D}^A(x; \mathcal{G})$ because $\sup_{\xi \in \mathcal{G}^c} h(x, \xi) \leq \sup_{\xi \in \Omega} h(x, \xi)$. Suppose (μ^*, λ^*) are the optimal dual variables to the inner maximization problem in (DRNV-V) at x_γ^* . Thus, $(\mu^*, \lambda^*) \in \mathcal{D}^A(x_\gamma^*; \mathcal{G})$ and we have

$$\begin{aligned} f_\gamma(x_\gamma^*) - f_\gamma^A(x_\gamma^*; \mathcal{G}) &\geq \mathcal{L}_\gamma(\mu^*, \lambda^*; x_\gamma^*) - \mathcal{L}_\gamma^A(\mu^*, \lambda^*; x_\gamma^*, \mathcal{G}) \\ &= \int_{\mathcal{G}} \left(h(x_\gamma^*, \xi) - \mu^* + \lambda^* \right)_+ p_0(\xi) d\xi \\ &= \int_{\mathcal{G}} \left(h(x_\gamma^*, \xi) - \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)] \right)_+ p_0(\xi) d\xi, \end{aligned}$$

where the last equality comes from the fact that $\mu^* - \lambda^* = \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]$ from the proof of Jiang and Guan (2015, Theorem 1). Observe that the lower bound is positive because $\mathcal{G}$ is a subset of $\mathcal{F}$.

Now, suppose $\mathbb{P}_0\{\mathcal{G}\} > \frac{\gamma}{2}$ (this case is possible because under (C2a) or (C3a) with $0 < \gamma < \frac{\gamma^{\text{cr}}}{2}$ we must have $\mathbb{P}_0\{\mathcal{F}\} > \frac{\gamma}{2}$, cf. the proof of Theorem 1). So, such $\mathcal{G}$ is effective by Definition 1. As a result, in either case, $\mathcal{F}$ and any of its subsets that belong to $\mathfrak{S}$ are effective. On the other hand, any singleton subset of $\mathcal{F}$ can be written as the countably infinite intersection of subsets of $\mathcal{F}$ that belongs to $\mathfrak{S}$. That is, any singleton subset of $\mathcal{F}$ is effective-in-limit. $\blacksquare$

A.2.3 Effective/Ineffective Subsets under Conditions (C1), (C2b), and (C3b)

Under any of the Conditions (C1), (C2b), and (C3b), we use the second proof-technique. We first characterize an ineffective subset in Proposition 5 using the convex analysis arguments on $f_\gamma^A(x; \mathcal{F})$. Then, we characterize an effective subset in Proposition 6 by comparing $f_\gamma^A(x_\gamma^*; \mathcal{F})$ and $f_\gamma(x_\gamma^*)$.

In the proof of Proposition 5, we evaluate $\sup_{\xi \in \Omega} h(x_\gamma^*, \xi)$, $\text{CVaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]$, $f_\gamma(x_\gamma^*)$, and their subdifferentials and compare them to $\sup_{\xi \in \mathcal{F}^c} h(x_\gamma^*, \xi)$, $\text{CVaR}_{\gamma'}[h(x_\gamma^*, \xi)|\mathcal{F}^c]$, $f_\gamma^A(x_\gamma^*; \mathcal{F})$, and their subdifferentials, respectively, to check the optimality of x_γ^* to the assessment problem of $\mathcal{F}$.

Proposition 5. *Consider (DRNV-V) with cost function defined in (1). Suppose Assumption (A1) holds. Then, under any of the Conditions (C1), (C2b), and (C3b)*

1. *with $0 < \gamma < 2$ and $\mathcal{F} = [h(x_\gamma^*, \xi) < \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]]$, or*
2. *with $\gamma = 2$ and $\mathcal{F} = [h(x_\gamma^*, \xi) < \sup_{\xi \in \Omega} h(x_\gamma^*, \xi) - \frac{1}{n}]$, $n \in \mathbb{N}$,*

there exists an optimal solution to (12) that satisfies $x_\gamma^(\mathcal{F}) = x_\gamma^*$ and $f_\gamma^A(x_\gamma^*(\mathcal{F}); \mathcal{F}) = f_\gamma(x_\gamma^*)$. Consequently, the corresponding $\mathcal{F}$ is ineffective for (DRNV-V).*

Proof. Suppose $0 < \gamma < 2$ and let $\mathcal{F} = [h(x_\gamma^*, \xi) < \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]]$. So, $\mathbb{P}_0\{\mathcal{F}\} = \frac{\gamma}{2}$ by definition of $\text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]$ and the fact that $\mathbb{P}_0\{h(x_\gamma^*, \xi) = \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]\} = 0$. This implies $\gamma' = 0$ as defined in Proposition 1. Thus, we have

$$\left(1 - \frac{\gamma}{2}\right) \mathbb{E}_{\mathbb{P}_0}[h(x_\gamma^*, \xi)|\mathcal{F}^c] = \int_{\mathcal{F}^c} h(x_\gamma^*, \xi) p_0(\xi) d\xi = \left(1 - \frac{\gamma}{2}\right) \text{CVaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]. \quad (41)$$

