ARTICLE IN PRESS
JID: EOR [m5G;July 8, 2019;13:52]
European Journal of Operational Research xxx (xxxx) xxx
Contents lists available at ScienceDirect
European Journal of Operational Research
journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing, Transportation and Logistics

# Controlling risk and demand ambiguity in newsvendor models

Hamed Rahimian[a,*], Güzin Bayraksan[b], Tito Homem-de-Mello[c]

[a] Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston IL 60208, USA
[b] Department of Integrated Systems Engineering, The Ohio State University, Columbus OH 43210, USA
[c] School of Business, Universidad Adolfo Ibañez, Santiago, Chile

ARTICLE INFO

*Article history:*
Received 19 October 2018
Accepted 17 June 2019
Available online xxx

*Keywords:*
Inventory
Distributionally robust optimization
Operating room time reservation problem
Newsvendor problem
Calibration of level of robustness

ABSTRACT

We use distributionally robust optimization (DRO) to model a general class of newsvendor problems with unknown demand distribution. 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 total variation distance—from a nominal distribution. The maximum distance allowed in the ambiguity set (called *level of robustness*) places the DRO between the risk-neutral stochastic programming and robust optimization models. An important problem a decision maker faces is how to determine the level of robustness—or, equivalently, how to find an appropriate level of risk-aversion. We answer this question in two ways. Our first approach relates the level of robustness and risk to the regions of demand that are critical (in a precise sense we introduce) to the optimal cost. Our second approach establishes new quantitative relationships between the DRO model and the corresponding risk-neutral and classical robust optimization models. To achieve these goals, we first focus on a single-product setting and derive explicit formulas and properties of the optimal solution as a function of the level of robustness. Then, we demonstrate the practical and managerial relevance of our results by applying our findings to a healthcare problem to reserve operating room time for cardiovascular surgeries. Finally, we extend some of our results to the multi-product setting and illustrate them numerically.



## 1. Introduction

The newsvendor problem is fundamental to many operations management models. It has been used, for instance, in production of influenza vaccines (Chick, Mamani, & Simchi-Levi, 2008), staffing problems (Harrison & Zeevi, 2005), reservation of operating room time (Olivares, Terwiesch, & Cassorla, 2008), and 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 determine an optimal quantity. For a review on the newsvendor problem, we refer the readers to Qin, Wang, Vakharia, Chen, and Seref (2011).

There are multiple ways of formulating a newsvendor problem. For instance, one could consider "classical" stochastic programming (SP) or robust optimization (RO) approaches[1]. 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.

In practice, 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. 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

* Corresponding author.
*E-mail addresses:* hamed.rahimian@northwestern.edu (H. Rahimian), bayraksan.1@osu.edu (G. Bayraksan), tito.hmello@uai.cl (T. Homem-de-Mello).

[1] There 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.

https://doi.org/10.1016/j.ejor.2019.06.036
0377-2217/

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 yield a suboptimal production quantity that is overly conservative, and therefore results in a large 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 optimization* (DRO for short). In such models, the goal is to find a decision that minimizes the *worst-case expected cost*, where "worst-case" is taken with respect to a set of distributions, called the *ambiguity set*. The classical SP and RO models can be viewed as special cases of DRO: When the ambiguity set of DRO 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, DRO lies between the two approaches.

Under mild conditions (e.g., real-valued costs, convex ambiguity sets), DRO is equivalent to a risk-averse stochastic program with a coherent measure of risk (see, e.g., Artzner, Delbaen, Eber, & Heath 1999). Our setting satisfies this equivalence relation; we will discuss this in more detail in Section 3. 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*.

The motivation for our study arises from the following observations. First, a 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. Establishing such a direct relationship yields multiple benefits. For example, this paper proposes to use critical regions of demand to determine an appropriate level of robustness. Such analysis may also help decision makers understand their risk attitude better and encourage them to collect more accurate information surrounding critical demands since the tails of distributions are often neglected in standard statistical methods for distribution fitting.

Second, in many practical applications—including the one studied in this paper—available data is not independently and identically distributed (i.i.d.), or might be confounded by other variables, thereby rendering many of the data-driven methods to determine the level of robustness unsuitable. In these cases, in addition to using critical regions, we propose to calibrate the level of robustness by balancing the regrets and prices—to be defined precisely in Section 6—of DRO relative to SP and RO. To be able to achieve these two goals, we first characterize the optimal solution as a function of the level of robustness.

### *1.1. The setting*

We study a distributionally robust newsvendor model (DRNV for short) with the following characteristics. The decision maker (i) has some "belief" about the demand distribution (based perhaps on available data), termed the *nominal distribution*, (ii) reckons the underlying demand distribution is close enough, in the sense of the *total variation distance*, to the nominal distribution (thereby forms the ambiguity set as the set of distributions whose total variation distances from the nominal distribution are bounded above by a level-of-robustness parameter), and (iii) minimizes the worst-case expected cost with respect to this ambiguity set. We present a formal definition and assumptions in Section 3 for the single-product, single-period newsvendor. We refer to this problem as DRNV-V. We then study the multi-product setting in Section 8.

Using total variation distance to form DRNV has several advantages. The first one is the intuitive meaning of the total variation distance, which gives a direct value to how close the distributions are. From the perspective of a decision maker, an important 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. This relationship also helps decision makers relate the results to their level of risk-aversion. A third advantage of using the total variation distance is its 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 total variation distance enables us to fully characterize the regions of demand that are most critical to the problem and obtain prices and regrets.

### *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 to DRNV-V as a function of the level of robustness and show some of its properties.

(ii) *Characterization of maximal effective demand regions.* 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 analytically characterize these subsets at different levels of robustness and discuss their interpretations.

(iii) *Calibration of the level of robustness.* We propose two methods to calibrate the level of robustness. One is based on the maximal effective subsets and the other one is based on the optimal order quantities' performance in the DRNV-V, SP, and RO settings.
We discuss the measures *price of optimism/pessimism*, *nominal/worst-case regret*, and further propose the *indifferent-to-solution/indifferent-to-distribution level of robustness* to balance the performance of DRNV-V relative to SP and RO.

(iv) *Application to operating room time reservation.* We illustrate the practical relevance of our results by applying them to a real-world healthcare problem, where a hospital has to reserve a certain amount of operating room (OR) time to specific cardiac surgeries (Olivares et al., 2008). Reserving too much OR time to a surgery is likely to incur excessive idle time for the hospital staff and capacity. Reserving too little OR time, however, leads to more frequent schedule overruns and overtime hours for the hospital staff and decreased service quality. Reserving OR time can be modeled as a newsvendor problem, where the hospital balances under- and over-utilization costs. As we shall see, the tools discussed in this paper allow the hospital managers to understand their tolerance to risk according to their evaluation of critical lengths of surgeries (in terms of cost). Such understanding, in turn, guides the choice of a proper OR reservation time.

(v) *Extension to multi-product setting.* Finally, we extend some of our results for the single-product setting to the multi-product setting and illustrate them numerically.

Please cite this article as: H. Rahimian, G. Bayraksan and T. Homem-de-Mello, Controlling risk and demand ambiguity in newsvendor models, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2019.06.036

### 1.3. Organization

The rest of the paper is organized as follows. The next section reviews the literature. We formally define the problem setting and list the assumptions and conditions used throughout the paper in Section 3. We also provide examples of different variants covered by our setting. Section 4 characterizes the optimal order quantity, and Section 5 introduces the maximal effective subsets of demand. Section 6 introduces the concepts of the price of optimism/pessimism and nominal/worst-case regrets for a general DRO. In Section 7, we apply the ideas set forth in the paper to the OR reservation time problem described in Olivares et al. (2008). We also present further insights on how to choose a level of robustness based on our analysis of maximal effective subsets and the indifference levels. In Section 8, we extend some of our results in Sections 4–6 to the multi-product setting and illustrate them numerically. Finally, we end the paper with conclusions and future research directions in Section 9. All omitted results and proofs are provided in the Online Supplement.

## 2. Literature review

Because the majority of the paper focuses on a single-period, single-product newsvendor problem, we review this class of problems in this section. We provide a review of the class of multi-product newsvendor problems in Section 8, where we present an extension to the multi-product setting. Within this class, 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.

Most studies on DRNV form the ambiguity set by all probability distributions with sufficiently close moments (typically up to second-order moments). Scarf (1958) proposed the first such model; for other studies, see, e.g., Gallego and Moon (1993), Mostard, de Koster, and Teunter (2005), and Natarajan, Sim, and Uichanco (2018). Among these studies, some papers consider cost functionals other than the expectation. For instance, Perakis and Roels (2008) use regret-based cost functionals, and Han, Du, and Zuluaga (2014) and Yu, Zhai, and Chen (2016) use 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, Glynn, & Ye, 2016), variation distance (Jiang & Guan, 2018), $\zeta$-structure probability metrics (Zhao & Guan, 2015), and Kolmogorov–Smirnov distance (Bertsimas, Gupta, & Kallus, 2018) have been used as "distances" between probability distributions. Our work belongs to this category and subsumes the specific newsvendor model in Jiang and Guan (2018) as a special case (see Section 3.2 for details). It also provides a significantly more detailed study, including the critical demand regions and various novel ways to calibrate the model.

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, Çakmak, and Shapiro (2007) study mean-risk objective functions, where risk is either the $p$th semideviation or CVaR. Choi and Ruszczyński (2008) derive an equivalent mean-risk model for risk-averse newsvendor with a law invariant coherent measure of risk. Wu, Zhu, and Teunter (2013b) study two newsvendor problems with the CVaR objective and a Value-at-Risk (VaR) constraint, where the production capacity and demand are random. Wu, Zhu, and Teunter (2013a) address similar problems as the ones studied in Wu et al. (2013b), but they assume the shortage cost and demand are random. Chen, Xu, and Zhang (2009) and Wu, Zhu, and Teunter (2014) study both ordering and pricing decisions with the CVaR objective. For an earlier detailed overview of risk-averse newsvendor problems, we refer to Choi, Ruszczyński, and Zhao (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, similar to the works on the risk-averse newsvendor. However, unlike most of these works, we present an interpretation of this optimal order quantity in terms of the solutions to the classical SP- and RO-based newsvendor problems. Furthermore, we show a *critical level of robustness*, where the optimal order quantity stabilizes at the RO solution. Furthermore, 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 3.2 for details).

Another contribution of our work to the literature is the study of critical demand regions. Building on the work of Rahimian, Bayraksan, and Homem-de Mello (2019), we introduce the notion of maximal effective subsets. We expand upon Rahimian et al. (2019) in several other ways. First, in this paper we have continuous support of the random variable (instead of discrete), which is a nontrivial extension. Second, by exploiting the properties of the newsvendor problem, we are able to *fully* characterize the effective subsets of DRNV-V (unlike previous work). Importantly, we relate the maximal effective sets to the level of robustness, which allows us to use this new notion to determine an appropriate level of robustness based on the decision maker's preferences. We believe using maximal effective subsets for this purpose may have independent interest.

Because price of optimism/pessimism and nominal/worst-case regrets are defined for a general DRO problem, we contrast our investigation on these notions to other works beyond the newsvendor model. Price of optimism has been used in the context of DRO (Analui & Pflug, 2014) and risk-averse optimization (Zhang, Rahimian, & Bayraksan, 2016). Nominal regret has also been used in the literature in various contexts: e.g., DRO (Analui & Pflug, 2014; Gallego & Moon, 1993; Perakis & Roels, 2008), RO (Averbakh, 2001), and risk-averse optimization (Shapiro, Tekaya, da Costa, & Soares, 2013; Zhang et al., 2016). However, to the best of our knowledge, the price of pessimism and worst-case regret are new. Furthermore, this is the first paper to propose indifference levels of robustness for DRO that balance prices and regrets with respect to the SP and RO models. Again, we believe this may have independent interest beyond the class of problems considered here.

