On the heavy-tail behavior of the distributionally robust newsvendor

Bikramjit Das* Anulekha Dhara† Karthik Natarajan‡§

Submitted: June 2018, First Revision: August 2019, Second Revision: February 2020, Third Revision: June 2020, Accepted: September 2020

Abstract

Since the seminal work of Scarf (1958) [A min-max solution of an inventory problem, Studies in the Mathematical Theory of Inventory and Production, pages 201-209] on the newsvendor problem with ambiguity in the demand distribution, there has been a growing interest in the study of the distributionally robust newsvendor problem. The model is criticized at times for being conservative since the worst-case distribution is discrete with a few support points. However, it is the order quantity prescribed by the model that is of practical relevance. Interestingly, the order quantity from Scarf's model is optimal for a heavy-tailed distribution. In this paper, we generalize this observation by showing a "heavy-tail optimality" property of the distributionally robust order quantity for an ambiguity set where information on the first and the αth moment is known, for any real $\alpha > 1$. We show that the optimal order quantity for the distributionally robust newsvendor is also optimal for a regularly varying distribution with parameter α. In the high service level regime, when the original demand distribution is given by an exponential or a lognormal distribution and contaminated with a regularly varying distribution, the distributionally robust order quantity is shown to outperform the optimal order quantity of the original distribution, even with a small amount of contamination.

1 Introduction

Since the pioneering work of Scarf [34], there has been a growing interest in the study of the distributionally robust newsvendor problem where the probability distribution of the demand is ambiguous. Formally, the problem is stated as follows. A newsvendor needs to decide on the number of units of an item to order before the actual demand is observed. The unit purchase cost is c and the unit revenue

*Engineering Systems and Design, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372. Email: bikram@sutd.edu.sg

†TCS Research, New Delhi, India. Email: anulekha.dhara@tcs.com

‡Engineering Systems and Design, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372. Email: karthik_natarajan@sutd.edu.sg

§The research of the first author was partly supported by the MOE Academic Research Fund Tier 2 grant MOE2017-T2-2-161, "Learning from common connections in social networks" second and third authors was partly supported by MOE Academic Research Fund Tier 2 grant T2MOE1706, "On the Interplay of Choice, Robustness and Optimization in Transportation" and the SUTD-MIT International Design Center grant IDG21700101 on "Design of the Last Mile Transportation System: What Does the Customer Really Want?".

is p where $p > c > 0$. Any unsold item at the end of the selling period has zero salvage value. The demand $\tilde{d}$ for the item is random and unknown before the order is placed. The probability distribution of the demand, denoted by $F(w) := \mathbb{P}(\tilde{d} \leq w)$ is ambiguous and only assumed to lie in a set of possible distributions $\mathcal{F}$. Among other reasons, this ambiguity may arise due to the introduction of a new product with no past demand data, or, when past data is used in terms of moments, probability distance metrics or other structural information to create a plausible set of demand distributions. The distributionally robust newsvendor orders the quantity that maximizes the minimum expected profit. Mathematically, this problem is formulated as choosing an order quantity q that solves:

$$\max_{q \geq 0} \quad \inf_{F \in \mathcal{F}} \quad \left(p\mathbb{E}_F[\min(q, \tilde{d})] - cq \right). \tag{1.1}$$

Using the relation $\min(d, q) = d - [d - q]_+$, where $[d - q]_+ = \max(0, d - q)$, the optimal order quantity in (1.1) is equivalently obtained by solving the problem:

$$\min_{q \geq 0} \quad \left((1 - \eta)q + \sup_{F \in \mathcal{F}} \mathbb{E}_F[\tilde{d} - q]_+ \right), \tag{1.2}$$

under the assumption that the mean value of demand is specified in $\mathcal{F}$, where $\eta = 1 - c/p \in [0, 1)$ denotes the critical ratio.

1.1 Scarf's Model

The earliest version of the model is attributed to Scarf [34] who assumed that the mean and the variance of the demand are specified in the set $\mathcal{F}$. The set of demand distributions is defined as:

$$\mathcal{F}_{1,2} = \left\{ F \in \mathbb{M}(\Re_+) : \int_0^\infty \mathrm{d}F(w) = 1, \ \int_0^\infty w \, \mathrm{d}F(w) = m_1, \ \int_0^\infty w^2 \, \mathrm{d}F(w) = m_2 \right\},$$

where $\mathbb{M}(\Re_+)$ is the set of finite positive Borel measures supported on the non-negative real line and m_1 and m_2 are the first and second moment which are assumed to satisfy $m_2 > m_1^2 > 0$. In the standard newsvendor model, where the set of distributions is a singleton, the optimal order quantity reduces to the well-known critical fractile formula $q^* = F^{-1}(\eta)$, where $F^{-1}(\cdot)$ is the generalized inverse of the cumulative distribution function. However, in the robust model, the worst-case demand distribution depends on the order quantity. Given an order quantity $q \geq m_2/2m_1$, Scarf [34] characterized the worst-case two point demand distribution in $\sup_{F \in \mathcal{F}_{1,2}} \mathbb{E}_F[\tilde{d} - q]_+$ as:

$$\tilde{d}_q = \begin{cases} q - \sqrt{q^2 - 2m_1 q + m_2}, & \text{w.p. } \dfrac{1}{2}\left(1 + \dfrac{q - m_1}{\sqrt{q^2 - 2m_1 q + m_2}}\right), \\ q + \sqrt{q^2 - 2m_1 q + m_2}, & \text{w.p. } \dfrac{1}{2}\left(1 - \dfrac{q - m_1}{\sqrt{q^2 - 2m_1 q + m_2}}\right), \end{cases}$$

where the support points and the probabilities are dependent on q (this can be viewed as the "power" of the adversary). In this case the worst-case demand distribution will have one support point strictly above the order quantity, even when the order quantity approaches $+\infty$. When the order quantity q lies in the range $[0, m_2/2m_1]$, the worst-case demand distribution is two point, but independent of q and given by $\tilde{d}_{m_2/2m_1}$, where:

$$\tilde{d}_{m_2/2m_1} = \begin{cases} 0, & \text{w.p. } 1 - \dfrac{m_1^2}{m_2}, \\ \dfrac{m_2}{m_1}, & \text{w.p. } \dfrac{m_1^2}{m_2}. \end{cases}$$

Combining these results, the worst-case bound is given as:

$$\sup_{F \in \mathcal{F}_{1,2}} \mathbb{E}_F[\tilde{d} - q]_+ = \begin{cases} \dfrac{1}{2}\left(\sqrt{q^2 - 2m_1 q + m_2} + m_1 - q\right), & \text{if } q \geq \dfrac{m_2}{2m_1}, \\ m_1 - \dfrac{q m_1^2}{m_2}, & \text{if } 0 \leq q < \dfrac{m_2}{2m_1}. \end{cases} \tag{1.3}$$

Plugging in the expression (1.3) into (1.2) and from calculus, the closed form solution for the optimal order quantity is given by:

$$q_\eta^{\text{scarf}} = \begin{cases} m_1 + \dfrac{\sqrt{m_2 - m_1^2}}{2} \dfrac{2\eta - 1}{\sqrt{\eta(1-\eta)}}, & \text{if } \dfrac{m_2 - m_1^2}{m_2} < \eta < 1, \\ 0, & \text{if } 0 \leq \eta < \dfrac{m_2 - m_1^2}{m_2}. \end{cases}$$

For $\eta = (m_2 - m_1^2)/m_2$, the optimal order quantity is given by the set of all values in the interval $[0, m_2/(2m_1)]$. The closed form solution leads to some interesting observations:

(i) The optimal order quantity is above, equal to, or below the mean depending on whether η is greater than, equal to or less than $1/2$ respectively.