Moreover, $\sup_{\xi \in \mathcal{F}^c} h(x_\gamma^*, \xi) = \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)$. So, $f_\gamma^A(x_\gamma^*; \mathcal{F}) = f_\gamma(x_\gamma^*)$. On the other hand, from (41) and the convexity of h in x , we have $\partial_x \mathbb{E}_{\mathbb{P}_0}[h(x_\gamma^*, \xi)|\mathcal{F}^c] = \partial_x \text{CVaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]$. Using the fact that $[h(x_\gamma^*, \xi) = \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)] \subset \mathcal{F}^c$, we have $\partial_x \sup_{\xi \in \mathcal{F}^c} h(x_\gamma^*, \xi) = \partial_x \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)$ according to Hiriart-Urruty and Lemaréchal (2001, Corollary 4.3.2). As a result, we have $\partial_x f_\gamma^A(x_\gamma^*; \mathcal{F}) = \frac{\gamma}{2} \partial_x \sup_{\xi \in \mathcal{F}^c} h(x_\gamma^*, \xi) + \left(1 - \frac{\gamma}{2}\right) \partial_x \mathbb{E}_{\mathbb{P}_0}[h(x_\gamma^*, \xi)|\mathcal{F}^c] = \frac{\gamma}{2} \partial_x \sup_{\xi \in \Omega} h(x_\gamma^*, \xi) + \left(1 - \frac{\gamma}{2}\right) \partial_x \text{CVaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)] = \partial_x f_\gamma(x_\gamma^*)$. By the optimality of x_γ^* to (DRNV-V), there exists a subgradient $s \in \partial_x f_\gamma^A(x_\gamma^*; \mathcal{F}) = \partial_x f_\gamma(x_\gamma^*)$ with $s^\top(x - x_\gamma^*) \geq 0$ for all $x \in \mathbb{X}$. Then, by the convexity of $f_\gamma^A(x; \mathcal{F})$ in x , we have $f_\gamma^A(x; \mathcal{F}) \geq f_\gamma^A(x_\gamma^*; \mathcal{F}) + s^\top(x - x_\gamma^*) \geq f_\gamma^A(x_\gamma^*; \mathcal{F})$ for any $x \in \mathbb{X}$. This implies $x_\gamma^*(\mathcal{F}) = x_\gamma^*$. Thus, $f_\gamma^A(x_\gamma^*(\mathcal{F}); \mathcal{F}) = f_\gamma(x_\gamma^*)$ and $\mathcal{F}$ is ineffective.

Now, suppose $\gamma = 2$ and $\mathcal{F} = [h(x_\gamma^*, \xi) < \sup_{\xi \in \Omega} h(x_\gamma^*, \xi) - \frac{1}{n}]$, $n \in \mathbb{N}$. By Proposition 1, we have $f_\gamma^A(x_\gamma^*; \mathcal{F}) = \sup_{\xi \in \mathcal{F}^c} h(x_\gamma^*, \xi)$ and by (3), we have $f_\gamma(x_\gamma^*) = \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)$. These, combined with $\sup_{\xi \in \mathcal{F}^c} h(x_\gamma^*, \xi) = \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)$, imply that $f_\gamma^A(x_\gamma^*; \mathcal{F}) = f_\gamma(x_\gamma^*)$. On the other hand, similar to above, we obtain $\partial_x \sup_{\xi \in \mathcal{F}^c} h(x_\gamma^*, \xi) = \partial_x \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)$. Now, following a similar argument as that for $0 < \gamma < 2$ completes the proof. $\blacksquare$

Proposition 6. *Consider (DRNV-V) with cost function defined in (1). Suppose Assumption (A1) holds. Under any of the Conditions (C1), (C2b), and (C3b) with $0 < \gamma < 2$ and $\mathcal{F} = [h(x_\gamma^*, \xi) \geq \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]]$, we have*

- (i) $\mathcal{F}$ and any of its subsets that belong to $\mathfrak{S}$ are effective for (DRNV-V),
- (ii) any singleton subset of $\mathcal{F}$ is effective-in-limit for (DRNV-V).

The proof follows similar steps as those in the proof of Proposition 4. We, therefore, skip it for brevity.

A.2.4 Proof of Theorem 3

Proof of Theorem 3. Under any of the Conditions (C1), (C2b), and (C3b), suppose $0 < \gamma < 2$. Let us define $\Phi_\gamma = \Lambda_\gamma \cup \Xi_\gamma$, where Λ_γ and Ξ_γ are defined in (32) and (33), respectively. By Proposition 5, $\mathcal{E}_\gamma \subseteq \Phi_\gamma$ and by Proposition 6, $\mathcal{E}_\gamma \supseteq \Phi_\gamma$. Consequently, $\mathcal{E}_\gamma = \Phi_\gamma$. Now, similar to the argument in the proof of Proposition 3, by replacing the expressions of $h(x_\gamma^*, \xi)$ and setting $\alpha_\gamma^* = \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]$, we can equivalently write $\mathcal{E}_\gamma$ as stated in the theorem.

Now, under any of the Conditions (C1), (C2b), and (C3b), suppose $\gamma = 2$ and consider Γ_n , $n \in \mathbb{N}$, defined in (35). Let $\mathcal{F}_n := \Gamma_n$, $n \in \mathbb{N}$. Observe that $\mathcal{F}_n \in \mathfrak{S}$, $n \in \mathbb{N}$. Also, by (3), we have $f_\gamma(x_\gamma^*) = \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)$ and by Proposition 1, we have $f_\gamma^A(x_\gamma^*; \mathcal{F}_n) = \sup_{\xi \in \mathcal{F}_n^c} h(x_\gamma^*, \xi)$. Thus, for any $n \in \mathbb{N}$,

$$f_\gamma^A(x_\gamma^*(\mathcal{F}_n); \mathcal{F}_n) \leq f_\gamma^A(x_\gamma^*, \mathcal{F}_n) < \sup_{\xi \in \Omega} h(x_\gamma^*, \xi) - \frac{1}{n} < \sup_{\xi \in \Omega} h(x_\gamma^*, \xi) = f_\gamma(x_\gamma^*).$$

This shows $\mathcal{F}_n$ is effective by Definition 1. So, we can conclude M_γ , defined in (35), is effective-in-limit because $M_\gamma = \bigcap_{n \in \mathbb{N}} \mathcal{F}_n = \lim_{n \rightarrow \infty} \mathcal{F}_n$. As a result, $\mathcal{E}_\gamma \supseteq M_\gamma$. On the other hand, by

Proposition 5, $\mathcal{E}_\gamma \subseteq M_\gamma$. Consequently, $\mathcal{E}_\gamma = M_\gamma$. That is, under Condition (C1), $\{\underline{\xi}, \bar{\xi}\}$; under Condition (C2b), $\{\underline{\xi}\}$; and under Condition (C3b), $\{\bar{\xi}\}$ is maximal effective at $\gamma = 2$.