Finally, we contrast our approaches to calibrate the level of robustness with those in the literature, again going beyond the newsvendor model. Data-driven DROs typically propose a level of robustness that is inversely proportional to the number of observations, assuming i.i.d. data. One common approach is to choose the level of robustness based on the large-sample analysis of the corresponding distance; see, e.g., Jiang and Guan (2018), and Zhao and Guan (2015). Hypothesis test-based DROs, e.g., Ben-Tal, Den Hertog, De Waegenaere, Melenberg, and Rennen (2013), Bayraksan and Love (2015), and Bertsimas et al. (2018), on the other hand, propose to choose the level of robustness based on the critical threshold of the corresponding hypothesis test. In a recent work, Gotoh, Kim, and Lim (2017) propose to choose the level of robustness by trading off between the mean and variance of the out-of-sample objective function value. We refer the readers to that paper for a review of calibration approaches in DRO. A common feature of many of these calibration methods is the assumption of i.i.d. data. However, in many real-life applications this assumption is violated, thereby rendering these approaches unsuitable. So, we do not rely on this assumption. Instead, we use the notions of maximal effective subsets and prices of optimism/pessimism and nominal/worst-case regrets to calibrate the level of robustness. Thus, our work can be valuable in many practical applications where the observations are not independent, the distribution of data changes over time, and other cases where data-driven methods may be unsuitable.

## 3. Problem formulation and assumptions

### 3.1. Problem setup

Consider the classical stochastic programming model for the single-product newsvendor problem

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

where

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

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

Throughout the paper, we make the following assumption on the problem parameters:

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

We do not impose any restriction on the sign of $V$. Assumption (A1) ensures two important features. First, it implies that $h$ is jointly convex in $x$ and $\xi$ on $\mathbb{R}\times\mathbb{R}$, so we have a convex optimization problem in all problem settings studied in this paper. Second, it ensures that the critical ratio $Q := \frac{U}{U+W}$ satisfies $0<Q<1$, so we have a well-defined solution to (Risk Neutral). We will recall the solution to (Risk Neutral) shortly.

In problem (Risk Neutral), the decision maker assumes that demand $\xi$ follows the continuous distribution $\mathbb{P}_0$. We adopt the following notation: $\Omega$ denotes the support of $\xi$, $\underline{\xi} := \inf\{\hat\xi : \hat\xi\in\Omega\}$, and $\overline{\xi} := \sup\{\hat\xi : \hat\xi\in\Omega\}$ with $0\le\underline{\xi}<\overline{\xi}$. We assume $\Omega$ is closed and bounded. We also assume $\mathbb{X} := \Omega$ for theoretical convenience, since it is never optimal to order less than the lowest demand or more than the highest demand.

From (1), we see that the compact support assumption of $\xi$ 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 $\mathbb{P}_0$: $F(t) := \mathbb{P}_0\{\xi\le t\}$. 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 newsvendor problem. We will use the notation $x^{\text{neut}}$, $Q$, and $F$ throughout the paper.

Consider the distributionally robust version of the newsvendor problem, following the motivation outlined in Section 1. In such a problem, the decision maker has some belief that $\mathbb{P}_0$ is the distribution of demand but would like to allow for perturbations of $\mathbb{P}_0$. Throughout the paper we refer to $\mathbb{P}_0$ as the nominal distribution and the set of possible perturbations of the nominal distribution $\mathbb{P}_0$ as 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, which is denoted by $\nu$. Let $p=\frac{d\mathbb{P}}{d\nu}$ denote the associated density function of $\mathbb{P}$ with respect to $\nu$. Similarly, $p_0=\frac{d\mathbb{P}_0}{d\nu}$ denotes the corresponding density function of $\mathbb{P}_0$. We also assume the support of $\mathbb{P}_0$ is $\Omega$, i.e., $p_0$ is strictly positive on $[\underline{\xi},\overline{\xi}]$. Given this setup, recall that the total variation distance between $\mathbb{P}$ and $\mathbb{P}_0$ is defined as $\frac{1}{2}\int_\Omega |p(s)-p_0(s)|\,ds$.

DRNV-V, a distributionally robust version of (Risk Neutral) formed via the total variation distance, can then be formulated as

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

where

$$\mathcal{P}_\gamma := \left\{ p : \frac{1}{2}\int_\Omega |p(s)-p_0(s)|\,ds \le \gamma,\ \int_\Omega p(s)\,ds = 1,\ p\ge 0 \right\}. \tag{2}$$

The ambiguity set of distributions $\mathcal{P}_\gamma$ contains all probability distributions $\mathbb{P}$ whose total variation distance to the nominal probability distribution $\mathbb{P}_0$ is limited by the level of robustness $\gamma$. The total variation distance in (2) has a maximum value of 1; therefore, $0\le\gamma\le1$ covers the whole spectrum[2]. In particular, when $\gamma=0$, we recover the classical SO model and when $\gamma=1$, we recover the classical RO model. 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 $x\in\mathbb{X}$. The worst-case expected value in (DRNV-V) can be written as the following risk measure

$$f_\gamma(x) = \gamma \operatorname{ess\,sup}_{\xi\in\Omega} h(x,\xi) + (1-\gamma)\text{CVaR}_\gamma[h(x,\xi)], \tag{3}$$

all with respect to the nominal distribution $\mathbb{P}_0$ (Jiang and Guan, 2018, Theorem 2). 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 $\text{VaR}_\rho[h(x,\xi)] := \inf\{u : \mathbb{P}_0\{h(x,\xi)\le u\}\ge\rho\}$ is the Value-at-Risk (VaR) at level $\rho$. Per usual convention, we set $\text{CVaR}_0[h(x,\xi)] := \mathbb{E}_{\mathbb{P}_0}[h(x,\xi)]$ and $\text{CVaR}_1[h(x,\xi)] := \operatorname{ess\,sup}_{\xi\in\Omega} h(x,\xi)$, where $\operatorname{ess\,sup}_{\xi\in\Omega} h(x,\xi) = \inf\left\{a\in\mathbb{R} : \mathbb{P}_0\{\xi\in\Omega : h(x,\xi)>a\}=0\right\}$. Because of the equivalence between the worst-case expected value in (DRNV-V) and (3), we refer to $\gamma$ as the level of robustness or level of risk-aversion interchangeably.

### 3.2. Conditions and examples

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

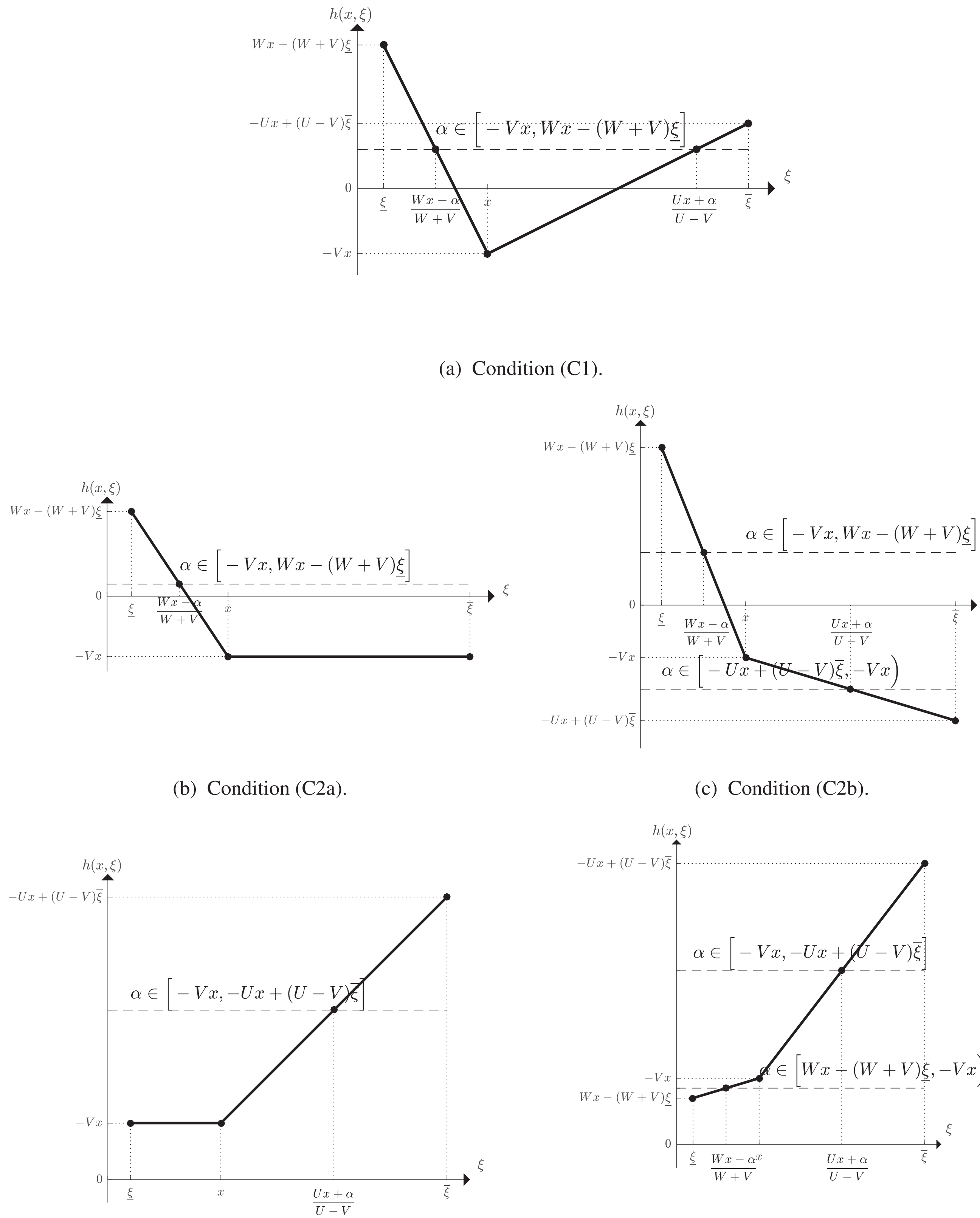
(**C1**) $W+V>0$ and $U-V>0$,
(**C2a**) $W+V>0$ and $U-V=0$,
(**C2b**) $W+V>0$ and $U-V<0$,
(**C3a**) $W+V=0$ and $U-V>0$,
(**C3b**) $W+V<0$ and $U-V>0$.

We refer to Conditions (C2a) and (C2b) collectively as Condition (C2). Similarly, we refer to Conditions (C3a) and (C3b) collectively as Condition (C3). Fig. 1 depicts the shape of $h(x,\cdot)$ under each condition. The differences between the shapes of $h(x,\cdot)$ affect our analysis of $\operatorname{ess\,sup}_{\xi\in\Omega} h(x,\xi)$ and $\text{CVaR}_\gamma[h(x,\xi)]$. These, in turn, play a crucial role in characterizing the optimal solution and critical demand regions under each condition.

Below, we present four examples that are covered by our setup 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, $h(x,\xi) := cx + b(\xi-x)_+ + m(x-\xi)_+$,

[2] It can be shown with the Minkowski inequality that the total variation distance between any two distributions is bounded above by 1 as follows: $\frac{1}{2}\int_\Omega |p(s)-p_0(s)|\,ds \le \frac{1}{2}(\int_\Omega |p(s)|\,ds + \int_\Omega |p_0(s)|\,ds) = \frac{1}{2}(\int_\Omega p(s)\,ds + \int_\Omega p_0(s)\,ds) = 1$.




(a) Condition (C1).

(b) Condition (C2a).

(c) Condition (C2b).

(d) Condition (C3a).

(e) Condition (C3b).