(ii) As the critical ratio $\eta \to 1$, the order quantity $q_\eta^{\text{scarf}} \to +\infty$. Moreover, the limiting behavior of the optimal order quantity satisfies $\lim_{\eta \uparrow 1} \left(q_\eta^{\text{scarf}} \sqrt{1-\eta}\right) = \sqrt{m_2 - m_1^2}/2$.

While the optimal order quantity is a simple closed form expression, this model is criticized at times for being conservative[1]. In fact the accuracy of the model depends on the actual distribution, if known, and the critical ratio η. Scarf [34] observed that for a large range of critical ratios (specifically $\eta \in [0.05, 0.95]$), the optimal order quantity for the two moment model is very close to the optimal order quantity for a normal approximation of a Poisson distribution, while for higher critical ratios, the model prescribed higher order quantities (see Gallego and Moon [21] for related results). In contrast, Wang, Glynn and Ye [40] found that the difference between the order quantity from Scarf's model and the optimal order quantity under the exponential distribution is more significant, for certain choices of the critical ratio (see Figure 1(c) for $\eta = 1/2$). In Figure 1, we provide a comparison of the optimal order quantities for two demand distributions (normal and exponential) and Scarf's model where only the first

[1] We cite from page 243 in Wang, Glynn and Ye [40]: "In the distributionally robust optimization approach, the worst-case distribution for a decision is often unrealistic. Scarf (1958) shows that the worst-case distribution in the newsvendor context is a two point distribution. This raises the concern that the decision chosen by this approach is guarding under some overly conservative scenarios, while performing poorly in more likely scenarios. Unfortunately, these drawbacks seem to be inherent in the model choice and cannot be remedied easily."

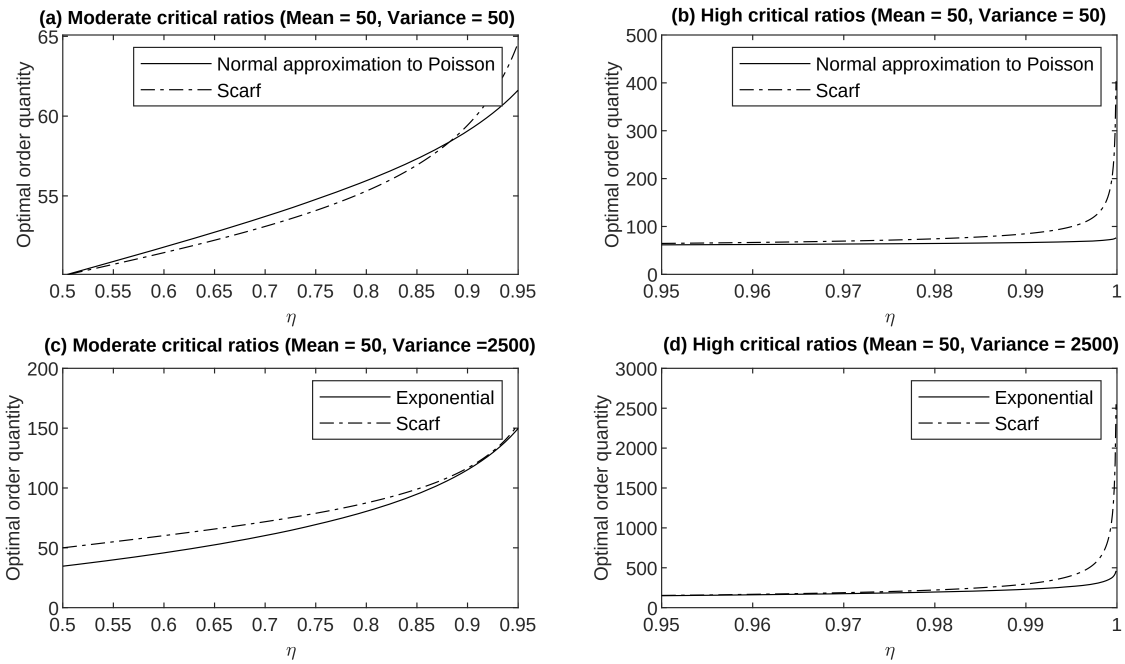

Figure 1: The plots at the top compare the optimal order quantities for a normal approximation to a Poisson demand distribution and Scarf's model as in the original paper of Scarf [34] with mean demand 50 and variance 50. The plots at the bottom compare the optimal order quantities for an exponential demand distribution and Scarf's model with mean demand 50 and variance 2500.

two moments are assumed to be known. While the figure suggests the optimal order quantities from Scarf's model is comparatively close to the optimal order quantities for the normal and exponential distributions for moderate critical ratios, it prescribes substantially higher order quantities for high critical ratios. In this paper, we provide an analytical characterization of this numerical insight and generalize it to a larger class of distributionally robust newsvendor models.

1.2 Motivation and Contributions

Our interest in the analysis of the tail behavior is motivated by its relevance in capacity management problems with high service level guarantees. There is growing evidence in the literature that stockouts for retailer have significant short-term and long-term effects that needs to be minimized (see Anderson, Fitzsimons and Simester [1]). In cloud computing, the service level agreements of major companies such as Google and Amazon guarantee service levels at 99.9% and even higher for many of their products[2]. The high service levels have direct implications on the capacity planning of resources. Another example is the outbreak of the COVID-19 pandemic of 2019-20 which led to a severe shortage of products such as hand sanitizers and thermometers at both offline and online retailer locations in China and Singapore[3]. While the demand for such products are typically moderate, planning inventory in the face of a big

[2]See https://cloud.google.com/terms/sla and https://aws.amazon.com/legal/service-level-agreements/

[3]See https://www.dailymail.co.uk/health/article-7926111/Chinese-e-commerce-sites-running-medical-masks.html and https://www.straitstimes.com/singapore/coronavirus-no-shortage-of-food-here-say-leaders-in-appeal-for-calm-amid-panic-buying/

spike in demand is a serious challenge. At the same time, precisely modeling the tail of such demand distributions is challenging. This brings us to the format of the paper and the main contributions:

(a) In Section 2, we review the distributionally robust newsvendor problem and show that Scarf's order quantity is optimal for a heavy-tailed censored student-t distribution. While this observation has been made in prior research (see Müller and Stoyan [30] and Gallego [20]), the result has not been extended to generalizations of Scarf's model, to the best of our knowledge.