Under Condition (C2a), suppose $0 < \gamma < \gamma^{\text{cr}}$ and consider Ξ_γ . First, similar to the argument in the proof of Proposition 3, we can equivalently write $\Xi_\gamma = [\xi < x_\gamma^*]$. Observe that Ξ_γ does not belong to $\mathfrak{S}$. By Proposition 4, we can conclude $\mathcal{E}_\gamma \supset \Xi_\gamma$. Moreover, because any subset of $[\xi < x_\gamma^*]$ that belongs to $\mathfrak{S}$ is effective, then any subset of $[\xi \leq x_\gamma^*]$ that belongs to $\mathfrak{S}$ is also effective. Now, if we show $\{x_\gamma^*\}$ is effective-in-limit, then this ensures $\mathcal{E}_\gamma \supseteq [\xi \leq x_\gamma^*]$. Let us consider $\mathcal{F}_n := [x_\gamma^* - \frac{1}{n}, x_\gamma^*] \in \mathfrak{S}$, $n \in \mathbb{N}$. Observe that $x_\gamma^* = \bigcap_{n \in \mathbb{N}} \mathcal{F}_n = \lim_{n \rightarrow \infty} \mathcal{F}_n$; so $\{x_\gamma^*\}$ is effective-in-limit. On the other hand, there is no subset $\mathcal{G}$ of $[\xi > x_\gamma^*]$ such that all of its subsets that belong to $\mathfrak{S}$ are effective because by Proposition 3, we can always find an ineffective subset $\mathcal{F} \subset \mathcal{G}$ with $0 < \mathbb{P}_0\{\mathcal{F}\} < \frac{\gamma}{2}$. So, $\mathcal{E}_\gamma \subseteq [\xi \leq x_\gamma^*]$. Consequently, $\mathcal{E}_\gamma = [\xi \leq x_\gamma^*]$.

Now, under Condition (C2a), suppose $\gamma^{\text{cr}} \leq \gamma \leq 2$. Let us consider $\mathcal{F}_n := [\underline{\xi}, a_n] \in \mathfrak{S}$, $n \in \mathbb{N}$, such that $a_n \searrow \underline{\xi}$ and $\mathbb{P}_0\{\mathcal{F}_1\} < \frac{\gamma}{2}$. By Theorem 1, we have $f_\gamma(x_\gamma^*) = -V\underline{\xi}$. Also, by Proposition 2, we have $x_\gamma^*(\mathcal{F}_n) = a_n > \underline{\xi}$. Hence, $f_\gamma^A(x_\gamma^*(\mathcal{F}_n); \mathcal{F}_n) = -Vx_\gamma^*(\mathcal{F}_n) < -V\underline{\xi}$ (note that $V > 0$ in this case). Consequently, $\mathcal{F}_n$ is effective by Definition 1. These imply $\{\underline{\xi}\}$ is effective-in-limit because it can be written as $\bigcap_{n \in \mathbb{N}} \mathcal{F}_n = \lim_{n \rightarrow \infty} \mathcal{F}_n$. On the other hand, there is no subset $\mathcal{G}$ of Ω such that all of its subsets that belong to $\mathfrak{S}$ are effective because by Proposition 3, we can always find an ineffective subset $\mathcal{F} \subset \mathcal{G}$ with $\underline{\xi}_{|\mathcal{F}^c} = \underline{\xi}$ and $0 < \mathbb{P}_0\{\mathcal{F}\} < \frac{\gamma}{2}$. As a result, we conclude $\{\underline{\xi}\}$ is the largest set that is effective-in-limit, and hence it is maximal effective. The proof under (C3a) is similar to that under (C2a) and is skipped for brevity. ■

A.2.5 Proof of Theorem 4

Proof of Theorem 4. (i) According to Remark 7, under any of the Conditions (C2b) and (C3b), $\mathcal{E}_\gamma$ is decreasing in γ . Additionally, under any of the Conditions (C1), (C2a), and (C3a), $\mathcal{E}_\gamma$ is decreasing in γ if $\gamma < \gamma^{\text{cr}}$. Under (C1), if $\gamma \geq \gamma^{\text{cr}}$, $\alpha_\gamma^* = \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]$ increases as γ increases. On the other hand, if $\gamma \geq \gamma^{\text{cr}}$, $x_\gamma^* = x^{\text{rob}}$ by Theorem 1. Then, Theorem 3(ii) implies if $\gamma \geq \gamma^{\text{cr}}$, then $\mathcal{E}_\gamma$ is decreasing in γ . Under (C2a) (or (C3a)), with $\gamma \geq \gamma^{\text{cr}}$, $\mathcal{E}_\gamma$ is $\{\underline{\xi}\}$ (or $\{\bar{\xi}\}$) by Theorem 3. Consequently, $\mathcal{E}_\gamma$ is non-increasing in γ under any of the Conditions (C2a) and (C3a).