**Fig. 1.** Cost function $h(x,\cdot)$ under Conditions (C1)–(C3) for a fixed $x\in\Omega$. Observe that $-(W+V)$ is the slope of $h(x,\cdot)$ when $\xi<x$ and $U-V$ is the slope of $h(x,\cdot)$ when $\xi\geq x$.

Please cite this article as: H. Rahimian, G. Bayraksan and T. Homem-de-Mello, Controlling risk and demand ambiguity in newsvendor models, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2019.06.036

**Table 1**
Expressions for robust order quantity $x^{\text{rob}}$, critical level of robustness $\gamma^{\text{cr}}$, and parameter $\theta_\gamma$ under Conditions (C1)–(C3).

| Condition | $x^{\text{rob}}$ | $\gamma^{\text{cr}}$ | $\theta_\gamma$ |
|---|---|---|---|
| (C1) if $x^{\text{neut}} > x^{\text{rob}}$ | $\frac{W+V}{W+U}\underline{\xi} + \frac{U-V}{W+U}\overline{\xi}$ | $Q - F\left(\frac{W+U}{W+V}x^{\text{rob}} - \frac{U-V}{W+V}x^{\text{neut}}\right)$ | $\frac{W+V}{W+U}\left(\frac{x^{\text{neut}} - F^{-1}(Q-\gamma)}{x^{\text{neut}} - x^{\text{rob}}}\right)$ |
| (C1) if $x^{\text{neut}} < x^{\text{rob}}$ | | $F\left(\frac{W+U}{U-V}x^{\text{rob}} - \frac{W+V}{U-V}x^{\text{neut}}\right) - Q$ | $\frac{U-V}{W+U}\left(\frac{F^{-1}(Q+\gamma) - x^{\text{neut}}}{x^{\text{rob}} - x^{\text{neut}}}\right)$ |
| (C1) if $x^{\text{neut}} = x^{\text{rob}}$ | | 0 | 0 |
| (C2) | $\underline{\xi}$ | $Q$ | $\frac{x^{\text{neut}} - F^{-1}(Q-\gamma)}{x^{\text{neut}} - x^{\text{rob}}}$ |
| (C3) | $\overline{\xi}$ | $1-Q$ | $\frac{F^{-1}(Q+\gamma) - x^{\text{neut}}}{x^{\text{rob}} - x^{\text{neut}}}$ |

and one can show $W = c + m$, $U = b - c$, 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 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 $\xi$ 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). Conditions (C1) and (C2a) (e.g., Examples 1 and 2) are the most studied instances in the literature, e.g., Wang et al. (2016), Jiang and Guan (2018), Gotoh and Takano (2007), and 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).

## 4. Characterization of optimal solution

In this section, we characterize the optimal solution to (DRNV-V) as a function of the level of robustness[3], and discuss some of its properties. We analyze the sensitivity of this solution with respect to the model parameters in Section O-4 in the Online Supplement.

In our theorem below that characterizes the optimal order quantity $x^*_\gamma$ to (DRNV-V), three values will play an important role: (i) $x^{\text{rob}}$, referred to as *robust order quantity*, (ii) $\gamma^{\text{cr}}$, referred to as the *critical level of robustness*, and (iii) $\theta_\gamma$ that relates $x^*_\gamma$ to $x^{\text{neut}}$ and $x^{\text{rob}}$. These values are presented in Table 1. We state the theorem first; then we explain the role of each quantity.

[3] Note that $f_\gamma(x)$, as defined in (DRNV-V), is a continuous function on $\mathbb{R}$, and $\mathbb{X}$ is compact. Thus, by Weierstrass extreme value theorem, (DRNV-V) has a finite optimal solution.

**Theorem 1.** *Consider (DRNV-V) with cost function defined in (1), satisfying Assumption (A1). Let $x^{\text{rob}}$, $\gamma^{\text{cr}}$, and $\theta_\gamma$ be defined as in Table 1 according to Conditions (C1), (C2), and (C3). Then, for $0 \le \gamma < \gamma^{\text{cr}}$, there exists a unique optimal solution to (DRNV-V) given by*

$$x^*_\gamma = (1-\theta_\gamma)x^{\text{neut}} + \theta_\gamma x^{\text{rob}}.$$

*For $\gamma^{\text{cr}} \le \gamma \le 1$, $x^*_\gamma = x^{\text{rob}}$.*

To explain the theorem, first consider the case that the decision maker is extremely conservative, i.e., $\gamma = 1$. In this case, (DRNV-V) reduces to the following classical robust optimization model

$$\min_{x\in\mathbb{X}} \ \text{ess sup}_{\xi\in\Omega}\ h(x,\xi) \qquad \text{(Robust)}$$

with optimal solution $x^{\text{rob}}$, provided in Table 1. Under either (C1), (C2a), or (C3a), $x^{\text{rob}}$ is the order quantity where the costs at $\underline{\xi}$ and $\overline{\xi}$ equalize.

Second, the critical level of robustness $\gamma^{\text{cr}}$ is the smallest level of robustness $0 \le \gamma \le 1$ at which the optimal order quantity $x^*_\gamma$ to (DRNV-V) becomes the robust order quantity, $x^*_\gamma = x^{\text{rob}}$, and it remains optimal for larger values of $\gamma$. Note that $\gamma^{\text{cr}} < 1$ because $0 < Q < 1$. Finally, $\theta_\gamma$, which connects $x^*_\gamma$ to $x^{\text{neut}}$ and $x^{\text{rob}}$, is a nondecreasing, continuous, and not necessarily linear function in $\gamma$.

Putting these together, Theorem 1 implies that as the level of robustness increases, the optimal order quantity moves from the risk-neutral order quantity to the robust order quantity, either monotonically decreasing (if $x^{\text{neut}} > x^{\text{rob}}$, e.g., under (C2)) or monotonically increasing (if $x^{\text{neut}} < x^{\text{rob}}$, e.g., under (C3)). It eventually reaches $x^{\text{rob}}$ at $\gamma = \gamma^{\text{cr}}$ and stabilizes at $x^{\text{rob}}$ for $\gamma > \gamma^{\text{cr}}$.

In addition to optimal solution, we have also derived the optimal worst-case probability distribution. For brevity, we relegate this result and its proof to Section O-3.1 in the Online Supplement.

## 5. Characterization of maximal effective subsets

The (DRNV-V) yields the optimal order quantity for a given level of robustness $\gamma$, as studied in the previous section. But, how can the decision maker define a proper value for $\gamma$? It is intuitive that, the more risk-averse the decision maker is, the more (s)he is concerned about the values further in the tails of the demand distribution. Thus, our goal in this section is to precisely define the notion of "to be concerned about," which will allow us to directly express the critical regions in terms of the level of robustness $\gamma$, and vice versa.

The motivation for this study is twofold. First, this analysis can provide a deeper understanding of how risk and demand ambiguity affect the decisions. Second, it can provide a way to choose the level of robustness $\gamma$. In this section and the next one, we investigate two ways to achieve the second goal, both of which may be useful in situations where the assumption of data-driven methods are violated.

Data-driven approaches typically assume a set of i.i.d. sample data is available from the unknown distribution. In many situations, however, there is no guarantee that the future uncertainty is drawn from the same distribution. For instance, when a new product is launched, one can use historical data on similar products to construct the ambiguity set. The new product's demand, however, is unlikely to be distributed identically from this historical data. Similarly, in an influenza vaccine production problem, one can use historical data from previous years to form the ambiguity set. But, again, because influenza viruses mutate in unexpected ways from year to year, this data may violate the i.i.d assumption for the current year. These and many other important applications violate the typical assumptions of data-driven methods to choose $\gamma$.

In this section, we introduce the concept of critical regions of demand and interpret it as a way to choose the level of robustness $\gamma$. Here and in Section 6, we present our main results and demonstrate their applications in Section 7.

To be more focused on our goal, we present the results only under Condition (C1) in this section. All the results and proofs under Conditions (C2) and (C3) are relegated to the Online Supplement.

### 5.1. Definitions

Effective sets were first introduced in Rahimian et al. (2019) for finite $\Omega$. Here, we generalize that definition for the setting of continuous distributions in (DRNV-V) and introduce maximal effective subsets. As we shall see shortly, this requires dealing with new technical issues.

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. That is, if, for a given value of $\gamma$, the removal of $\mathcal{F}$ changes the optimal value, this means that the demand values in $\mathcal{F}$ are indeed critical for the problem at the decision maker's level of risk-aversion.

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

$$\min_{x \in \mathbb{X}} \left\{ f^{A}_{\gamma}(x; \mathcal{F}) := \sup_{p \in \mathcal{P}^{A}_{\gamma}(\mathcal{F})} \int_{\mathcal{F}^c} h(x, s) p(s)\, ds \right\}, \tag{4}$$

where

$$\mathcal{P}^{A}_{\gamma}(\mathcal{F}) := \left\{ p : \frac{1}{2} \int_{\mathcal{F}^c} |p(s) - p_0(s)|\, ds \le \gamma - \frac{1}{2} \int_{\mathcal{F}} p_0(s)\, ds, \right. \\ \left. \int_{\mathcal{F}^c} p(s)\, ds = 1, \ \ p \ge 0 \right\} \tag{5}$$

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 (4) is equal to $-\infty$ if the set $\mathcal{P}^{A}_{\gamma}(\mathcal{F})$ is empty. The ambiguity set $\mathcal{P}^{A}_{\gamma}(\mathcal{F})$ of the assessment problem is a rearrangement of $\mathcal{P}_\gamma \cap \{p : p = 0 \text{ on } \mathcal{F} \text{ Lebesgue-almost surely}\}$. Compare (5) to (2) and observe the adjustments to the variation distance constraint and the next constraint that ensure $p$ to be a probability density on $\mathcal{F}^c$.

It is worthwhile noting that, since $\mathcal{P}^{A}_{\gamma}(\mathcal{F}) \subseteq \mathcal{P}_\gamma$, then for any $x \in \mathbb{X}$ the optimal value of the inner problem in (4) is less than or equal to the optimal value of the inner problem in (DRNV-V). Thus, the optimal value of (4) is less than or equal to that of (DRNV-V).

We recall now the definition of effective sets for (DRNV-V) from Rahimian et al. (2019).

**Definition 1.** A subset $\mathcal{F} \subset \Omega$ is effective for (DRNV-V) if the optimal value of the corresponding assessment problem (4) is strictly smaller than the optimal value of (DRNV-V). A subset $\mathcal{F} \subset \Omega$ is called ineffective if it is not effective.

Intuitively, if the decision maker would not change their decision even if they were told that demand cannot lie within a certain range of values, then that range is ineffective.

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 (4) and (DRNV-V) are equal. Therefore, such $\mathcal{F}$ is ineffective according to Definition 1. 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. All definitions from this point on are new.

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

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

In other words, $\mathfrak{S}$ contains all finite unions of closed intervals on $\Omega$.

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 $\mathfrak{S}$.

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

In this paper, we aim to characterize a subset of $\Omega$ that is maximal effective. Roughly speaking, it is the largest[4] set such that any of its subsets is effective.

**Definition 3.** A subset $\mathcal{F} \subset \Omega$ is called maximal effective for (DRNV-V) 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 5.2).

### 5.2. Maximal effective subsets of DRNV-V

In this section, we first identify the maximal effective subsets of (DRNV-V) under Condition (C1). Then, we study the properties of these subsets in terms of the level of robustness.

To ease our presentation, we use the set $\mathcal{E}_\gamma$ to denote the maximal effective subset of (DRNV-V) for a given value of $\gamma$.

Also, for any $B \subset \Omega$, we use the shorthand notation $[\xi \in B]$ to represent the set $\{\xi \in \Omega : \xi \in B\}$.

First, observe that when $\gamma = 0$, the maximal effective subset is given by $\mathcal{E}_0 = \Omega$. This is because the ambiguity set (5) is empty for any subset $\mathcal{F}$ of $\Omega$ that belongs to $\mathfrak{S}$. Hence, the optimal value of the corresponding assessment problem (4) 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 2 presented below.