(b) In Section 3, we propose a generalization of the ambiguity set from the first and the second moment to the first and the αth moment for any real $\alpha > 1$. The ambiguity set is simple while providing flexibility in allowing for new distributions to be considered. However unlike Scarf's model, the solution does not appear to have a closed form representation. We show that for low critical values it is optimal to order nothing. For high critical values (resulting in high order quantities) we characterize the worst-case expected cost with asymptotically tight upper and lower bounds. The proof technique should be of independent interest in other contexts where closed form solutions for moment problems are not readily available.

(c) In Section 4, we provide a characterization of the optimal order quantity in the high service level regime for the distributionally robust newsvendor model by showing that it is optimal for a regularly varying (heavy-tailed) distribution with tail parameter α. This provides an explicit link between the solution of a robust optimization problem which accounts for worst-case behavior and heavy-tails which are used to model extreme events.

(d) In Section 5, we use numerical examples to illustrate the performance of the model for high critical ratios. We show: (1) the optimal order quantities for certain distributions are better approximated by incorporating moment information beyond Scarf's model; (2) the behavior of the ratio of the optimal order quantities to the optimal costs as the critical ratio increases; (3) the robustness of the models under contamination by heavy-tailed distributions. We conclude in Section 6 by identifying future research directions.

2 Literature Review

In this section, we review some of the key results for the distributionally robust newsvendor problem with a focus on ambiguity sets where demand might take any value in $[0, \infty)$. We also review prior research that provides empirical evidence on heavy-tailed demand distributions.

2.1 Distributionally Robust Newsvendor Models

Shapiro and Kleywegt [37] and Shapiro and Ahmed [36] reformulated the distributionally robust newsvendor as a classical newsvendor problem through the construction of a new probability demand distribution. Their key insight (see Section 3.1 on page 532 in Shapiro and Kleywegt [37]) was that given a set $\mathcal{F}$ of non-negative random variables with a finite mean, there exists a non-negative random variable $\tilde{d}^*$

with probability distribution F^* such that the following equality holds:

$$\sup_{F\in\mathcal{F}} \mathbb{E}_F[\tilde{d}-q]_+ = \mathbb{E}_{F^*}[\tilde{d}^*-q]_+, \ \forall q. \tag{2.1}$$

This is seen by noting that the function $\Pi(q) := \sup_{F\in\mathcal{F}} \mathbb{E}_F[\tilde{d}-q]_+$ is a non-increasing, convex function that satisfies the following properties: (i) For all $q \leq 0$, $\Pi(q)+q = \sup_{F\in\mathcal{F}} \mathbb{E}_F[\tilde{d}]$ which is a constant under the assumption that the ambiguity set contains distributions of non-negative random variables with a finite mean (equal to m_1 when the mean is specified in the ambiguity set), and (ii) $\lim_{q\to\infty} \Pi(q) = 0$. This implies that there exists a non-negative random variable $\tilde{d}^*$ with a distribution given by $F^*(q) = 1 + \Pi'_+(q)$ where $\Pi'_+(\cdot)$ is the right derivative of $\Pi(\cdot)$, such that condition (2.1) is satisfied. It is easy to verify that when the ambiguity set $\mathcal{F}$ consists of random variables with a fixed mean m_1, the new random variable $\tilde{d}^*$ also has mean m_1, since $\Pi(0) = m_1$. The random variable $\tilde{d}^*$ dominates all the random variables $\tilde{d}$ in the set $\mathcal{F}$ in an increasing convex order sense (see Müller and Stoyan [30], Shaked and Shanthikumar [35]). Moreover, unlike the extremal distribution on the left hand side of the equation (2.1) which might vary with q, the random variable $\tilde{d}^*$ with distribution F^* does not depend on q. This equivalence helps convert the distributionally robust newsvendor problem to the classical newsvendor problem as follows:

$$\min_{q\geq 0} \ \left((1-\eta)q + \mathbb{E}_{F^*}[\tilde{d}^*-q]_+\right), \tag{2.2}$$

where F^* does not depend on q and η. However, the challenge here lies in finding F^*, which in many cases does not have an explicit expression and may not even lie within the set $\mathcal{F}$. Nevertheless, the equivalence provides an important insight by identifying a distribution F^* for which the optimal order quantity from the distributionally robust newsvendor model in (1.2) is optimal for any the critical ratio.

In Scarf's model, it is straightforward to construct the distribution F^* as the worst-case bound is known in closed form for each value of q. Taking the right-derivatives in (1.3) we can compute

$$F^*(w) = \mathbb{P}(\tilde{d}^* \leq w) = \begin{cases} \dfrac{1}{2}\left(1 + \dfrac{w-m_1}{\sqrt{w^2-2m_1w+m_2}}\right), & \text{if } w \geq \dfrac{m_2}{2m_1}, \\ 1 - \dfrac{m_1^2}{m_2}, & \text{if } 0 \leq w < \dfrac{m_2}{2m_1}. \end{cases} \tag{2.3}$$

The distribution in (2.3) defines a censored student-t random variable with a mixture of discrete and continuous terms as follows:

$$\tilde{d}^* = \begin{cases} \tilde{t}_2\left(m_1, (m_2-m_1^2)/2\right), & \text{if } \tilde{t}_2\left(m_1, (m_2-m_1^2)/2\right) \geq \dfrac{m_2}{2m_1}, \\ 0, & \text{otherwise}, \end{cases} \tag{2.4}$$

where $\tilde{t}_\nu(\mu, \sigma^2)$ is a three parameter student-t random variable with location μ, scale $\sigma > 0$, degrees of freedom $\nu > 0$, and probability density function given by:

$$g(w) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\sigma\Gamma(\frac{\nu}{2})}\left(1+\frac{1}{\nu}\left(\frac{w-\mu}{\sigma}\right)^2\right)^{-(\nu+1)/2}, \quad \forall w \in \Re. \tag{2.5}$$

It is easy to check here that $\mathbb{E}_{F^*}[\tilde{d}^{*2}] = \infty$. Hence to recreate the optimal order quantity of Scarf's model which contains distributions with finite second moments, we need to solve a standard newsvendor problem for a random variable $\tilde{d}^*$ as described in (2.4), which has infinite second moment. To the best of our knowledge, this observation has been made for real-valued random variables with known first and second moments in Theorem 1.10.7, Müller and Stoyan [30], with its application to inventory problems discussed in Gallego [20].

Several generalizations of the distributionally robust newsvendor model to other ambiguity sets have been studied. While for some of these ambiguity sets (see Ben-Tal and Hochman [5], Natarajan, Uichanco and Sim [31], Bertsimas, Gupta and Kallus [2], Chen, He and Zhang [12]), the problem is solvable in a closed form-manner, in most cases, numerical optimization techniques are needed. Bertsimas and Popescu [3] and Lasserre [27] developed semidefinite optimization techniques to compute the worst-case bound when the set of distributions is defined by a set of fixed moments up to an integer degree $n \in \mathbb{Z}_+$:

$$\mathcal{F}_{1,2,\ldots,n} = \left\{F \in \mathbb{M}(\Re_+) : \int_0^\infty \mathrm{d}F(w) = 1, \int_0^\infty w^i \mathrm{d}F(w) = m_i,\ i = 1,2,\ldots,n\right\}.$$

While some attempts has been made to solve this problem analytically for $n = 3$ and $n = 4$, the tight worst-case bounds have complicated expressions involving roots of cubic and quartic equations (see Jansen, Haezendonck, and Goovaerts [25], Zuluaga, Pena and Du [41]). In general, for these problems, there is an absence of closed form solutions and hence finding an explicit representation of the distribution F^* does not appear to be straightforward.