(ii) First, note that $\lim_{\gamma \rightarrow 2} \mathcal{E}_\gamma = \cap_{\gamma=0}^2 \mathcal{E}_\gamma$ because $\mathcal{E}_\gamma$ is non-increasing in γ . Consider $\mathcal{E}_\gamma$ as presented in Theorem 3 and Remark 7.

Under (C1), $\frac{Wx_\gamma^* - \alpha_\gamma^*}{W+V} \searrow \underline{\xi}$ and $\frac{Ux_\gamma^* + \alpha_\gamma^*}{U-V} \nearrow \bar{\xi}$ as $\gamma \nearrow 2$. Thus, $\mathcal{E}_\gamma \searrow \{\underline{\xi}, \bar{\xi}\}$ as $\gamma \nearrow 2$.

Under (C2a), $F^{-1}(Q - \frac{\gamma}{2}) \searrow \underline{\xi}$ as $\gamma \nearrow 2Q$ and $\mathcal{E}_\gamma = \{\underline{\xi}\}$ for $\gamma \geq 2Q$. Under (C2b), $F^{-1}(1 - \frac{\gamma}{2}) \searrow \underline{\xi}$ as $\gamma \nearrow 2$. Thus, $\mathcal{E}_\gamma \searrow \{\underline{\xi}\}$ as $\gamma \nearrow 2$ under (C2).

Under (C3a), $F^{-1}(Q + \frac{\gamma}{2}) \searrow \bar{\xi}$ as $\gamma \nearrow 2(1 - Q)$ and $\mathcal{E}_\gamma = \{\bar{\xi}\}$ for $\gamma \geq 2(1 - Q)$. Under (C3b), $F^{-1}(\frac{\gamma}{2}) \searrow \bar{\xi}$ as $\gamma \nearrow 2$. Thus, $\mathcal{E}_\gamma \searrow \{\bar{\xi}\}$ as $\gamma \nearrow 2$ under (C3). ■

A.2.6 Proof of Corollary 1

Proof of Corollary 1. Under the first set of conditions stated, first, suppose $0 < \gamma < 2$. Then, by Theorem 2, there exists a limiting worst-case probability distribution whose density function is given by (9) and supported on $\left[h(x_\gamma^*, \xi) \geq \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]\right]$. On the other hand, by the proof of Theorem 3, $\mathcal{E}_\gamma = \left[h(x_\gamma^*, \xi) \geq \text{VaR}_{\frac{\gamma}{2}}[h(x_\gamma^*, \xi)]\right]$.

Now, suppose $\gamma = 2$. Then, by Theorem 2, there exists a limiting worst-case probability distribution whose density function is given by (11) and supported on $\left[h(x_\gamma^*, \xi) = \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)\right]$. On the other hand, by the proof of Theorem 3, $\mathcal{E}_\gamma = \left[h(x_\gamma^*, \xi) = \sup_{\xi \in \Omega} h(x_\gamma^*, \xi)\right]$.

Under either (C2a) or (C3a) with $\gamma^{\text{cr}} \leq \gamma \leq 2$, by Theorem 2, there exists a limiting worst-case probability distribution that puts a probability mass of one on $\{\underline{\xi}\}$ or $\{\bar{\xi}\}$, respectively. On the other hand, under (C2a) (or (C3a)), with $\gamma^{\text{cr}} \leq \gamma \leq 2$, $\mathcal{E}_\gamma$ is $\{\underline{\xi}\}$ (or $\{\bar{\xi}\}$) by Theorem 3. ■

A.3 Proof of Section 5 (Price of Optimism/Pessimism and Regrets)

Proof of Theorem 5. If $x^{\text{neut}} = x^{\text{rob}}$, the price of optimism, the price of pessimism, and regrets are all zero for $0 \leq \gamma \leq 2$ because $x_\gamma^* = x^{\text{neut}} = x^{\text{rob}}$ from Theorem 1. So, $\gamma^{\text{S}} = \gamma^{\text{D}} = 0$.

Suppose $x^{\text{neut}} \neq x^{\text{rob}}$. We first show γ^{S} is well defined. Note that at any level of robustness γ , x_γ^* is the unique optimal order quantity to (DRNV-V). Thus, $\text{PO}_0 = 0$ and $\text{PP}_0 > 0$ according to (16). On the other hand, $\text{PO}_2 > 0$ and $\text{PP}_2 = 0$. These imply $\text{PO}_0 - \text{PP}_0 < 0$ and $\text{PO}_2 - \text{PP}_2 > 0$. So, if we show $\text{PO}_\gamma - \text{PP}_\gamma$ is continuous in γ , there exists some $0 < \gamma < 2$ such that $\text{PO}_\gamma - \text{PP}_\gamma = 0$, i.e., γ^{S} is well defined. Consider a fixed $x \in \mathbb{X}$. We show $f_\gamma(x)$ is continuous in γ . It is well

known that $\text{CVaR}_\beta[h(x, \xi)]$ is continuous in $\beta \in (0, 1)$, $\lim_{\beta \rightarrow 0} \text{CVaR}_\beta[h(x, \xi)] = \mathbb{E}_{\mathbb{P}_0}[h(x, \xi)]$, and $\lim_{\beta \rightarrow 1} \text{CVaR}_\beta[h(x, \xi)] = \sup_{\xi \in \Omega} h(x, \xi)$ (see, e.g., [Rockafellar and Uryasev 2002](#); [Krokhmal et al. 2011](#)). Consequently, $(1 - \frac{\gamma}{2}) \text{CVaR}_{\frac{\gamma}{2}}[h(x, \xi)]$ is continuous in γ . Now, because $\frac{\gamma}{2} \sup_{\xi \in \Omega} h(x, \xi)$ is linear in γ , $f_\gamma(x)$ is continuous in γ . In particular, $f_\gamma(x^{\text{neut}})$ and $f_\gamma(x^{\text{rob}})$ are continuous in γ , and so is $\text{PP}_\gamma - \text{WR}_\gamma$.