**Theorem 2.** *Consider (DRNV-V) with cost function defined in (1), $\gamma^{cr}$ defined in Table 1, and $x^*_\gamma$ defined in Theorem 1 as the optimal*

[4] With respect to Lebesgue measure.

*solution to (DRNV-V). Suppose Assumption (A1) holds. Then, for problem (DRNV-V) under Condition (C1), for all $0 < \gamma \le 1$, we have:*

- *if* $x^{\text{neut}} > x^{\text{rob}}$ *and* $0 < \gamma < \gamma^{\text{cr}}$, *then* $\mathcal{E}_\gamma = \left[\xi \le F^{-1}(Q-\gamma)\right] \cup [\xi \ge F^{-1}(Q)]$,
- *if* $x^{\text{neut}} < x^{\text{rob}}$ *and* $0 < \gamma < \gamma^{\text{cr}}$, *then* $\mathcal{E}_\gamma = [\xi \le F^{-1}(Q)] \cup [\xi \ge F^{-1}(Q+\gamma)]$,
- *if* $\gamma \ge \gamma^{\text{cr}}$, *then* $\mathcal{E}_\gamma = [\xi \le \frac{Wx^{\text{rob}} - \text{VaR}_\gamma[h(x^{\text{rob}},\xi)]}{W+V}] \bigcup [\xi \ge \frac{Ux^{\text{rob}} + \text{VaR}_\gamma[h(x^{\text{rob}},\xi)]}{U-V}]$.

Theorem 2 indicates that under (C1), both “low” and “high” demand regions are critical, and we see a different behavior once $\gamma$ crosses $\gamma^{\text{cr}}$. For instance, for $0 < \gamma < \gamma^{\text{cr}}$, the critical demand regions are determined by the $Q$-quantile on one side and an adjusted $(Q \pm \gamma)$-quantile on the other (the sign of the adjustment depends on the relationship between $x^{\text{neut}}$ and $x^{\text{rob}}$). For $\gamma \ge \gamma^{\text{cr}}$, the critical demand regions behave differently, changing with $\gamma$ according to the VaR of costs at stabilized solution $x^{\text{rob}}$, $\text{VaR}_\gamma[h(x^{\text{rob}}, \xi)]$. In particular, at $\gamma = 1$, $\text{VaR}_\gamma\left[h(x^{\text{rob}}, \xi)\right] = \text{ess sup}_{\xi \in \Omega} h(x^{\text{rob}}, \xi)$, and we have $\mathcal{E}_\gamma = \{\underline{\xi}, \overline{\xi}\}$.

Maximal effective subsets can also be interpreted in terms of the optimal worst-case probability distribution for a given level of robustness $\gamma$. In fact, as shown in Proposition O-7 in the Online Supplement, under certain conditions, the critical regions of demand given by the maximal effective subsets constitute the support of the worst-case probability distributions. Thus, values of demand outside the critical region are deemed unimportant because the worst-case probability assigns zero mass to such values.

Our final result shows that the sequence of maximal effective subsets $\mathcal{E}_\gamma$ are nested and monotonically non-increasing as $\gamma \to 1$. 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. (2019) show that such monotonicity and nestedness of effective subsets are not true in general (cf. Section 5.4 of that paper for a counterexample of a newsvendor problem with a finite support). However, we are able to obtain these properties for (DRNV-V).

**Theorem 3.** *Consider (DRNV-V) with cost function defined in (1), satisfying Assumption (A1). Then,*

(i) $\mathcal{E}_{\gamma_1} \supseteq \mathcal{E}_{\gamma_2}$ *for any* $0 \le \gamma_1 < \gamma_2 \le 1$.
(ii) *As* $\gamma \to 1$, *under Condition (C1),* $\mathcal{E}_\gamma$ *converges to* $\{\underline{\xi}, \overline{\xi}\}$.

Theorem 3 shows that the more risk-averse the decision maker, the smaller the size of the critical demand region. When $\gamma$=0, the maximal effective subset is $\Omega$.

As the level of risk-aversion increases, the critical demand region becomes smaller. This is because as the problem becomes more robust, it focuses more and more on the worst-case costs. When $\gamma = 1$, the maximal effective subset only contains the extreme points of $\Omega$. Using Theorems 2 and 3, the decision maker can then choose a value for the level of robustness $\gamma$ based on his/her perception of what the actual critical regions are. We illustrate this approach in Section 7 in the context of our application. In Section 8, we present how to use these ideas in the multi-product setting.

## 6. Price of optimism/pessimism and regrets

We now propose another way to choose the level of robustness $\gamma$ by quantifying the relationships between DRO and the classical SP and RO models. In particular, we propose measures to evaluate the performance of optimal solution to DRO in the classical SP and the classical RO models and vice versa. The motivation for this approach is as follows. In case of full trust in the nominal distribution, one may pick $\gamma = 0$ (the SP model), and in the case of no trust (and most conservatism), one may pick $\gamma = 1$ (the RO model). In the absence of full trust—but *some* trust—and knowing that data-driven assumptions are violated, one may choose to balance regrets and prices relative to the full- and no-trust models. We shall see shortly that these measures may help decision makers understand how valuable these different optimal solutions are and choose an appropriate level of robustness.

Recall the function $f_\gamma(x)$ for a fixed $x \in \mathbb{X}$ and $\gamma$ in (DRNV-V). For $0 \le \gamma \le 1$, the following two sets of inequalities hold:

$$f_\gamma(x^{\text{neut}}) \ge f_\gamma(x^*_\gamma) \ge f_0(x^*_\gamma) \ge f_0(x^{\text{neut}}), \tag{6}$$

$$f_\gamma(x^*_\gamma) \le f_\gamma(x^{\text{rob}}) \le f_1(x^{\text{rob}}) \le f_1(x^*_\gamma). \tag{7}$$

In (6), $f_\gamma(x^*_\gamma) \ge f_0(x^*_\gamma)$ because for any $0 < \gamma \le 1$, 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 quantities in (6) are equal. Similarly in (7), $f_1(x^{\text{rob}}) \ge f_\gamma(x^{\text{rob}})$ because, for $\gamma = 1$, the worst-case expected problem at $x \in \mathbb{X}$ is a relaxation of the worst-case expected problem at $x$ for any $0 \le \gamma < 1$. At $\gamma = 1$, all quantities in (7) are equal. The other inequalities in (6) and (7) are justified by suboptimality.

To conduct our analysis, we define the following measures for $0 \le \gamma \le 1$:

$$\text{PO}_\gamma := f_\gamma(x^{\text{neut}}) - f_\gamma(x^*_\gamma), \qquad \text{PP}_\gamma := f_\gamma(x^{\text{rob}}) - f_\gamma(x^*_\gamma), \tag{8}$$

$$\text{NR}_\gamma := f_0(x^*_\gamma) - f_0(x^{\text{neut}}), \qquad \text{WR}_\gamma := f_1(x^*_\gamma) - f_1(x^{\text{rob}}). \tag{9}$$

The first equation in (8) 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. Similarly, the second equation in (8) measures the *Price of Pessimism (PP)*—what we lose by being overly conservative when (DRNV-V) accurately represents the ambiguity in the distribution.

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 (9) 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 (9) measures the *Worst-case Regret (WR)*—what we lose by using (DRNV-V) with $\gamma < 1$ 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 & Moon, 1993; Perakis & 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

$$\text{PO}_\gamma - \text{PP}_\gamma = f_\gamma(x^{\text{neut}}) - f_\gamma(x^{\text{rob}}). \tag{10}$$

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 solutions 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 $\gamma^{\text{S}}$.

Similarly,

$$\text{NR}_\gamma - \text{WR}_\gamma = \left(f_0(x^*_\gamma) - f_0(x^{\text{neut}})\right) - \left(f_1(x^*_\gamma) - f_1(x^{\text{rob}})\right). \tag{11}$$

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 regrets 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 $\gamma^{\mathrm{D}}$.

In the following theorem, we show that the indifference levels of robustness are well defined for (DRNV-V). We further elaborate on these notions and comment on how they can be used to choose an appropriate level of robustness in Section 7. In Section 8, we present a partial extension and numerically illustrate how to use these notions in the multi-product setting.

**Theorem 4.** *Consider (DRNV-V) with cost function defined in (1) and the price of optimism, price of pessimism, nominal and worst-case regrets defined in (8)–(9). Let* $\gamma^{\mathrm{S}} = \min\{\gamma \in [0,1] : \mathrm{PO}_\gamma - \mathrm{PP}_\gamma = 0\}$ *be the indifferent-to-solution level of robustness and* $\gamma^{\mathrm{D}} = \min\{\gamma \in [0,1] : \mathrm{NR}_\gamma - \mathrm{WR}_\gamma = 0\}$ *be the indifferent-to-distribution level of robustness. Suppose Assumption (A1) holds. Then,* $\gamma^{\mathrm{S}}$ *and* $\gamma^{\mathrm{D}}$ *are well defined for problem (DRNV-V), and both are smaller than or equal to the critical level of robustness* $\gamma^{\mathrm{cr}}$.

Observe that the prices and regrets can be calculated using the analytical expressions in Theorem 1 and numerical evaluation of $f$ based on (3). Hence, the indifference levels of robustness can be obtained fairly easily by a numerical (root-finding) method.

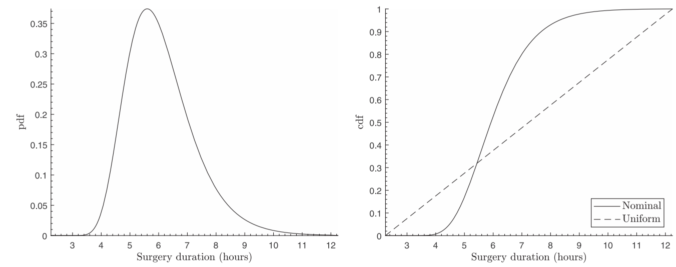
## 7. Application to operating room time reservation

We now apply the ideas set forth in the paper to an operating room (OR) time reservation problem, motivated by Olivares et al. (2008).

Reserving OR time for each surgery case is a task performed as part of the OR management problem, typically during the preoperative planning phase (Gupta, 2007; May, Spangler, Strum, & Vargas, 2011; Samudra et al., 2016). Reserving OR time involves balancing the under- and over-utilization costs of a hospital's surgical capacities. Thus, the newsvendor model can be used for this purpose, and there exists significant evidence in the medical literature that this model is often used in practice to reserve OR time; see, e.g., Lehtonen, Torkki, Peltokorpi, and Moilanen (2013), Olivares et al. (2008), and Wachtel and Dexter (2010).

When deciding how much OR time to reserve for a surgery case, the decision maker can use historical data in combination with surgeons' estimate of surgery duration. These estimations, however, may not be reliable due to two reasons: (i) data scarcity and (ii) data mistrust. First, too few relevant historical cases may be available (Dexter, Traub, Fleisher, & Rock, 2002). When surgical procedure and surgeon—the two most important factors that determine surgery duration (Strum, Sampson, May, & Vargas, 2000b; Zhou, Dexter, Macario, & Lubarsky, 1999)—are fixed, only 5 or fewer data points are available in approximately half of the surgeries performed in the U.S. (Macario, 2009). It is well known in the medical literature that pooling data from several hospitals, multiple years, and similar procedures usually does not improve the estimation because they pose the risk that surgical durations are confounded by other variables (Macario, 2010; Zhou et al., 1999). Even when such data is available, the sample is not necessarily identically distributed due to patient characteristics and case severity (Olivares et al., 2008), thereby violating data-driven assumptions.

On the data mistrust issues, some papers have reported that surgical services deviate from acting according to economic incentives and show systematic biases in estimating surgery durations




(a) The pdf. (b) The cdf.