Lam and Mottet [26] have recently characterized worst-case distributions for ambiguity sets containing random variables whose tail probability beyond a given threshold is known and are assumed to have convex tail density functions among other mild assumptions. In the context of the newsvendor problem, their result implies that the worst-case density function is in fact a piecewise linear function beyond the threshold, with at most two linear pieces and is a extremely light-tailed distribution. Ben-Tal et al. [4] studied newsvendor problems with ϕ-divergence based ambiguity sets around a reference discrete distribution and proposed a convex optimization formulation to solve the distributionally robust optimization problem. Building on this, Blanchet, He and Murthy [8] found that for certain reference distributions using the Kullback-Leibler distance leads to worst-case distributions with very heavy-tails; in contrast using the Renyi divergence provides less conservative solutions. A related recent stream of literature has focused on solving distributionally robust optimization problems, including the newsvendor model, using ambiguity sets defined around an empirical distribution with the Wasserstein distance (see Esfahani and Kuhn [17], Gao and Kleywegt [22] and Blanchet and Murthy [9]) with convex optimization methods; but most of these works address light-tailed distributions and do no extend easily to the heavy-tailed case.

2.2 Empirical Evidence of Heavy-Tailed Demand

There has been growing evidence in the recent years that heavy-tailed demand distributions occur in practice and has to be better accounted for in operational settings. Clauset, Shalizi and Newman [10] found strong evidence for power-law behavior in two datasets relevant to demand models: a) the number of calls received by customers of AT&T long distance telephone in the United States during a single day and b) the number of copies of bestselling books sold in the United States during the period 1895 to 1965. In another study, Gaffeo, Scorcu and Vici [19] analyzed the demand of books in Italy and found that for the three categories - local novels, foreign novels and non-fiction books, a power-law distribution where the exponent is typically less than 2 is a good fit to the right tail of the demand distribution. Bimpkis and Markakis [6] used the ratings of movies on Netflix as an approximation to the demand of a movie and estimated a power law distribution with an exponent of around 1.04 for the number of movies per number of distinct ratings. Natarajan, Sim and Uichanco [31] used data from a European automotive manufacturer with 36 spare part SKUs over a one year period and found that the best-fit among 17 different families was often obtained by heavy-tailed distributions such as Pareto, extreme value or t-distributions. Chevalier and Goolsbee [13] used publicly available data on sales ranks of books on Amazon.com to obtain estimates on the sales quantity of the books, and observed that a Pareto distribution with a parameter of 1.2 was a reasonable approximation to the demand data. Empirical evidence in this literature seems to suggest that when Pareto distributions are used to model the demand, the exponent is strictly greater than 1 and possesses finite mean but might not necessarily possess finite variance. The ambiguity set we consider in the next section is inspired by such empirical evidence.

3 Model with the First and αth Moment

Consider an ambiguity set defined as follows for a fixed $\alpha > 1$:

$$\mathcal{F}_{1,\alpha} = \left\{ F \in \mathbb{M}(\Re_+) : \int_0^\infty \mathrm{d}F(w) = 1,\ \int_0^\infty w\,\mathrm{d}F(w) = m_1,\ \int_0^\infty w^\alpha\,\mathrm{d}F(w) = m_\alpha \right\}, \tag{3.1}$$

where m_1 and m_α are fixed numbers satisfying $m_\alpha > m_1^\alpha > 0$. We discuss a few features of this ambiguity set next:

(a) The ambiguity set $\mathcal{F}_{1,\alpha}$ can be used with any real value $\alpha > 1$, not necessarily just an integer. Clearly, when $\alpha = 2$, this corresponds to the original model of Scarf [34]. This allows for the possibility of the ambiguity set to specify more light-tailed ($\alpha > 2$) or more heavy-tailed distributional information ($\alpha < 2$) than Scarf's model allows.

(b) The ambiguity set preserves the simplicity of Scarf's [34] moment ambiguity set as it is parameterized by the choice of only three parameters - α, m_1 and m_α. The choice of α can be estimated from sample data using nonparametric hypothesis tests (see Fedotenkov [18]).

Under this ambiguity set, we are interested in solving:

$$\min_{q\geq 0}\left((1-\eta)q+\sup_{F\in\mathcal{F}_{1,\alpha}}\mathbb{E}_F[\tilde{d}-q]_+\right). \tag{3.2}$$

The primal moment formulation for computing $\sup_{F\in\mathcal{F}_{1,\alpha}}\mathbb{E}_F[\tilde{d}-q]_+$ is given as:

$$\sup\left\{\int_0^\infty [w-q]_+dF(w)\ \Big|\ \int_0^\infty \mathrm{d}F(w)=1,\int_0^\infty w\mathrm{d}F(w)=m_1,\int_0^\infty w^\alpha\mathrm{d}F(w)=m_\alpha, F\in\mathbb{M}(\Re_+)\right\}, \tag{3.3}$$

and the corresponding dual formulation given as:

$$\begin{aligned} \inf \quad & y_0+y_1m_1+y_\alpha m_\alpha \\ \text{s.t.} \quad & y_0+y_1w+y_\alpha w^\alpha\geq 0, && \forall w\geq 0, \\ & y_0+y_1w+y_\alpha w^\alpha\geq w-q, && \forall w\geq 0, \end{aligned} \tag{3.4}$$

where y_0 is the dual variable for the constraint that the total probability is equal to 1 and y_1 and y_α are the dual variables for the first and the αth moment constraints respectively. In the next section, we consider a special case where for a range of small values of q, the primal and dual formulations are solvable in closed form. This helps us characterize the optimal order quantity in the low service level regime by showing that there exists a threshold below which for all critical ratios, it is optimal to order nothing. The flexibility of allowing for any value $\alpha>1$ leads to challenges in solving the inner moment problem in closed form for higher values of q. In the Appendix, we discuss why finding a simple closed form solution to this problem is very unlikely. Instead, we focus on finding lower and upper bounds on the worst-case expected value that is valid beyond a certain value of the order quantity q. Our approach is based on constructing approximately optimal primal-dual solutions in this regime.

3.1 Closed Form for Small Values of q

For small values of q, the moment problem (3.3) can be solved in closed form. Building on this, we provide a characterization of the optimal solution to the distributionally robust newsvendor problem (3.2) for small values of η.