Now, we show γ^D is well defined. Similar to above, it is easy to verify that $\text{NR}_0 - \text{WR}_0 < 0$ and $\text{NR}_2 - \text{WR}_2 > 0$. So, continuity of $\text{NR}_\gamma - \text{WR}_\gamma$ in γ proves there exists some $0 < \gamma < 2$ such that $\text{NR}_\gamma - \text{WR}_\gamma = 0$. Note that x_γ^* is continuous in γ by Theorem 1 and also $f_0(\cdot)$ is continuous. So, the composite function $f_0(x_\gamma^*)$ is continuous in γ . Moreover, because $f_0(x^{\text{neut}})$ is constant, NR_γ is continuous in γ . Similarly, WR_γ is continuous in γ because $f_2(x_\gamma^*)$ is continuous in γ and $f_2(x^{\text{rob}})$ is constant. So, $\text{NR}_\gamma - \text{WR}_\gamma$ is continuous in γ . $\blacksquare$

B Appendix B

Appendix B provides the proof of Lemma 3. For completeness, it also provides the proof of Proposition 1 for DRNV-V based on the proof of [Jiang and Guan \(2015, Theorem 1\)](#) for a general DRSP formed via the variation distance.

B.1 Proof of Lemma 3

Proof of Lemma 3. When $\gamma = 0$, we have $\underline{\xi} < x_0^* = x^{\text{neut}} = F^{-1}(Q) < \bar{\xi}$ because $0 < Q < 1$. When $\gamma = 2$, $\underline{\xi} \leq x_2^* = x^{\text{rob}} \leq \bar{\xi}$ according to (5). For the rest of the proof, suppose $0 < \gamma < 2$, and consider a fixed $x \in \mathbb{R}$. We will show that if x is not in the support of Ω ,

$$0 \in \partial_\alpha \mathcal{T}_\gamma(x, \alpha), \quad 0 \in \partial_x \mathcal{T}_\gamma(x, \alpha) \quad (42)$$

has no solution.

First, we show $x_\gamma^* \geq \underline{\xi}$ by contradiction; so suppose $x < \underline{\xi}$. Then, we have $h(x, \xi) = U(\xi - x) - V\xi$ for all $\xi \in \Omega$. Under either Conditions (C1) or (C3), if $\alpha > U(\underline{\xi} - x) - V\underline{\xi}$, we have $\frac{\partial}{\partial \alpha} \mathcal{T}_\gamma(x, \alpha) = F\left(\frac{Ux + \alpha}{U - V}\right) - \frac{\gamma}{2}$ and $\frac{\partial}{\partial x} \mathcal{T}_\gamma(x, \alpha) = UF\left(\frac{Ux + \alpha}{U - V}\right) - \left(1 + \frac{\gamma}{2}\right)U$ by (22) and Lemma 2. However, (42) has

no solution because $U > 0$ by Assumption (A1). Otherwise, if $\alpha \leq U(\underline{\xi} - x) - V\underline{\xi}$, $F\left(\frac{Ux+\alpha}{U-V}\right) = 0$. Again, there is no solution to (42). Consequently, we have $x_\gamma^* \geq \underline{\xi}$ under either (C1) or (C3).

Under (C2), if $\alpha > U(\underline{\xi} - x) - V\underline{\xi}$, $\mathfrak{E}(x, \alpha) = 0$, and hence $\frac{\partial}{\partial \alpha} \mathcal{T}_\gamma(x, \alpha) = 1 - \frac{\gamma}{2}$ and $\frac{\partial}{\partial x} \mathcal{T}_\gamma(x, \alpha) = -\frac{\gamma}{2}U$ by (20). As a result, (42) has no solution because $U > 0$ by Assumption (A1). Otherwise, if $\alpha \leq U(\underline{\xi} - x) - V\underline{\xi}$ under (C2b), we have $\frac{\partial}{\partial \alpha} \mathcal{T}_\gamma(x, \alpha) = 1 - F\left(\frac{Ux+\alpha}{U-V}\right) - \frac{\gamma}{2}$ and $\frac{\partial}{\partial x} \mathcal{T}_\gamma(x, \alpha) = -U\left(\frac{\gamma}{2} + F\left(\frac{Ux+\alpha}{U-V}\right)\right)$ by (22) and Lemma 2. As a result, (42) has no solution because $U > 0$ by Assumption (A1). Under (C2a), if $\alpha = U(\underline{\xi} - x) - V\underline{\xi}$, then $\partial_x \mathcal{T}_\gamma(x, \alpha) = [-(1 + \frac{\gamma}{2})U, -\frac{\gamma}{2}U]$ and similarly, if $\alpha < U(\underline{\xi} - x) - V\underline{\xi}$, then $\frac{\partial}{\partial x} \mathcal{T}_\gamma(x, \alpha) = -(1 + \frac{\gamma}{2})U$ by Lemmas 1 and 2. As a result, (42) has no solution because $0 < \gamma < 2$ and $U > 0$ by Assumption (A1). Consequently, we have $x_\gamma^* \geq \underline{\xi}$ under (C2).