**Fig. 2.** The pdf and cdf of the nominal distribution of an average surgery duration, supported between (2.25, 12.25).

**Table 2**
Problem parameters.

| $W$ | $U$ | $V$ | $\rho$ | Condition | Distribution |
|---|---|---|---|---|---|
| 0.5 | 1 | 0 | 0.5 | (C1) | $2.25 + \log\mathcal{N}(1.303, 0.0922, 0, 10)$ |

(Dexter, Macario, Epstein, & Ledolter, 2005; Fügener, Schiffels, & Kolisch, 2017; Macario, 2009; Olivares et al., 2008). For instance, some surgeons systematically underestimate their case durations to get their cases to fit into their allocated OR time. Similarly, some surgeons intentionally overestimate their case durations to keep control of their allocated OR time.

The lack of data and lack of trust in data imply there is ambiguity about the underlying distribution of surgery duration. As a result, DRNV is a more realistic approach to study the OR time reservation problem. In this paper, we use (DRNV-V) to model this problem.

### 7.1. Problem parameters

We conducted our numerical experiments based on the OR time reservation problem described in Olivares et al. (2008) for cardiac surgery cases[5]. There is one OR with regular time from 7:30 A.M. to 7 P.M. (11.5 hours). This application is modeled as a newsvendor satisfying (C1) with $V = 0$. We will discuss specific parameters next.

It is commonly believed in the medical literature that *log-normal distribution* provides a good statistical fit for the surgery duration; see, e.g., Strum, May, and Vargas (2000a). However, Olivares et al. (2008) conclude cardiovascular surgery procedures are not well fitted by the standard log-normal distribution because they are much longer than general surgery cases. To remedy this statistical issue they use a shift parameter of $\delta \simeq 2.25$ hours, so that $\xi - \delta$ follows a log-normal distribution. We estimated[6] the parameters of this log-normal to be $\mu = 1.303$ and $\sigma^2 = 0.0922$.

Because our model requires bounded distributions, we truncated this distribution at $\mathrm{VaR}_{0.9995}[\xi - \delta] \simeq 10$ hours. As a result, we obtain $\xi - \delta \sim \log\mathcal{N}(1.303, 0.0922, 0, 10)$.

For reference, we plot the probability density function (pdf) and cdf of the nominal distribution of surgery duration in Fig. 2. This distribution is a shifted, truncated log-normal supported on the interval (2.25, 12.25) hours.

Next, we need to choose parameters $U$ and $W$. Let us denote the cost ratio $\frac{W}{U}$ by $\rho$. Olivares et al. (2008) estimate[7] an average

[5] Olivares et al. (2008) study an inverse optimization approach to estimate the unobserved ratio between the overage and underage costs based on 258 cardiac surgery cases and their characteristics. To conduct our experiments, we estimated necessary parameters from their results.

[6] This is the distribution of an average surgery duration obtained from models N1 and N2 in Olivares et al. (2008).

[7] Model N4 in Olivares et al. (2008).

Please cite this article as: H. Rahimian, G. Bayraksan and T. Homem-de-Mello, Controlling risk and demand ambiguity in newsvendor models, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2019.06.036

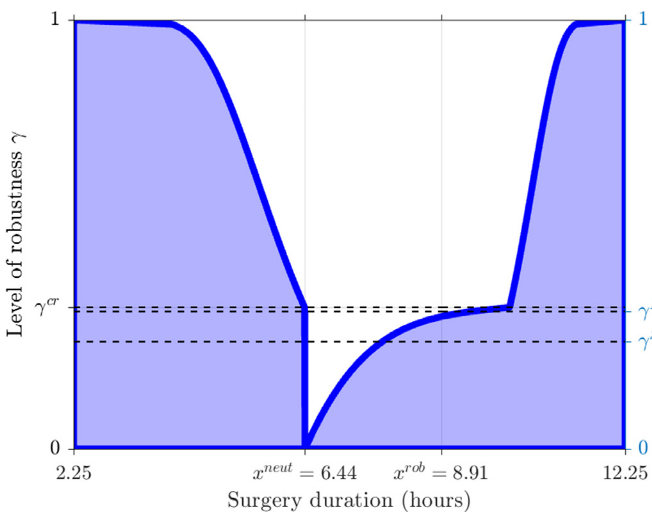



**Fig. 3.** Maximal effective subsets.

cost ratio $\rho$ for a surgery case to be 0.5; so, we use $\rho = 0.5$ for our analysis. It can be verified that with a fixed $\rho$, our main results are insensitive to the magnitudes of $U$ and $W$ when $V = 0$. So, without loss of generality, we assume $U = 1$ and $W = 0.5$. Given our choice of the nominal distribution $\mathbb{P}_0$ and the cost ratio $\rho$, this application satisfies (C1) with $x^{\text{neut}} = 6.44 < x^{\text{rob}} = 8.91$. Table 2 summarizes the problem parameters.

We now remark on why $x^{\text{neut}} < x^{\text{rob}}$ for this application. One can interpret $x^{\text{rob}} = \frac{\rho\underline{\xi}+\overline{\xi}}{1+\rho}$ as $U^{-1}_{[\underline{\xi},\overline{\xi}]}(\frac{1}{1+\rho})$, where $U_{[\underline{\xi},\overline{\xi}]}(\cdot)$ denotes the cdf of a uniform distribution on $[\underline{\xi},\overline{\xi}]$ (as opposed to the nominal distribution $F$). On the other hand, because $Q = \frac{1}{1+\rho}$, it follows that $x^{\text{neut}} = F^{-1}(Q)$ is greater than the median of the surgery duration for all $\rho < 1$, including $\rho = 0.5$. Observe from Fig. 2 that the right-skewed nominal distribution of surgery duration is (strictly) stochastically dominated by a uniform distribution on [2.25,12.25] for values greater than the median. Consequently, $F^{-1}(\frac{1}{1+\rho}) < U^{-1}_{[\underline{\xi},\overline{\xi}]}(\frac{1}{1+\rho})$, i.e., $x^{\text{neut}} < x^{\text{rob}}$, holds for all $\rho < 1$, including $\rho = 0.5$. Intuitively, this is a consequence of longer surgery durations being more costly than shorter durations (i.e., $\rho < 1$) plus the right-skewness of the nominal distribution toward longer surgeries.

### 7.2. Maximal effective subsets

The shaded area in Fig. 3 depicts the maximal effective subsets for $0 \le \gamma \le 1$. This figure reveals several interesting properties. First, the region of maximal effective subset moves away from $x^{\text{neut}}$ as $\gamma$ increases. This is predicted by Theorem 2, but it can be interpreted as follows. For the studied problem, the goal of the decision maker is to find a trade-off between the costs of under- and over-utilization so as to balance the costs of surgery durations that are too low or too high. Of course, "too low" and "too high" are relative terms, which depend on the decision maker's level of risk-aversion (or, equivalently, level of robustness). The less conservative the decision maker is, the farther such thresholds are from the lowest and highest surgery durations, until of course they coincide—which is precisely where $x^{\text{neut}}$ is attained.

Second, for almost all levels of robustness, surgery durations approximately smaller than 4 hours and greater than 11.5 hours are in the maximal effective subset for the available data.

This can be explained as follows. Surgery durations less than 4 hours or longer than 11.5 for this application are rare according to the nominal distribution (see Fig. 2). Moreover, we know from Theorem 2 that, for larger values of $\gamma$ (more specifically, $\gamma \ge \gamma^{\text{cr}}$), the critical region is dictated by $\text{VaR}_\gamma[h(x^{\text{rob}}, \xi)]$. Since $h$ is piecewise linear in $\xi$, it follows that this region is largely determined by the tails of the distribution of surgery durations $\xi$. This means that rare events in the tails of $h(x^{\text{rob}}, \xi)$—and therefore, of $\xi$—will be in the maximal effective subset for almost all values of $\gamma$, except for those very close to 1.

Third, recall from Theorem 3 that the size of maximal effective subsets is inversely related to the level of risk-aversion.

As predicted by Theorem 2, for all $0 < \gamma < \gamma^{\text{cr}}$, the model's risk attitude toward surgery durations below $x^{\text{neut}}$ is the same. That is, as Fig. 3 depicts, for this range of $\gamma$'s, the left piece of the maximal effective subsets are the same. However, the model increasingly adjusts the optimal solution—and hence the critical regions through $(Q + \gamma)$-quantile adjustment—by being more risk averse against longer surgeries. In other words, surgery durations in the right tail are more critical than the surgery durations in the left tail for the decision maker's level of risk-aversion and costs.

In this sense, our results on maximal effective subsets are consistent with the OR management literature that suggests schedule overrun and long working hours are more critical to reserving OR time than idle capacity; see, e.g., Strum, Vargas, and May (1999). These results imply predictable work hours and schedule stability are key drivers of employees' satisfaction and wellness in the healthcare industry.

Let us now discuss how to use maximal effective subsets to determine $\gamma$. Decision makers can infer their risk attitude by choosing a value of $\gamma$ for which the corresponding maximal effective subset is closest to their evaluation of critical surgery durations. Because the optimal solution stabilizes after $\gamma^{\text{cr}}$, we recommend a value below $\gamma^{\text{cr}}$. Also, because large surgery durations are relatively more costly and the model is equally risk-averse to durations below $x^{\text{neut}}$ for $0 < \gamma < \gamma^{\text{cr}}$, we ask the decision maker what durations on the right of $x^{\text{neut}}$ are critical. If, for instance, the decision maker deems surgeries longer than 8 hours as critical, then we recommend $\gamma \simeq 0.31$ from Fig. 3. Once $\gamma$ is chosen, the decision maker can then use $x^*_\gamma$ for the OR reservation time; for example, with $\gamma = 0.31$ we have $x^*_{0.31} = 8.12$.

### 7.3. Price of optimism/pessimism and regrets

Fig. 4 shows the price of optimism, price of pessimism, and regrets for the studied problem. Several properties can be seen from this figure.

First, the price of optimism and nominal regret are nondecreasing in $\gamma$, whereas the price of pessimism and worst-case regret are non-increasing in $\gamma$.

Second, even the most conservative decision maker has no reason to choose a level of robustness higher than $\gamma^{\text{cr}}$. For $\gamma \ge \gamma^{\text{cr}}$, the price of pessimism and worst-case regret are zero because $x^*_\gamma$ is stabilized at $x^{\text{rob}} = x^*_1$. On the other hand, the nominal regret is a constant in these cases because $f_0(x^*_\gamma)$ is a constant.

Price of optimism/pessimism and regrets can also help the decision maker choose the value of $\gamma$ based on the indifferent-to-solution and indifferent-to-distribution levels of robustness (recall these notions from Section 6). For the studied problem, we have $0.25 \simeq \gamma^{\text{S}} < \gamma^{\text{D}} \simeq 0.32$, where $\gamma^{\text{D}}$ is very close to $\gamma^{\text{cr}} \simeq 0.33$ (see also Fig. 3). If the decision maker wants to be indifferent regarding the error from using either the robust or the risk-neutral order quantities, then (s)he chooses $\gamma^{\text{S}}$. 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 (s)he chooses $\gamma^{\text{D}}$. In this particular application, the indifference levels of robustness are close to each other and close to $\gamma^{\text{cr}}$. This suggests a small range of choices ($\gamma \in [0.25, 0.33]$) for the decision maker if such criteria are used to find a desired value of $\gamma$.