Proposition 3.1. *Given an ambiguity set $\mathcal{F}_{1,\alpha}$, the worst-case expected value is given as:*

$$\sup_{F\in\mathcal{F}_{1,\alpha}}\mathbb{E}_F[\tilde{d}-q]_+=m_1-q\left(\frac{m_1^\alpha}{m_\alpha}\right)^{1/(\alpha-1)},\quad \text{if } 0\leq q\leq\left(\frac{\alpha-1}{\alpha}\right)\left(\frac{m_\alpha}{m_1}\right)^{1/(\alpha-1)}. \tag{3.5}$$

The worst-case demand distribution in this case is independent of q and is given as:

$$\tilde{d}=\begin{cases} 0, & w.p.\ 1-\left(\dfrac{m_1^\alpha}{m_\alpha}\right)^{1/(\alpha-1)}, \\ \left(\dfrac{m_\alpha}{m_1}\right)^{1/(\alpha-1)}, & w.p.\ \left(\dfrac{m_1^\alpha}{m_\alpha}\right)^{1/(\alpha-1)}. \end{cases} \tag{3.6}$$

Proof. We show tightness of the value in (3.5) by constructing a primal and dual feasible solution that attains it for the specified range of q. Observe that the demand distribution in (3.6) is a feasible distribution in the set $\mathcal{F}_{1,\alpha}$ and attains the expected value in (3.5). Construct a dual solution as follows:

$$y_0 = 0,\ y_1 = 1 - q\left(\frac{\alpha}{\alpha-1}\right)\left(\frac{m_1}{m_\alpha}\right)^{1/(\alpha-1)} \text{ and } y_\alpha = \frac{q}{\alpha-1}\left(\frac{m_1}{m_\alpha}\right)^{\alpha/(\alpha-1)}. \tag{3.7}$$

We first verify that this solution is dual feasible for $q \in [0, ((\alpha-1)/\alpha)(m_\alpha/m_1)^{1/(\alpha-1)}]$. Note that for such a q, we have $y_1 \geq 0$. Moreover, $y_0 = 0$ and $y_\alpha \geq 0$, implying that the first dual feasibility constraint in (3.4) is satisfied. The second dual feasibility constraint can be expressed as:

$$\min_{w \geq 0} \left(y_0 + q + (y_1 - 1)w + y_\alpha w^\alpha\right) \geq 0. \tag{3.8}$$

This constraint is satisfied at equality for $q = 0$, since $y_1 = 1$ and $y_\alpha = 0$ in this case. Next, we focus on the case with $q > 0$. Since $y_0 = 0$, $y_\alpha > 0$ and $y_1 \in [0, 1)$, the minimum value in (3.8) is attained at:

$$w^* = \left(\frac{1-y_1}{\alpha y_\alpha}\right)^{1/(\alpha-1)} = \left(\frac{m_\alpha}{m_1}\right)^{1/(\alpha-1)}.$$

Substituting in the given choice of the dual variables and w^*, the left hand side of (3.8) reduces to:

$$y_0 + q + (y_1 - 1)w^* + y_\alpha w^{*\alpha} = q - q\left(\frac{\alpha}{\alpha-1}\right) + \frac{q}{\alpha-1} = 0,$$

implying the feasibility of the second dual constraint. Finally, we verify that this dual feasible solution is optimal, since the objective value of the dual feasible solution is given as:

$$y_0 + y_1 m_1 + y_\alpha m_\alpha = m_1 - q\left(\frac{m_1^\alpha}{m_\alpha}\right)^{1/(\alpha-1)},$$

which is equal to the objective value of the primal solution. □

When $\alpha = 2$, the formula in Proposition 3.1 reduces precisely to the second term in the worst-case expected value in (1.3) as developed by Scarf [34]. Proposition 3.1 indicates that for any $\alpha > 1$, there is always a range of q around 0, where the worst-case value decreases linearly with q. Building on this, the next proposition identifies a range of critical ratio η around zero, where it optimal to order nothing.

Proposition 3.2. *Define* $\eta_0 := 1 - (m_1^\alpha/m_\alpha)^{1/(\alpha-1)} \in (0, 1)$*. The optimal order quantity to the distributionally robust newsvendor problem in (3.2) satisfies the following properties:*

(a) For any critical ratio in the range $\eta \in [0, \eta_0)$*, the optimal order quantity is* $q_\eta^* = 0$.

(b) For any critical ratio in the range $\eta \in (\eta_0, 1)$*, the optimal order quantity is strictly positive and satisfies the condition:*

$$q_\eta^* \geq q_0 := \left(\frac{\alpha-1}{\alpha}\right)\left(\frac{m_\alpha}{m_1}\right)^{1/(\alpha-1)}.$$

(c) When $\eta = \eta_0$*, the set of optimal order quantities contains all the values in the range* $[0, q_0]$.

Proof. The optimal solution to the distributionally robust newsvendor problem in (3.2) is given as:

$$\min\left\{\min_{0\le q\le q_0}(1-\eta)q+\Pi_{1,\alpha}(q),\ \min_{q>q_0}(1-\eta)q+\Pi_{1,\alpha}(q)\right\}, \tag{3.9}$$

where $\Pi_{1,\alpha}(q)=\sup_{F\in\mathcal{F}_{1,\alpha}}\mathbb{E}_F[\tilde{d}-q]_+$. Using Proposition 3.1, the first term in (3.9) reduces to:

$$\min_{0\le q\le q_0}(1-\eta)q+\Pi_{1,\alpha}(q) \quad = \quad \min_{0\le q\le q_0} m_1-q\left(\eta-\eta_0\right). \tag{3.10}$$

The minimizer in (3.10) is 0 when $\eta<\eta_0$ and the corresponding objective value is m_1. Since the objective function in (3.2) is convex in q and increasing in $[0,q_0]$, the global minimum is also attained at $q^*_\eta=0$. Similarly, the minimizer in (3.10) is q_0 when $\eta>\eta_0$, since the function is strictly decreasing in this range. Moreover, since the objective function is convex, the global minimum in this case is attained at some $q^*_\eta\ge q_0>0$. Finally, when $\eta=\eta_0$, the minimum value in (3.10) is attained for all $q\in[0,q_0]$. Since the objective function is convex, the global minimum must also be attained at these values. □

Proposition 3.2 shows that for any finite α, regardless of how high the order of the specified moment in the ambiguity set is, there is always a certain range of small critical ratios where it is optimal to order zero for the moment based ambiguity set. The corresponding F^* distribution thus has a finite probability mass at a demand of zero. When critical ratios are larger than the threshold value, the optimal order quantity is strictly positive. In the next two sections, we identify lower and upper bounds on the worst-case value that help characterize F^* for large values of q.