Now, we show $x_\gamma^* \leq \bar{\xi}$, assuming $\bar{\xi} < \infty$, by contradiction. So, suppose $x > \bar{\xi}$. Then, we have $h(x, \xi) = W(x - \xi) - V\xi$ for all $\xi \in \Omega$. Under either (C1) or (C2), if $\alpha > W(x - \bar{\xi}) - V\bar{\xi}$, we have $\frac{\partial}{\partial \alpha} \mathcal{T}_\gamma(x, \alpha) = 1 - F\left(\frac{Wx-\alpha}{W+V}\right) - \frac{\gamma}{2}$ and $\frac{\partial}{\partial x} \mathcal{T}_\gamma(x, \alpha) = W\left(\frac{\gamma}{2} + F\left(\frac{Wx-\alpha}{W+V}\right)\right)$ by (22) and Lemma 2. However, (42) has no solution because $W > 0$ by Assumption (A1). Otherwise, if $\alpha \leq W(x - \bar{\xi}) - V\bar{\xi}$, $F\left(\frac{Wx-\alpha}{W+V}\right) = 0$. Again, there is no solution to (42). Consequently, we conclude $x_\gamma^* \leq \bar{\xi}$ under either (C1) or (C2).

Under (C3), if $\alpha > W(x - \bar{\xi}) - V\bar{\xi}$, $\mathfrak{E}(x, \alpha) = 0$, and hence $\frac{\partial}{\partial \alpha} \mathcal{T}_\gamma(x, \alpha) = 1 - \frac{\gamma}{2}$ and $\frac{\partial}{\partial x} \mathcal{T}_\gamma(x, \alpha) = \frac{\gamma}{2}W$ by (20). As a result, (42) has no solution because $W > 0$ by Assumption (A1). Otherwise, if $\alpha \leq W(x - \bar{\xi}) - V\bar{\xi}$ under (C3b), we have $\frac{\partial}{\partial \alpha} \mathcal{T}_\gamma(x, \alpha) = F\left(\frac{Wx-\alpha}{W+V}\right) - \frac{\gamma}{2}$ and $\frac{\partial}{\partial x} \mathcal{T}_\gamma(x, \alpha) = W\left(\frac{\gamma}{2} + 1 - F\left(\frac{Wx-\alpha}{W+V}\right)\right)$ by (22) and Lemma 2. As a result, (42) has no solution because $W > 0$ by Assumption (A1). Under (C3a), if $\alpha = W(x - \bar{\xi}) - V\bar{\xi}$, then $\partial_x \mathcal{T}_\gamma(x, \alpha) = [\frac{\gamma}{2}W, (1 + \frac{\gamma}{2})W]$ and similarly, if $\alpha < W(x - \bar{\xi}) - V\bar{\xi}$, then $\frac{\partial}{\partial x} \mathcal{T}_\gamma(x, \alpha) = (1 + \frac{\gamma}{2})W$. As a result, (42) has no solution because $0 < \gamma < 2$ and $W > 0$ by Assumption (A1). Consequently, we have $x_\gamma^* \leq \bar{\xi}$ under (C3).

Now, under (C2), assume the support of Ω is unbounded, i.e., $\bar{\xi} = \infty$. We show $x_\gamma^* < \infty$. If $x = \infty$, then $h(x, \xi) = \infty$ for all $\xi < \infty$. As a result, the worst-case expected value to (DRNV-V) at $x = \infty$ is ∞ ; hence, $x_\gamma^* < \infty$. ■

B.2 Proof of Proposition 1

Proof of Proposition 1. Consider a fixed $x \in \mathbb{X}$ and a subset $\mathcal{F} \subset \Omega$. Observe that by the reverse triangle inequality, $\int_{\mathcal{F}^c} |p(s) - p_0(s)| ds \geq |\int_{\mathcal{F}^c} p(s) ds - \int_{\mathcal{F}^c} p_0(s) ds|$. So, if $\int_{\mathcal{F}^c} p(s) ds = 1$, then $\int_{\mathcal{F}^c} |p(s) - p_0(s)| \geq \mathbb{P}_0\{\mathcal{F}\}$ because $\int_{\mathcal{F}^c} p_0(s) ds = 1 - \mathbb{P}_0\{\mathcal{F}\}$. On the other hand, if $0 < \gamma < 2\mathbb{P}_0\{\mathcal{F}\}$, $\gamma - \mathbb{P}_0\{\mathcal{F}\} < \mathbb{P}_0\{\mathcal{F}\}$. So, the set $\mathcal{P}_\gamma^A(\mathcal{F})$ is empty because the first constraint in (13) is violated.

Now, choose $\gamma \geq 2\mathbb{P}_0\{\mathcal{F}\}$. Let λ and μ denote Lagrange multipliers for the first and second constraints in (13), respectively. By relaxing the first and second constraints in (13) we can write the Lagrangian dual formulation of the inner maximization problem in (12) as

$$\begin{aligned} f_\gamma^A(x; \mathcal{F}) &= \min_{\lambda \geq 0, \mu} \mu - \lambda(1 - \gamma) + \int_{\mathcal{F}^c} \left(h(x, \xi) - \mu + \lambda \right)_+ p_0(\xi) d\xi \\ &\quad \text{s.t. } \text{ess sup}_{\mathcal{F}^c} h(x, \xi) \leq \mu + \lambda \\ &= \min_{\lambda \geq 0, \mu} \mu - \lambda(1 - \gamma) + \int_{\mathcal{F}^c \cap [h(x, \xi) - \mu + \lambda > 0]} \left(h(x, \xi) - \mu + \lambda \right) p_0(\xi) d\xi \end{aligned} \quad (43)$$