Please cite this article as: H. Rahimian, G. Bayraksan and T. Homem-de-Mello, Controlling risk and demand ambiguity in newsvendor models, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2019.06.036

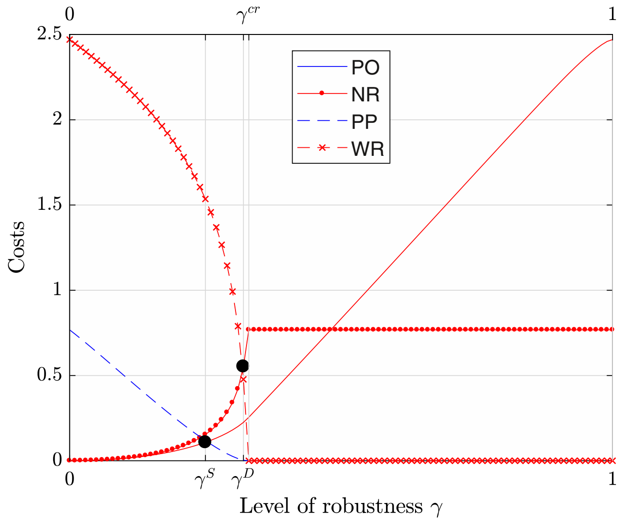



**Fig. 4.** Price of optimism, price pessimism, nominal regret, and worst-case regret. *Notes*: If the magnitude of $U$ (and $W$) changes while $\rho$ is kept at 0.5, then all values on the y-axis should be scaled by the magnitude of $U$ (or $W$), but $\gamma^S$ and $\gamma^D$ remain the same.

## 8. Extension to the multi-product setting

In the previous sections, we concentrated on a single-product setting. In this section, we extend the results in Sections 3–6 to a multi-product newsvendor problem. Throughout this section, we denote a vector by a boldface symbol. For example, a vector $[x_1, \ldots, x_k]^\top$ is denoted by $\boldsymbol{x} \in \mathbb{R}^k$. In order to refer to a particular product $i$, we use subscript $i$. For example, $x_i^{\text{neut}}$ and $x_i^{\text{rob}}$ denote the risk-neutral and robust order-quantities for product $i$, $i = 1, \ldots, k$, obtained by solving the corresponding single-product problems (Risk Neutral) and (*Robust*), respectively. On the other hand, we use superscript M to denote solutions in the multi-product setting. For instance, $\boldsymbol{x}_\gamma^{\mathrm{M}*}$ denotes the optimal order quantity vector in the multi-product setting. Otherwise, we use the exact same notation as before to refer to a joint property of the products. For example, $\mathbb{P}_0$ and $p_0$ denote the joint nominal probability distribution of $\boldsymbol{\xi}$ and its associated density function, respectively.

To introduce the problem, consider $k$ products, with parameters $W_i$, $U_i$, and $V_i$. Let

$$h(\boldsymbol{x}, \boldsymbol{\xi}) := \sum_{i=1}^{k} h_i(x_i, \xi_i) \tag{12}$$

represent the total net loss of the newsvendor for a vector of fixed order quantities $\boldsymbol{x} \in \mathbb{R}^k$ and a vector of uncertain demand realizations $\boldsymbol{\xi} \in \mathbb{R}^k$, where

$$h_i(x_i, \xi_i) := W_i(x_i - \xi_i)_+ + U_i(\xi_i - x_i)_+ - V_i\xi_i$$

denotes the total net loss of the newsvendor for product $i$. We let $\mathbb{X}$ denote the feasibility set for the decision vector $\boldsymbol{x}$. Other constraints characterizing problem characteristics can be added to the problem through $\mathbb{X}$.

We make the following assumption on the parameters

(**A2**) $U_i > 0$ and $W_i > 0$ for $i = 1, \ldots, k$,

and suppose that the support $\Omega$ of $\boldsymbol{\xi}$ has the following property

(**A3**) $\Omega = \Omega_1 \times \cdots \times \Omega_k$,

where $\Omega_i$ is the compact support of $\xi_i$. Then, a distributionally robust version for the multi-product newsvendor problem formed via the total variation distance can be formulated as

$$\min_{\boldsymbol{x} \in \mathbb{X}} \left\{ f_\gamma(\boldsymbol{x}) := \sup_{\mathbb{P} \in \mathcal{P}_\gamma} \mathbb{E}_{\mathbb{P}}\left[h(\boldsymbol{x}, \boldsymbol{\xi})\right] \right\}, \tag{13}$$

where

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

Distributionally robust multi-product newsvendor problems are studied in the literature, see, e.g., Hanasusanto, Kuhn, Wallace, and Zymler (2015), Ardestani-Jaafari and Delage (2016), and Natarajan et al. (2018) for models formed via moment-based ambiguity sets, and Bertsimas et al. (2018) for models formed via distance-based ambiguity sets. These DRO models are usually reformulated or approximated as a linear program, second-order cone program, or semi-definite program, depending on the structure of the ambiguity set. Risk-averse multi-product newsvendor problems are also studied in Tomlin and Wang (2005), Van Mieghem (2007), Gotoh and Takano (2007), Chen and Sim (2009), and Choi et al. (2011).

### 8.1. Main results

We investigate problem (13) in further detail by studying its optimal solution, maximal effective subsets, and price of optimism/pessimism and regrets.

As pointed out in the literature, closed-form solutions to the distributionally robust and/or risk-averse multi-product newsvendor problems do not exist in general and the solutions to these problems are usually obtained numerically (Choi et al., 2011; Tomlin & Wang, 2005). Note that (13) has a similar form as (DRNV-V). Thus, the result in (3) follows similarly, albeit with the cost function $h(\boldsymbol{x}, \boldsymbol{\xi})$ defined in (12). Assuming that $\Omega$ is finite (or approximated with finitely many realizations $\{\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^m\}$), one can reformulate (or approximate) (13) as follows

$$\min_{\boldsymbol{x}, \alpha, t, \boldsymbol{\eta}} \ \gamma t + (1-\gamma)\alpha + \sum_{j=1}^{m} p_0^j \eta^j \tag{14a}$$

$$\text{s.t.} \quad t - \sum_{i=1}^{k} \left\{ W_i u_i^j + U_i v_i^j \right\} \ge -\sum_{i=1}^{k} V_i \xi_i^j, \quad j = 1, \ldots, m, \tag{14b}$$

$$x_i - u_i^j + v_i^j = \xi_i^j, \quad i = 1, \ldots, k,\ j = 1, \ldots, m, \tag{14c}$$

$$\eta^j - \sum_{i=1}^{k} \left\{ W_i u_i^j + U_i v_i^j \right\} + \alpha \ge -\sum_{i=1}^{k} V_i \xi_i^j, \quad j = 1, \ldots, m, \tag{14d}$$

$$\boldsymbol{x} \in \mathbb{X}, \tag{14e}$$

$$\boldsymbol{\eta} \ge \boldsymbol{0}, \tag{14f}$$

where $p_0^j$ denotes the nominal probability of the event $\{\boldsymbol{\xi} = \boldsymbol{\xi}^j\}$, $j = 1, \ldots, m$. The key to the reformulation (14) is to substitute $\mathrm{CVaR}_\gamma[h(\boldsymbol{x}, \boldsymbol{\xi})]$ with a univariate convex optimization problem as follows due to Rockafellar and Uryasev (2002, Theorem 10)

$$\mathrm{CVaR}_\gamma\left[h(\boldsymbol{x}, \boldsymbol{\xi})\right] = \min_\alpha \left\{ \alpha + \frac{1}{1-\gamma} \sum_{j=1}^{m} p_0^j \left( h(\boldsymbol{x}, \boldsymbol{\xi}^j) - \alpha \right)_+ \right\}.$$

Moreover, because $h(\boldsymbol{x}, \cdot)$ is continuous in $\boldsymbol{\xi}$, we replaced $\operatorname{ess\,sup}_{\boldsymbol{\xi} \in \Omega} h(\boldsymbol{x}, \boldsymbol{\xi})$ by $\sup_{\boldsymbol{\xi} \in \Omega} h(\boldsymbol{x}, \boldsymbol{\xi})$ due to Phu and Hoffmann (1996, Proposition 3.5).

If $\mathbb{X}$ is a Ployhedral set, the above formulation is a linear program, and one can obtain the optimal order quantities $\boldsymbol{x}_\gamma^{\mathrm{M}*}$ to (13) efficiently. Alternatively, if the number of realizations is large, one can use a variant of the L-shaped method capable of handling the CVaR function to solve the problem. We refer to Zhang et al.

(2016) for a survey of different decomposition schemes to handle the CVaR function.

We begin our main results by studying the properties of the optimal solution $\boldsymbol{x}_\gamma^{\mathrm{M}*}$ to (13). We make the following additional assumption

(**A4**) $\mathbb{X}$ is separable in products, e.g., $\mathbb{X} = \{\boldsymbol{x} : \boldsymbol{x} \geq \boldsymbol{0}\}$.

Similar to the single-product setting, let $\boldsymbol{x}^{\mathrm{Mneut}}$ and $\boldsymbol{x}^{\mathrm{Mrob}}$ denote the risk-neutral order quantity and robust order quantity to (13), when $\gamma = 0$ and $\gamma = 1$, respectively.

**Theorem 5.** *Consider (13) with cost function defined in (12), satisfying Assumptions (A2)–(A4). Let $x_i^{\mathrm{neut}}$ denote the risk-neutral order quantity for product i, $i = 1, \ldots, k$. Moreover, let $x_i^{\mathrm{rob}}$ and $\gamma_i^{\mathrm{cr}}$ be defined as in Table 1 according to Conditions (C1), (C2), and (C3), for $i = 1, \ldots, k$.*

*Then,*

(i) *When $\gamma = 0$, (13) is separable in products and there exists a unique optimal solution to (13) given by $\boldsymbol{x}^{\mathrm{Mneut}} = [x_1^{\mathrm{neut}}, \ldots, x_k^{\mathrm{neut}}]$.*
(ii) *When $\gamma = 1$, (13) is separable in products and there exists a unique optimal solution to (13) given by $\boldsymbol{x}^{\mathrm{Mrob}} = [x_1^{\mathrm{rob}}, \ldots, x_k^{\mathrm{rob}}]$.*

Note that (13) is not necessarily separable in products for $0 < \gamma < 1$ even if Assumption (A4) holds. Under Assumption (A3), $\sup_{\Omega} \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\xi})$ is separable in products. However, $\mathrm{CVaR}_\gamma[\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\xi})]$ is not necessarily separable in products even when we have identical products with mutually independent demands. Moreover, if $\mathbb{X}$ is not separable in products, e.g., there is a budget constraint of form $\sum_{i=1}^{k} x_i \leq b$, then problem (13) is not necessarily separable in products even if $\gamma = 0$ or $\gamma = 1$.

Now, we turn our attention to the maximal effective subsets of (13). Definition 1 in Section 5.1 is applicable to (13) and its corresponding assessment problem. We do not have closed-form analytical expressions for the maximal effective subsets of (13) as in Section 5.2. However, for our numerical analysis in Section 8.2, we use a discretization of the support $\Omega$ that satisfies Assumption (A3) and utilize the techniques in Rahimian et al. (2019) to identify the effective scenarios.

Finally, as we mentioned before, the notions of price of optimism/pessimism and nominal/worst-case regrets, defined in Section 6, hold for a general DRO problem. In particular, they hold for problem (13) and its associated SO and RO models. Moreover, because $\boldsymbol{x}^{\mathrm{Mneut}}$ and $\boldsymbol{x}^{\mathrm{Mrob}}$ are unique, one can obtain the indifferent-to-solution level of robustness for problem (13) using (10). However, because we do not have a result on the uniqueness of $\boldsymbol{x}_\gamma^{\mathrm{M}*}$, we cannot guarantee that a unique indifferent-to-distribution level of robustness can be obtained using (11). As a result, a partial analogue of Theorem 4 holds for (13).