3.2 Lower Bound for Large Values of q

To develop the lower bound, we first consider a related ambiguity set that was studied by Grundy [23] with fixing only the αth moment:

$$\mathcal{F}_\alpha=\left\{F\in\mathbb{M}(\Re_+):\int_0^\infty \mathrm{d}F(w)=1,\ \int_0^\infty w^\alpha\,\mathrm{d}F(w)=m_\alpha\right\}, \tag{3.11}$$

where $\mathcal{F}_{1,\alpha}\subseteq\mathcal{F}_\alpha$. Grundy [23] characterized the unique two point distribution that attains the bound in $\sup_{F\in\mathcal{F}_\alpha}\mathbb{E}_F[\tilde{d}-q]_+$. Given a value $q>(\alpha-1)m_\alpha^{1/\alpha}/\alpha$, the worst-case demand distribution was characterized as follows:

$$\tilde{d}_q=\begin{cases}\dfrac{q\alpha}{\alpha-1}, & \text{w.p. } \dfrac{(\alpha-1)^\alpha m_\alpha}{\alpha^\alpha q^\alpha},\\ 0, & \text{otherwise,}\end{cases} \tag{3.12}$$

while for $0 \leq q \leq (\alpha - 1)m_\alpha^{1/\alpha}/\alpha$, the worst-case demand distribution is degenerate with the mass at the point $m_\alpha^{1/\alpha}$. The corresponding worst-case expected value is given as:

$$\sup_{F \in \mathcal{F}_\alpha} \mathbb{E}_F[\tilde{d} - q]_+ = \begin{cases} \dfrac{m_\alpha}{\alpha}\left(\dfrac{\alpha - 1}{\alpha q}\right)^{\alpha - 1}, & \text{if } q > \dfrac{\alpha - 1}{\alpha} m_\alpha^{1/\alpha}, \\ m_\alpha^{1/\alpha} - q, & \text{if } 0 \leq q \leq \dfrac{\alpha - 1}{\alpha} m_\alpha^{1/\alpha}. \end{cases} \tag{3.13}$$

The worst-case distribution in this ambiguity set also depends on q, as in Scarf's model. In the next proposition, we derive a lower bound on the worst-case expected value by modifying the two point distribution in (3.12) to a three point distribution to make it feasible for the ambiguity set $\mathcal{F}_{1,\alpha}$. This gives a lower bound on the worst-case expected value for large values of q.

Proposition 3.3. *Given an ambiguity set $\mathcal{F}_{1,\alpha}$, the following lower bound is valid:*

$$\sup_{F \in \mathcal{F}_{1,\alpha}} \mathbb{E}_F[\tilde{d} - q]_+ \quad \geq \quad \frac{(m_\alpha - m_1^\alpha)}{\alpha^\alpha q^{\alpha - 1}}(\alpha - 1)^{\alpha - 1}, \quad \forall q > \underline{q}(m_1, m_\alpha, \alpha), \tag{3.14}$$

where:

$$\underline{q}(m_1, m_\alpha, \alpha) \quad := \quad \left(\left(\frac{m_\alpha - m_1^\alpha}{m_1}\right)\left(\frac{\alpha - 1}{\alpha}\right)^{\alpha - 1} + m_1^{\alpha - 1}\right)^{1/(\alpha - 1)}. \tag{3.15}$$

Proof. We derive the lower bound through the construction of a three point feasible distribution. In Step 1, we provide a three point distribution by a modification of the two point worst-case distribution in (3.12) such that the moment constraints are met while in Step 2, we show that this defines a valid probability distribution for large values of q. Evaluating the objective value provides the lower bound. *Step 1:* Consider a three point random variable $\tilde{d}$ with a distribution defined as follows:

$$\tilde{d} = \begin{cases} \dfrac{q\alpha}{\alpha - 1}, & \text{w.p. } \dfrac{(m_\alpha - m_1^\alpha)}{\alpha^\alpha q^\alpha}(\alpha - 1)^\alpha, \\ w, & \text{w.p. } p, \\ 0, & \text{w.p. } 1 - p - \dfrac{(m_\alpha - m_1^\alpha)}{\alpha^\alpha q^\alpha}(\alpha - 1)^\alpha, \end{cases} \tag{3.16}$$

where we choose particular values of w and p to ensure feasibility. We choose a w such that $0 < w < q$, and along with it a p satisfying the first and αth moment constraints as discussed hence. To ensure the αth moment constraint:

$$\begin{aligned} m_\alpha &= \mathbb{E}[\tilde{d}^\alpha], \\ &= \left(\frac{q\alpha}{\alpha - 1}\right)^\alpha \frac{(m_\alpha - m_1^\alpha)}{\alpha^\alpha q^\alpha}(\alpha - 1)^\alpha + w^\alpha p, \\ &= m_\alpha - m_1^\alpha + w^\alpha p. \end{aligned}$$

This gives rise to a condition that w and p must satisfy:

$$w^\alpha p \quad = \quad m_1^\alpha. \tag{3.17}$$

We next ensure the first moment constraint for the distribution is met as follows:

$$\begin{aligned} m_1 &= \mathbb{E}[\tilde{d}], \\ &= \left(\frac{q\alpha}{\alpha-1}\right)\frac{(m_\alpha - m_1^\alpha)}{\alpha^\alpha q^\alpha}(\alpha-1)^\alpha + wp, \\ &= \frac{(m_\alpha - m_1^\alpha)}{\alpha^{\alpha-1}q^{\alpha-1}}(\alpha-1)^{\alpha-1} + wp. \end{aligned}$$

This gives rise to a second condition that w and p must satisfy:

$$wp = m_1 - \frac{(m_\alpha - m_1^\alpha)}{\alpha^{\alpha-1}q^{\alpha-1}}(\alpha-1)^{\alpha-1}. \quad (3.18)$$

Solving the two simultaneous equations (3.17) and (3.18) gives:

$$w = \frac{m_1^{\alpha/(\alpha-1)}}{\left(m_1 - \frac{(m_\alpha - m_1^\alpha)}{\alpha^{\alpha-1}q^{\alpha-1}}(\alpha-1)^{\alpha-1}\right)^{1/(\alpha-1)}}, \quad (3.19)$$

$$p = \frac{\left(m_1 - \frac{(m_\alpha - m_1^\alpha)}{\alpha^{\alpha-1}q^{\alpha-1}}(\alpha-1)^{\alpha-1}\right)^{\alpha/(\alpha-1)}}{m_1^{\alpha/(\alpha-1)}}. \quad (3.20)$$

Step 2: Note that the demand realization w in (3.19) is a strictly decreasing function of q, and we can check that $w < q$ when $q > \underline{q}(m_1, m_\alpha, \alpha)$. Hence for $q > \underline{q}(m_1, m_\alpha, \alpha)$ the expected value of the objective function is computed with one support point above q and given by:

$$\begin{aligned} \mathbb{E}[\tilde{d} - q]_+ &= \left(\frac{q\alpha}{\alpha-1} - q\right)\frac{m_\alpha - m_1^\alpha}{\alpha^\alpha q^\alpha}(\alpha-1)^\alpha, \\ &= \frac{(m_\alpha - m_1^\alpha)(\alpha-1)^{\alpha-1}}{\alpha^\alpha q^{\alpha-1}}, \end{aligned}$$

which corresponds to the lower bound on the expected value. To complete the proof, we need to ensure that (3.16) corresponds to a valid probability measure for the chosen w and p for all $q > \underline{q}(m_1, m_\alpha, \alpha)$. Since by construction the probabilities add up to one, we need only check that they are all non-negative:

(a) The probability of the first atom is trivially strictly positive.