$$\begin{aligned} &\quad \text{s.t. } \text{ess sup}_{\mathcal{F}^c} h(x, \xi) \leq \mu + \lambda \\ &= \min_{\lambda \geq 0} -\lambda(2 - \gamma) + m^* + \int_{\mathcal{F}^c \cap [h(x, \xi) - m^* + 2\lambda > 0]} \left(h(x, \xi) - m^* + 2\lambda \right) p_0(\xi) d\xi, \end{aligned} \quad (44)$$

where $m^* := \text{ess sup}_{\mathcal{F}^c} h(x, \xi)$. Equality (44) holds because $\mu^* = m^* - \lambda$ solves the optimization problem (43) over μ for a fixed $\lambda \geq 0$. To see this claim, we consider another feasible solution $\bar{\mu} > \mu^*$ and show it is suboptimal to (43) for a fixed λ . We have

$$\begin{aligned} &\bar{\mu} + \int_{\mathcal{F}^c \cap [h(x, \xi) - \bar{\mu} + \lambda > 0]} \left(h(x, \xi) - \bar{\mu} + \lambda \right) p_0(\xi) d\xi - \mu^* - \int_{\mathcal{F}^c \cap [h(x, \xi) - \mu^* + \lambda > 0]} \left(h(x, \xi) - \mu^* + \lambda \right) p_0(\xi) d\xi \\ &= (\bar{\mu} - \mu^*) + \int_{\mathcal{F}^c \cap [h(x, \xi) - \bar{\mu} + \lambda > 0]} \left(h(x, \xi) - \bar{\mu} + \lambda \right) p_0(\xi) d\xi - \int_{\mathcal{F}^c \cap [h(x, \xi) - \bar{\mu} + \lambda > 0]} \left(h(x, \xi) - \mu^* + \lambda \right) p_0(\xi) d\xi \\ &\quad - \int_{\mathcal{F}^c \cap [\mu^* - \lambda < h(x, \xi) \leq \bar{\mu} - \lambda]} \left(h(x, \xi) - \mu^* + \lambda \right) p_0(\xi) d\xi \end{aligned} \quad (45)$$

$$\begin{aligned} &= (\bar{\mu} - \mu^*) + \int_{\mathcal{F}^c \cap [h(x, \xi) - \bar{\mu} + \lambda > 0]} (\mu^* - \bar{\mu}) p_0(\xi) d\xi - \int_{\mathcal{F}^c \cap [\mu^* - \lambda < h(x, \xi) \leq \bar{\mu} - \lambda]} \left(h(x, \xi) - \mu^* + \lambda \right) p_0(\xi) d\xi, \end{aligned} \quad (46)$$

where (45) holds because $[h(x, \xi) - \mu^* + \lambda > 0] = [h(x, \xi) - \bar{\mu} + \lambda > 0] \cup [\mu^* - \lambda < h(x, \xi) \leq \bar{\mu} - \lambda]$.

We show the value of (46) is greater than or equal to zero to complete the proof of the claim.

Suppose $\mu^* - \lambda < h(x, \xi) \leq \bar{\mu} - \lambda$ or equivalently, $0 < h(x, \xi) - \mu^* + \lambda \leq \bar{\mu} - \mu^*$. This implies

$\int_{\mathcal{F}^c \cap [\mu^* - \lambda < h(x, \xi) \leq \bar{\mu} - \lambda]} (h(x, \xi) - \mu^* + \lambda) p_0(\xi) d\xi \leq \int_{\mathcal{F}^c \cap [\mu^* - \lambda < h(x, \xi) \leq \bar{\mu} - \lambda]} (\bar{\mu} - \mu^*) p_0(\xi) d\xi$. As a result,

$$\begin{aligned} & \int_{\mathcal{F}^c \cap [h(x, \xi) - \bar{\mu} + \lambda > 0]} (\mu^* - \bar{\mu}) p_0(\xi) d\xi - \int_{\mathcal{F}^c \cap [\mu^* - \lambda < h(x, \xi) \leq \bar{\mu} - \lambda]} (h(x, \xi) - \mu^* + \lambda) p_0(\xi) d\xi \\ & \geq (\mu^* - \bar{\mu}) \int_{\mathcal{F}^c \cap [h(x, \xi) - \bar{\mu} + \lambda > 0]} p_0(\xi) d\xi + (\mu^* - \bar{\mu}) \int_{\mathcal{F}^c \cap [\mu^* - \lambda < h(x, \xi) \leq \bar{\mu} - \lambda]} p_0(\xi) d\xi \\ & = (\mu^* - \bar{\mu}) \int_{\mathcal{F}^c \cap [h(x, \xi) - \mu^* + \lambda > 0]} p_0(\xi) d\xi \\ & \geq \mu^* - \bar{\mu}. \end{aligned}$$

Now, we solve (44). Let $B(\lambda)$ represent the objective function of (44). We can rewrite $B(\lambda)$ as

$$B(\lambda) = -\lambda(2 - \gamma) + m^* + (1 - \mathbb{P}_0\{\mathcal{F}\}) \int_{\mathcal{F}^c \cap [h(x, \xi) - m^* + 2\lambda > 0]} (h(x, \xi) - m^* + 2\lambda) p_{0|\mathcal{F}^c}(\xi) d\xi. \quad (47)$$