**Theorem 6.** *Consider (13) with cost function defined in (12), and the price of optimism and price of pessimism defined in a similar fashion to (8). Let $\gamma^{\mathrm{S}} = \min\{\gamma \in [0, 1] : \mathrm{PO}_\gamma - \mathrm{PP}_\gamma = 0\}$ be the indifferent-to-solution level of robustness. Suppose Assumptions (A2)–(A4) hold. Then, $\gamma^{\mathrm{S}}$ is well defined for problem (13), and is smaller than or equal to the critical level of robustness $\gamma^{\mathrm{cr}}$.*

### 8.2. Numerical experiments

To illustrate the results, we suppose that there are two products, i.e., $k = 2$, and we consider two problems. For both problems, the demand of product 1 and 2 are independent from each other, and we have $V_1 = V_2 = 0$. Similarly to the single-product setting, our main results are insensitive to the magnitudes of $U_i$ and $W_i$ when $V_i = 0$. We assume the marginal nominal probability distribution of each product follows a normal distribution, truncated on [0,4]. We approximated the joint nominal probability distribution on the grid $[0, 4] \times [0, 4]$ by $101 \times 101 = 10,201$ points. Table 3 summarizes the parameters for both problems. In problem 1, both products have identical parameters and identical marginal nominal distribution, whereas in problem 2, both products have identical parameters and different marginal nominal distributions. Given our choice of the nominal distribution $\mathbb{P}_0$ and the parameters, for problem 1, we have $x_i^{\mathrm{neut}} < x_i^{\mathrm{rob}}$, $i = 1, 2$, and for problem 2, we have $x_1^{\mathrm{neut}} > x_1^{\mathrm{rob}}$ and $x_2^{\mathrm{neut}} < x_2^{\mathrm{rob}}$. In this section, we report the results for the problems listed in Table 3, with specific parameters and with independent demands. However, we observed similar trends for the optimal order quantities, maximal effective subsets, prices/regrets, and the indifference levels of robustness for other choices of the parameters and correlated demands.

To obtain optimal order quantities $\boldsymbol{x}_\gamma^{\mathrm{M}*}$, $\gamma \in [0, 1]$, we solved the linear program (14). We observed that each component of $\boldsymbol{x}_\gamma^{\mathrm{M}*}$ shows a similar behavior to that of the solution to the corresponding single-product (DRNV-V). That is, for problem 1, $x_{i,\gamma}^{\mathrm{M}*}$ monotonically increases from $x_i^{\mathrm{neut}}$ to $x_i^{\mathrm{rob}}$, $i = 1, 2$. On the other hand, for problem 2, $x_{1,\gamma}^{\mathrm{M}*}$ monotonically decreases from $x_1^{\mathrm{neut}}$ to $x_1^{\mathrm{rob}}$, while $x_{2,\gamma}^{\mathrm{M}*}$ monotonically increases from $x_2^{\mathrm{neut}}$ to $x_2^{\mathrm{rob}}$. As a result, there exists a critical level of robustness at which the optimal order quantity $\boldsymbol{x}_\gamma^{\mathrm{M}*}$ stabilizes at the robust order quantity $\boldsymbol{x}^{\mathrm{Mrob}}$. For problem 1, the critical level of robustness $\gamma^{\mathrm{cr}} \simeq 0.29$, whereas for problem 2, the critical level of robustness $\gamma^{\mathrm{cr}} \simeq 0.60$. It is interesting to observe that for both problems (and other problems we tested), the critical level of robustness $\gamma^{\mathrm{cr}}$ is equal to $\max_{i=1}^{k} \gamma_i^{\mathrm{cr}}$.

The closure of the colored plots in Fig. 5 depicts the maximal ineffective subsets (i.e., the complement of the maximal effective subsets) for selected values of $\gamma \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$. In other words, the colored plots form the boundaries between the maximal effective subsets and maximal ineffective subsets. Several comments are in order about the maximal effective subsets. It can be seen that, similar to the single-product setting, the maximal effective subsets are non-increasing as $\gamma$ increases. More interestingly, a similar behavior to what we predicted in Theorem 2 and depicted in Fig. 3 for the case $x^{\mathrm{neut}} < x^{\mathrm{rob}}$ happens for these two problems. Recall that in Fig. 3, the left piece of the maximal effective subsets were the same for all $0 < \gamma < \gamma^{\mathrm{cr}}$. As it is shown in Fig. 5, for the multi-product setting, there exists a face to the boundary of the maximal ineffective subsets that remains the same for all $0 < \gamma < \gamma^{\mathrm{cr}}$. If $\gamma > \gamma^{\mathrm{cr}}$, then the maximal ineffective subset starts growing in the direction normal to that face as well.

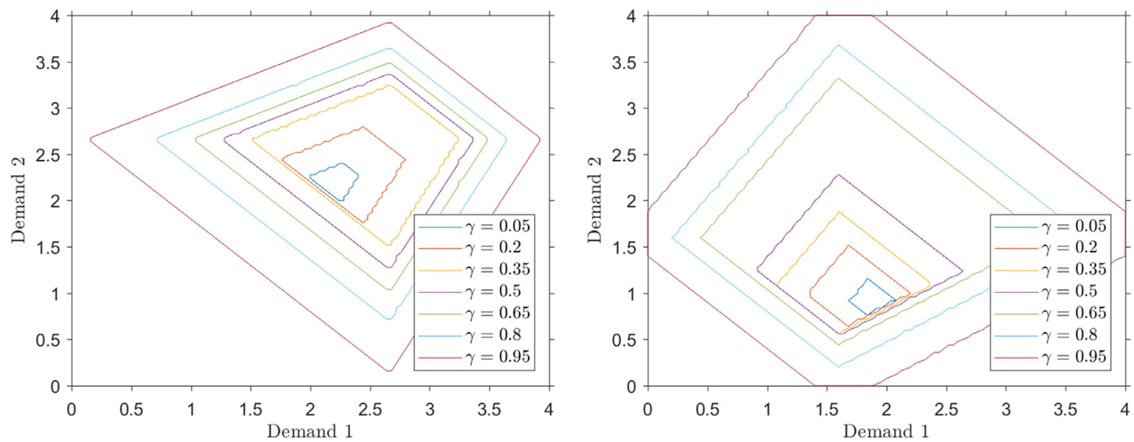
Let us now discuss how to use the maximal effective subsets to determine $\gamma$. Similarly to the single-product setting, the decision makers may infer their risk attitude by choosing a value of $\gamma$ for which the corresponding maximal effective subset is closest to their evaluation of the critical demand regions. Suppose the decision maker specifies the regions of demand that are *not* critical via a hyperrectangle $[a_1, b_1] \times [a_2, b_2]$. Then, we suggest choosing the smallest $\gamma$ whose corresponding maximal ineffective subset contains the region $[a_1, b_1] \times [a_2, b_2]$. If this $\gamma > \gamma^{\mathrm{cr}}$, then we suggest choosing $\gamma^{\mathrm{cr}}$. This approach is applicable to higher-dimensional spaces as well as the single-product setting.

Fig. 6 shows the price of optimism, price of pessimism, and regrets for the studied problems. Several properties can be seen from this figure. Observe that similarly to the single-product setting, the price of optimism and nominal regret are non-decreasing in $\gamma$, whereas the price of pessimism and worst-case regret are non-increasing in $\gamma$. Also, as before, even the most conservative decision maker has no reason to choose a level of robustness higher than $\gamma^{\mathrm{cr}}$. Price of optimism/pessimism and regrets can also help the decision maker choose the value of $\gamma$ based on the indifferent-to-solution and indifferent-to-distribution levels of robustness (recall these notions from Section 6). For problem 1, we

**Table 3**
Problem parameters.

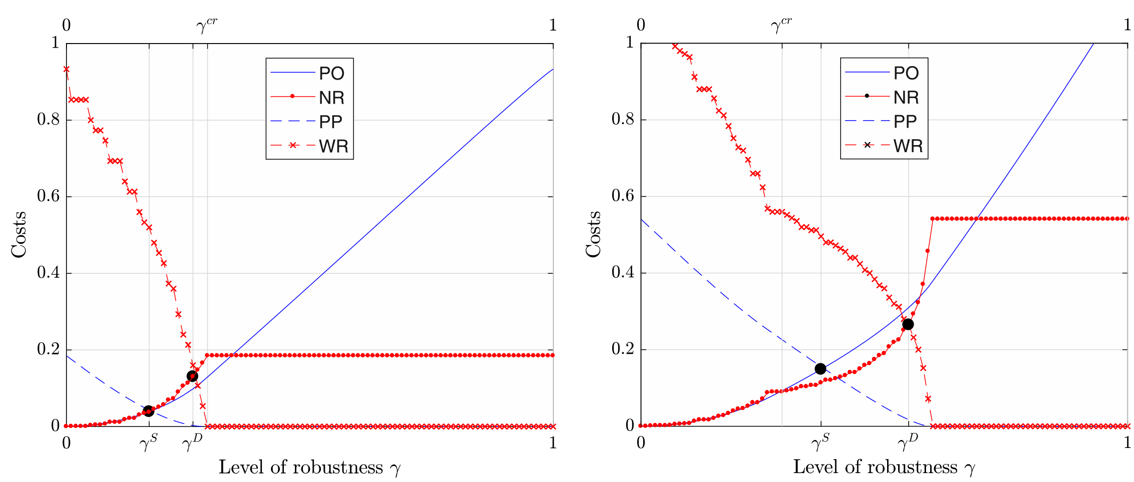
| Problem | $\boldsymbol{W}$ | $\boldsymbol{U}$ | $\boldsymbol{V}$ | $\boldsymbol{Q}$ | Condition | Distribution |
|---|---|---|---|---|---|---|
| 1 | $[0.5, 0.5]^\top$ | $[1, 1]^\top$ | $[0, 0]^\top$ | $[\frac{2}{3}, \frac{2}{3}]^\top$ | (C1) | $\xi_1 \sim \mathcal{N}(2, 0.25, 0, 4)$, $\xi_2 \sim \mathcal{N}(2, 0.25, 0, 4)$ |
| 2 | $[1.5, 1.5]^\top$ | $[1, 1]^\top$ | $[0, 0]^\top$ | $[0.4, 0.4]^\top$ | (C1) | $\xi_1 \sim \mathcal{N}(2, 0.25, 0, 4)$, $\xi_2 \sim \mathcal{N}(1, 0.25, 0, 4)$ |

$^\dagger\mathcal{N}(\mu, \sigma^2, a, b)$ denotes a normal distribution with mean $\mu$ and variance $\sigma^2$, truncated on $[a, b]$.




(a) Problem 1. (b) Problem 2.

**Fig. 5.** Maximal effective subsets.




(a) Problem 1. (b) Problem 2.

**Fig. 6.** Price of optimism, price of pessimism, nominal regret, and worst-case regret.

have we have $0.17 \simeq \gamma^S < \gamma^D \simeq 0.26$, whereas for problem 2, we have $0.37 \simeq \gamma^S < \gamma^D \simeq 0.55$. Note that although we did not present a theoretical result on the existence of $\gamma^D$ for the multi-product setting, an indifferent-to-distribution level of robustness exists for both problems in Table 3. We do not however make any claims about its uniqueness.

## 9. Conclusions and future research

Although the standard newsvendor model is used in a variety of settings, in many real-world applications the demand distribution is not known with certainty. In such cases, a distributionally robust version of the model—which minimizes the expected cost with respect to the worst-case demand distribution within an ambiguity set—is more realistic. There are of course multiple ways to choose the ambiguity set; in this paper we have studied one such way, formed by distributions which are within a certain variation-distance $\gamma$ from a reference distribution.

For the single-product setting, our results show that the optimal solution monotonically moves from $x^{\text{neut}}$ to $x^{\text{rob}}$ until a critical level of robustness, at which point the optimal solution stabilizes at $x^{\text{rob}}$. Practical use of this type of models requires choosing a value for $\gamma$, which can be interpreted as both the size of the ambiguity set as well as the level of risk-aversion of the model. We have introduced two tools to help the decision maker make that choice. One is based on the notion of maximal effective subsets for the problem, which correspond to regions of demand that are critical in the sense that their removal will change the optimal value. The other tool balances the price of optimism/pessimism and nominal/worst-case regrets resulting from using the distributionally robust model.