(b) The probability of the second atom p in (3.20) is strictly positive when:

$$q > \underline{q}_1 := \left(\frac{m_\alpha - m_1^\alpha}{m_1}\right)^{1/(\alpha-1)}\left(\frac{\alpha-1}{\alpha}\right). \quad (3.21)$$

Condition (3.21) is satisfied since $q > \underline{q}(m_1, m_\alpha, \alpha)$ and $m_1 > 0$ implies

$$q > \underline{q}(m_1, m_\alpha, \alpha) = \left(\underline{q}_1^{\alpha-1} + m_1^{\alpha-1}\right)^{1/(\alpha-1)} > \underline{q}_1.$$

(c) Finally to verify that the probability of the atom 0 given by $1 - p - (m_\alpha - m_1^\alpha)(\alpha-1)^\alpha/(\alpha^\alpha q^\alpha) > 0$, plugging in the value of p, we need to verify the following for $q > \underline{q}(m_1, m_\alpha, \alpha)$:

$$\left(1 - (m_\alpha - m_1^\alpha)\left(\frac{\alpha-1}{\alpha q}\right)^\alpha\right)^{\alpha-1} \quad > \quad \left(1 - \left(\frac{m_\alpha - m_1^\alpha}{m_1}\right)\left(\frac{\alpha-1}{\alpha q}\right)^{\alpha-1}\right)^\alpha ?$$

or equivalently:

$$1 - (m_\alpha - m_1^\alpha)\left(\frac{\alpha-1}{\alpha q}\right)^\alpha \quad > \quad \left(1 - \left(\frac{m_\alpha - m_1^\alpha}{m_1}\right)\left(\frac{\alpha-1}{\alpha q}\right)^{\alpha-1}\right)^{\alpha/(\alpha-1)} ?$$

Since $q > \underline{q}(m_1, m_\alpha, \alpha)$:

$$\begin{aligned} 1 - (m_\alpha - m_1^\alpha)\left(\frac{\alpha-1}{\alpha q}\right)^\alpha \quad &= \quad 1 - \left(\frac{m_\alpha - m_1^\alpha}{m_1}\right)\left(\frac{\alpha-1}{\alpha q}\right)^{\alpha-1}\left(\frac{m_1(\alpha-1)}{\alpha q}\right), \\ &> \quad 1 - \left(\frac{m_\alpha - m_1^\alpha}{m_1}\right)\left(\frac{\alpha-1}{\alpha q}\right)^{\alpha-1}, \\ &> \quad \left(1 - \left(\frac{m_\alpha - m_1^\alpha}{m_1}\right)\left(\frac{\alpha-1}{\alpha q}\right)^{\alpha-1}\right)^{\alpha/(\alpha-1)}, \end{aligned}$$

where the first inequality holds since $q > \underline{q}(m_1, m_\alpha, \alpha) > m_1 > m_1(\alpha-1)/\alpha$ and the second inequality holds since the term in the brackets is strictly less than 1 for $q > \underline{q}(m_1, m_\alpha, \alpha)$ and the exponent is greater than 1.

Hence the distribution is feasible in $\mathcal{F}_{1,\alpha}$ for large values of q. This leads to the desired result. $\square$

3.3 Upper Bound for Large Values of q

To develop the upper bound on the worst-case expected value, we consider the dual formulation for the moment problem.

Proposition 3.4. *Consider the ambiguity set $\mathcal{F}_{1,\alpha}$.*

(a) When $\alpha \in (2, \infty)$, the following upper bound is valid:

$$\sup_{F \in \mathcal{F}_{1,\alpha}} \mathbb{E}_F[\tilde{d} - q]_+ \quad \leq \quad \frac{(m_\alpha - m_1^\alpha)}{\alpha^\alpha q^{\alpha-1} - \alpha^2 m_1^{\alpha-1}(\alpha-1)^{\alpha-1}}(\alpha-1)^{\alpha-1}, \quad \forall q > \overline{q}(m_1, \alpha), \tag{3.22}$$

where:

$$\overline{q}(m_1, \alpha) \quad = \quad m_1(\alpha-1)\alpha^{(2-\alpha)/(\alpha-1)}. \tag{3.23}$$

(b) When $\alpha \in (1, 2)$, for all $\epsilon \in (0, (\alpha/(\alpha-1))^{\alpha-1} - \alpha)$, the following upper bound is valid:

$$\sup_{F \in \mathcal{F}_{1,\alpha}} \mathbb{E}_F[\tilde{d} - q]_+ \quad \leq \quad \frac{(m_\alpha - m_1^\alpha)}{\alpha^\alpha q^{\alpha-1} - (\alpha+\epsilon)\alpha m_1^{\alpha-1}(\alpha-1)^{\alpha-1}}(\alpha-1)^{\alpha-1}, \quad \forall q > \overline{q}(m_1, \alpha, \epsilon). \tag{3.24}$$

where $\overline{q}(m_1, \alpha, \epsilon) = m_1(\alpha - 1)x^*/\alpha$ *and* x^* *is defined as the unique root in the interval* $((\alpha + \epsilon)^{1/(\alpha-1)}, \infty)$ *to the equation:*

$$x^\alpha - (\alpha + \epsilon)\, x + 1 - \left(x^{\alpha-1} - \alpha - \epsilon + 1\right)^{\alpha/(\alpha-1)} \;=\; 0. \tag{3.25}$$

Proof. We derive the upper bound by constructing a dual feasible solution to (3.4) as follows. Define $y_0,\ y_1$ and y_α as:

$$y_0 = \frac{(\alpha-1)m_1^\alpha(\alpha-1)^{\alpha-1}}{\alpha^\alpha(q^{\alpha-1}-K)},\ y_1 = \frac{-\alpha m_1^{\alpha-1}(\alpha-1)^{\alpha-1}}{\alpha^\alpha(q^{\alpha-1}-K)} \text{ and } y_\alpha = \frac{(\alpha-1)^{\alpha-1}}{\alpha^\alpha(q^{\alpha-1}-K)}, \tag{3.26}$$

where we choose $K > 0$ in a manner to be specified later. First verify that this forms a dual feasible solution for q satisfying $q > K^{1/(\alpha-1)}$, by checking each of the dual constraints. Observe that the dual feasibility constraints are equivalent to the following conditions:

$$\min_{w\geq 0}\left(y_0 + y_1 w + y_\alpha w^\alpha\right) \geq 0 \text{ and } \min_{w\geq 0}\left(y_0 + q + (y_1 - 1)w + y_\alpha w^\alpha\right) \geq 0 \tag{3.27}$$

Since the values of the dual variables in (3.26) satisfy $y_\alpha > 0$ and $y_1 < 0$ for $q > K^{1/(\alpha-1)}$, the minimum value in the first dual constraint is obtained at $w^* = (-y_1/(\alpha y_\alpha))^{1/(\alpha-1)}$. Substituting this value, the first dual feasibility constraint is equivalent to verifying the condition:

$$\begin{aligned} y_0 \;&\geq\; -y_1\left(\frac{-y_1}{\alpha y_\alpha}\right)^{1/(\alpha-1)} - y_\alpha\left(\frac{-y_1}{\alpha y_\alpha}\right)^{\alpha/(\alpha-1)}, \\ &=\; \frac{(-y_1)^{\alpha/(\alpha-1)}}{(\alpha y_\alpha)^{1/(\alpha-1)}}\left(\frac{\alpha-1}{\alpha}\right). \end{aligned}$$