Because $\left(h(x, \xi) - m^* + 2\lambda\right)_+$ is convex in λ , by an application of Lemma 1 and chain rule, we have

$$\begin{aligned} \partial B(\lambda) &= -2 + \gamma - 2(1 - \mathbb{P}_0\{\mathcal{F}\}) [\Psi_{|\mathcal{F}^c}(x, m^* - 2\lambda^-) - 1, \Psi_{|\mathcal{F}^c}(x, m^* - 2\lambda) - 1] \\ &= \gamma - 2\mathbb{P}_0\{\mathcal{F}\} - 2(1 - \mathbb{P}_0\{\mathcal{F}\}) [\Psi_{|\mathcal{F}^c}(x, m^* - 2\lambda^-), \Psi_{|\mathcal{F}^c}(x, m^* - 2\lambda)], \end{aligned}$$

where $\Psi_{|\mathcal{F}^c}(x, \eta)$ and $\Psi_{|\mathcal{F}^c}(x, \eta^-)$ denote the conditional versions of $\Psi(x, \eta)$ and $\Psi(x, \eta^-)$ on $\mathcal{F}^c$, respectively. Note that $\partial B(\lambda)$ is non-decreasing in λ . Hence, $B(\lambda)$ is convex in λ . We discuss the cases $\gamma \geq 2$, $2\mathbb{P}_0\{\mathcal{F}\} < \gamma < 2$, and $\gamma = 2\mathbb{P}_0\{\mathcal{F}\}$ separately. Let $\gamma' = \frac{\gamma - \mathbb{P}_0\{\mathcal{F}\}}{1 - \mathbb{P}_0\{\mathcal{F}\}}$. If $\gamma \geq 2$, $2(1 - \mathbb{P}_0\{\mathcal{F}\})\partial B(\lambda) = \gamma' - [\Psi_{|\mathcal{F}^c}(x, m^* - 2\lambda^-), \Psi_{|\mathcal{F}^c}(x, m^* - 2\lambda)]$ is a non-negative set because $\gamma' \geq 1$. So, $\lambda^* = 0$ solves (47) and we can obtain $f_\gamma^A(x; \mathcal{F})$ as follows:

$$f_\gamma^A(x; \mathcal{F}) = B(0) = m^* + (1 - \mathbb{P}_0\{\mathcal{F}\}) \int_{\mathcal{F}^c} \left(h(x, \xi) - m^*\right)_+ p_{0|\mathcal{F}^c}(\xi) d\xi = m^* = \text{ess sup}_{\mathcal{F}^c} h(x, \xi).$$

Otherwise, if $2\mathbb{P}_0\{\mathcal{F}\} < \gamma < 2$, then $0 \in \partial B(\lambda)$ or equivalently, $\gamma' \in [\Psi_{|\mathcal{F}^c}(x, m^* - 2\lambda^-), \Psi_{|\mathcal{F}^c}(x, m^* - 2\lambda)]$, gives a sufficient and necessary condition to solve (44). That is, regardless of $\mathbb{P}_0\{h(x, \xi) = m^* - 2\lambda|\mathcal{F}^c\}$, we have

$$m^* - 2\lambda^* = \inf_{\eta \in \mathbb{R}} \{\Psi_{|\mathcal{F}^c}(x, \eta) \geq \gamma'\}, \quad (48)$$

where $\inf_{\eta \in \mathbb{R}} \{\Psi_{|\mathcal{F}^c}(x, \eta) \geq \gamma'\}$ is the VaR of $h(x, \xi)$ at level $0 < \gamma' < 1$ *conditioned on $\mathcal{F}^c$* and is denoted by $\text{VaR}_{\gamma'}[h(x, \xi)|\mathcal{F}^c]$. Now, we obtain $f_\gamma^A(x; \mathcal{F})$ by replacing λ^* from (48) in (47). We have

$$\begin{aligned} f_\gamma^A(x; \mathcal{F}) &= \frac{\gamma}{2}m^* + \left(1 - \frac{\gamma}{2}\right) \text{VaR}_{\gamma'}[h(x, \xi)|\mathcal{F}^c] + \\ &\quad (1 - \mathbb{P}_0\{\mathcal{F}\}) \int_{\mathcal{F}^c} \left(h(x, \xi) - \text{VaR}_{\gamma'}[h(x, \xi)|\mathcal{F}^c]\right)_+ p_{0|\mathcal{F}^c}(\xi) d\xi \\ &= \frac{\gamma}{2}m^* + \left(1 - \frac{\gamma}{2}\right) [\text{VaR}_{\gamma'}[h(x, \xi)|\mathcal{F}^c] + \\ &\quad \frac{1}{1 - \gamma'} \int_{\mathcal{F}^c} \left(h(x, \xi) - \text{VaR}_{\gamma'}[h(x, \xi)|\mathcal{F}^c]\right)_+ p_{0|\mathcal{F}^c}(\xi) d\xi] \end{aligned} \quad (49)$$

$$= \frac{\gamma}{2} \text{ess sup}_{\mathcal{F}^c} h(x, \xi) + \left(1 - \frac{\gamma}{2}\right) \text{CVaR}_{\gamma'}[h(x, \xi)|\mathcal{F}^c], \quad (50)$$

where (49) is true because $1 - \frac{\gamma}{2} = (1 - \gamma')(1 - \mathbb{P}_0\{\mathcal{F}\})$ and (50) is due to [Rockafellar and Uryasev \(2002, Theorem 10\)](#) for the conditional distribution $p_{0|\mathcal{F}^c}$. Finally, if $\gamma = 2\mathbb{P}_0\{\mathcal{F}\}$, i.e., $\gamma' = 0$, similar to above we can show that $f_\gamma^A(x; \mathcal{F}) = \frac{\gamma}{2} \text{ess sup}_{\mathcal{F}^c} h(x, \xi) + \left(1 - \frac{\gamma}{2}\right) \mathbb{E}_{\mathbb{P}_0}[h(x, \xi)|\mathcal{F}^c]$. $\blacksquare$