As we have seen in the OR reservation example, the concepts introduced in this paper can have a meaningful interpretation and ultimately help managers understand their risks and protect against the uncertainty in their demand distribution.

The main results in this paper can lay the foundation for studying other problem settings in the context of DRNV. For instance, we extended some of our results to the multi-product setting and showed that similar tools as in the single-product setting can help the decision maker to choose a level of robustness for their problem. Another direction for future research is studying multi-period DRNVs; see, e.g., Xin and Goldberg (2013, 2015). Another line of research would be to study distributionally robust versions of networks of risk-averse newsvendor models, a class of models introduced by Van Mieghem (2007) that yields useful guidelines for risk-pooling and safety capacity in inventory networks. Other future work includes studying other cost functionals and ambiguity sets, including regret-based cost functionals and moment-based ambiguity sets (Hanasusanto et al., 2015), possibly in conjunction with more general $\phi$-divergence based ambiguity sets (Bayraksan & Love, 2015) and Wasserstein-metric-based ambiguity sets (Blanchet & Murthy, 2017; Gao & Kleywegt, 2016; Mohajerin Esfahani & Kuhn, 2018). In all these settings, characterizing the maximal effective subsets and the price of optimism/pessimism and nominal/worst-case regrets could provide the decision makers with useful insights about the underlying uncertainty.

## Acknowledgments

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 grant CMMI-1563504 and the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR) under Contract DE-AC02-06CH11347. The third author acknowledges the support of grant FONDECYT 1171145, Chile.

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2019.06.036.

## References

Ahmed, S., Çakmak, U., & Shapiro, A. (2007). Coherent risk measures in inventory problems. *European Journal of Operational Research, 182*, 226–238.
Analui, B., & Pflug, G. C. (2014). On distributionally robust multiperiod stochastic optimization. *Computational Management Science, 11*, 197–220.
Ardestani-Jaafari, A., & Delage, E. (2016). Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research, 64*, 474–494.
Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (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*, 263–272.
Bayraksan, G., & Love, D. K. (2015). Data-driven stochastic programming using phi-divergences. In *Tutorials in operations research* (pp. 1–19). INFORMS.
Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., & Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science, 59*, 341–357.
Bertsimas, D., Gupta, V., & Kallus, N. (2018). Robust sample average approximation. *Mathematical Programming, 171*, 217–282.
Blanchet, J., & Murthy, K. R. (2019). Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*. doi:10.1287/moor.2018.0936.
Chen, W., & Sim, M. (2009). Goal-driven optimization. *Operations Research, 57*, 342–357.
Chen, Y. F., Xu, M., & Zhang, Z. G. (2009). Technical note- a risk-averse newsvendor model under the CVaR criterion. *Operations Research, 57*, 1040–1044.
Chick, S. E., Mamani, H., & Simchi-Levi, D. (2008). Supply chain coordination and influenza vaccination. *Operations Research, 56*, 1493–1506.
Choi, S., & Ruszczyński, A. (2008). A risk-averse newsvendor with law invariant coherent measures of risk. *Operations Research Letters, 36*, 77–82.
Choi, S., Ruszczyński, A., & Zhao, Y. (2011). A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. *Operations Research, 59*, 346–364.
Dexter, F., Macario, A., Epstein, R. H., & Ledolter, J. (2005). Validity and usefulness of a method to monitor surgical services' average bias in scheduled case durations. *Canadian Journal of Anesthesia, 52*, 935–939.
Dexter, F., Traub, R. D., Fleisher, L. A., & Rock, P. (2002). What sample sizes are required for pooling surgical case durations among facilities to decrease the incidence of procedures with little historical data? *Anesthesiology, 96*, 1230–1236.
Fügener, A., Schiffels, S., & Kolisch, R. (2017). Overutilization and underutilization of operating rooms - insights from behavioral health care operations management. *Health Care Management Science, 20*, 115–128.
Gallego, G., & Moon, I. (1993). The distribution free newsboy problem: Review and extensions. *Journal of the Operational Research Society, 44*, 825–834.
Gao, R., & Kleywegt, A. J. (2016). *Distributionally robust stochastic optimization with Wasserstein distance* arXiv preprint arXiv:1604.02199v2 [math.OC].
Gotoh, J. y., Kim, M. J., & Lim, A. E. B. (2017). *Calibration of distributionally robust empirical optimization models* arXiv preprint arXiv:1711.06565 [stat.ML].
Gotoh, J. y., & Takano, Y. (2007). Newsvendor solutions via conditional value-at-risk minimization. *European Journal of Operational Research, 179*, 80–96.
Gupta, D. (2007). Surgical suites' operations management. *Production and Operations Management, 16*, 689–700.
Han, Q., Du, D., & Zuluaga, L. F. (2014). Technical note- A risk-and ambiguity-averse extension of the max-min newsvendor order formula. *Operations Research, 62*, 535–542.
Hanasusanto, G. A., Kuhn, D., Wallace, S. W., & Zymler, S. (2015). Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming, 152*, 1–32.
Harrison, J. M., & Zeevi, A. (2005). A method for staffing large call centers based on stochastic fluid models. *Manufacturing & Service Operations Management, 7*, 20–36.
Jiang, R., & Guan, Y. (2018). Risk-averse two-stage stochastic program with distributional ambiguity. *Operations Research, 66*, 1390–1405.
Lehtonen, J. M., Torkki, P., Peltokorpi, A., & Moilanen, T. (2013). Increasing operating room productivity by duration categories and a newsvendor model. *International Journal of Health Care Quality Assurance, 26*, 80–92.
Littlewood, K. (1972). Forecasting and control of passenger bookings. In *Proceedings of the AGIFORS 12th annual symposium proceedings*. Nathanya, Israel
Macario, A. (2009). Truth in scheduling: Is it possible to accurately predict how long a surgical case will last? *Anesthesia & Analgesia, 108*, 681–685.
Macario, A. (2010). *Is it possible to predict how long a surgery will last?*. Medscape. Accessed: April 25, 2018, https://www.medscape.com/viewarticle/724756.
May, J. H., Spangler, W. E., Strum, D. P., & Vargas, L. G. (2011). The surgical scheduling problem: Current research and future opportunities. *Production and Operations Management, 20*, 392–405.
Mohajerin Esfahani, P., & Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming, 171*, 115–166.
Mostard, J., de Koster, R., & Teunter, R. (2005). The distribution-free newsboy problem with resalable returns. *International Journal of Production Economics, 97*, 329–342.
Natarajan, K., Sim, M., & Uichanco, J. (2018). Asymmetry and ambiguity in newsvendor models. *Management Science, 64*, 3146–3167.
Olivares, M., Terwiesch, C., & Cassorla, L. (2008). Structural estimation of the newsvendor model: An application to reserving operating room time. *Management Science, 54*, 41–55.
Perakis, G., & Roels, G. (2008). Regret in the newsvendor model with partial information. *Operations Research, 56*, 188–203.
Phu, H., & Hoffmann, A. (1996). Essential supremum and supremum of summable functions: Summable functions. *Numerical Functional Analysis and Optimization, 17*, 161–180.
Qin, Y., Wang, R., Vakharia, A. J., Chen, Y., & Seref, M. M. (2011). The newsvendor problem: Review and directions for future research. *European Journal of Operational Research, 213*, 361–374.
Rahimian, H., Bayraksan, G., & Homem-de Mello, T. (2019). Identifying effective scenarios in distributionally robust stochastic programs with total variation distance. *Mathematical Programming, 173*, 393–430.
Rockafellar, R. T., & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance, 26*, 1443–1471.
Samudra, M., Van Riet, C., Demeulemeester, E., Cardoen, B., Vansteenkiste, N., & Rademakers, F. E. (2016). Scheduling operating rooms: achievements, challenges and pitfalls. *Journal of Scheduling, 19*, 493–525.
Scarf, H. (1958). A min-max solution of an inventory problem. In H. Scarf, K. Arrow, & S. Karlin (Eds.), *Studies in the mathematical theory of inventory and production: 10* (pp. 201–209). Stanford University Press Stanford, CA.
Shapiro, A., Tekaya, W., da Costa, J. P., & Soares, M. P. (2013). Risk neutral and risk averse stochastic dual dynamic programming method. *European Journal of Operational Research, 224*, 375–391.
Strum, D. P., May, J. H., & Vargas, L. G. (2000a). Modeling the uncertainty of surgical procedure times: Comparison of log-normal and normal models. *Anesthesiology, 92*, 1160–1167.
Strum, D. P., Sampson, A. R., May, J. H., & Vargas, L. G. (2000b). Surgeon and type of anesthesia predict variability in surgical procedure times. *Anesthesiology, 92*, 1454–1466.
Strum, D. P., Vargas, L. G., & May, J. H. (1999). Surgical subspecialty block utilization and capacity planning a minimal cost analysis model. *Anesthesiology, 90*, 1176–1185.
Tomlin, B., & Wang, Y. (2005). On the value of mix flexibility and dual sourcing in unreliable newsvendor networks. *Manufacturing & Service Operations Management, 7*, 37–57.

Van Mieghem, J. A. (2007). Risk mitigation in newsvendor networks: Resource diversification, flexibility, sharing, and hedging. *Management Science, 53*, 1269–1288.

Wachtel, R. E., & Dexter, F. (2010). Review of behavioral operations experimental studies of newsvendor problems for operating room management. *Anesthesia & Analgesia, 110*, 1698–1710.

Wang, Z., Glynn, P. W., & Ye, Y. (2016). Likelihood robust optimization for data–driven problems. *Computational Management Science, 13*, 241–261.

Wu, M., Zhu, S. X., & Teunter, R. H. (2013a). Newsvendor problem with random shortage cost under a risk criterion. *International Journal of Production Economics, 145*, 790–798.

Wu, M., Zhu, S. X., & Teunter, R. H. (2013b). The risk-averse newsvendor problem with random capacity. *European Journal of Operational Research, 231*, 328–336.

Wu, M., Zhu, S. X., & Teunter, R. H. (2014). A risk-averse competitive newsvendor problem under the CVaR criterion. *International Journal of Production Economics, 156*, 13–23.

Xin, L., & Goldberg, D. A. (2013). *Time (in) consistency of multistage distributionally robust inventory models with moment constraints* arXiv preprint arXiv:1304.3074 [math.OC].

Xin, L., & Goldberg, D. A. (2015). *Distributionally robust inventory control when demand is a martingale* arXiv preprint arXiv:1511.09437 [math.PR].

Yu, H., Zhai, J., & Chen, G. Y. (2016). Robust optimization for the loss-averse newsvendor problem. *Journal of Optimization Theory and Applications, 171*, 1008–1032.

Zhang, W., Rahimian, H., & Bayraksan, G. (2016). Decomposition algorithms for risk-averse multistage stochastic programs with application to water allocation under uncertainty. *INFORMS Journal on Computing, 28*, 385–404.

Zhao, C., & Guan, Y. (2015). *Data-driven risk-averse two-stage stochastic program with $\zeta$-structure probability metrics*. Technical report, Available on Optimization Online http://www.optimization-online.org/DB_HTML/2015/07/5014.html/.

Zhou, J., Dexter, F., Macario, A., & Lubarsky, D. A. (1999). Relying solely on historical surgical times to estimate accurately future surgical times is unlikely to reduce the average length of time cases finish late. *Journal of Clinical Anesthesia, 11*, 601–605.

Please cite this article as: H. Rahimian, G. Bayraksan and T. Homem-de-Mello, Controlling risk and demand ambiguity in newsvendor models, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2019.06.036