The choice of dual variables in (3.26) satisfy this condition at equality since:

$$\begin{aligned} y_0 - \frac{(-y_1)^{\alpha/(\alpha-1)}}{(\alpha y_\alpha)^{1/(\alpha-1)}}\left(\frac{\alpha-1}{\alpha}\right) \;&=\; \frac{(\alpha-1)m_1^\alpha(\alpha-1)^{\alpha-1}}{\alpha^\alpha(q^{\alpha-1}-K)} - \left(\frac{\alpha-1}{\alpha}\right)\frac{(\alpha m_1^{\alpha-1}(\alpha-1)^{\alpha-1})^{\alpha/(\alpha-1)}}{(\alpha(\alpha-1)^{\alpha-1})^{1/(\alpha-1)}\alpha^\alpha(q^{\alpha-1}-K)}, \\ &=\; 0. \end{aligned}$$

Furthermore as $y_\alpha > 0$, the minimum value in the second dual constraint is obtained at:

$$w^* = ((1-y_1)/(\alpha y_\alpha))^{1/(\alpha-1)}.$$

Substituting this in, the second dual feasibility constraint is equivalent to verifying the following:

$$\delta(q) \;:=\; y_0 + q - \frac{(1-y_1)^{\alpha/(\alpha-1)}}{(\alpha y_\alpha)^{1/(\alpha-1)}}\left(\frac{\alpha-1}{\alpha}\right) \geq 0. \tag{3.28}$$

The choice of dual variables in (3.26) leads to the following expression:

$$\begin{aligned}\delta(q) &= \frac{(\alpha-1)m_1^\alpha(\alpha-1)^{\alpha-1}}{\alpha^\alpha(q^{\alpha-1}-K)}+q-\frac{\left(\alpha^\alpha(q^{\alpha-1}-K)+nm_1^{\alpha-1}(\alpha-1)^{\alpha-1}\right)^{\alpha/(\alpha-1)}}{(\alpha(\alpha-1))^{1/(\alpha-1)}\,\alpha^\alpha(q^{\alpha-1}-K)}\left(\frac{\alpha-1}{\alpha}\right),\\ &= \frac{m_1^\alpha(\alpha-1)^\alpha+\alpha^\alpha q(q^{\alpha-1}-K)-\left(\alpha^{\alpha-1}(q^{\alpha-1}-K)+m_1^{\alpha-1}(\alpha-1)^{\alpha-1}\right)^{\alpha/(\alpha-1)}}{\alpha^\alpha(q^{\alpha-1}-K)}.\end{aligned}$$

Hence we need to verify for $q > K^{1/(\alpha-1)}$ and large enough, if the following inequality holds:

$$m_1^\alpha(\alpha-1)^\alpha+\alpha^\alpha q(q^{\alpha-1}-K)-\left(\alpha^{\alpha-1}(q^{\alpha-1}-K)+m_1^{\alpha-1}(\alpha-1)^{\alpha-1}\right)^{\alpha/(\alpha-1)}>0?$$

Let $C = m_1(\alpha-1)/\alpha > 0$. Dividing by $m_1^\alpha(\alpha-1)^\alpha$, this condition is equivalent to verifying that for q large enough, the following inequality holds:

$$\triangle(q) \;:=\; \left(\left(\frac{q}{C}\right)^\alpha - q\left(\frac{K}{C^\alpha}\right)+1\right)-\left(\left(\frac{q}{C}\right)^{\alpha-1}-\left(\frac{K}{C^{\alpha-1}}\right)+1\right)^{\alpha/(\alpha-1)}>0?$$

We consider two cases:

(a) $\alpha \in (2,\infty)$: In this case, set the constant $K = \alpha C^{\alpha-1} = \alpha(m_1(\alpha-1)/\alpha)^{\alpha-1}$. Then, we need to verify that for $q > \overline{q}(m_1, m_\alpha, \alpha)$ (which will be identified next), the following inequality holds:

$$\triangle(q) \;=\; \left(\left(\frac{q}{C}\right)^\alpha-\alpha\left(\frac{q}{C}\right)+1\right)-\left(\left(\frac{q}{C}\right)^{\alpha-1}-\alpha+1\right)^{\alpha/(\alpha-1)}>0?$$

By setting $\overline{q}(m_1, m_\alpha, \alpha) = K^{1/(\alpha-1)} = \alpha^{1/(\alpha-1)}C = m_1(\alpha-1)\alpha^{(2-\alpha)/(\alpha-1)}$, we observe that the condition is satisfied at equality, since:

$$\begin{aligned}\triangle(\overline{q}(m_1,m_\alpha,\alpha)) &= \left(\alpha^{\alpha/(\alpha-1)}-\alpha^{\alpha/(\alpha-1)}+1\right)-(\alpha-\alpha+1)^{\alpha/(\alpha-1)},\\ &= 0.\end{aligned} \tag{3.29}$$

Furthermore, the derivative of the function $\triangle(q)$ with respect to q satisfies:

$$\begin{aligned}\frac{d}{dq}\triangle(q) &= \frac{\alpha}{C}\left(\left(\frac{q}{C}\right)^{\alpha-1}-1-\left(\frac{q}{C}\right)^{\alpha-2}\left(\left(\frac{q}{C}\right)^{\alpha-1}-\alpha+1\right)^{1/(\alpha-1)}\right),\\ &= \frac{\alpha q^{\alpha-1}}{C^\alpha}\left(1-\frac{1}{(q/C)^{\alpha-1}}-\left(1-\frac{\alpha-1}{(q/C)^{\alpha-1}}\right)^{1/(\alpha-1)}\right),\\ &> 0, \qquad \forall q > \overline{q}(m_1,m_\alpha,\alpha),\end{aligned} \tag{3.30}$$

where the first equality is obtained by differentiating the function $\triangle(q)$, the second equality is obtained by straightforward algebraic manipulations and the inequality is obtained by using Bernoulli's inequality $(1-x)^t > 1-tx$ which is valid for $t > 1$ and $0 < x \leq 1$ and setting $t = \alpha-1$ and $x = (C/q)^{\alpha-1}$. Note that since $\alpha > 2$, $q > \alpha^{1/(\alpha-1)}C \geq C$ and the conditions $t > 1$ and $0 \leq x \leq 1$ are satisfied. This implies that the derivative of the function is positive for all values of $q > \overline{q}(m_1, m_\alpha, \alpha)$. Combining (3.29) and (3.30) implies that $\triangle(q)$ is positive for all values of q

above $\overline{q}(m_1, m_\alpha, \alpha)$. Hence the constructed solution is dual feasible for q above $\overline{q}(m_1, m_\alpha, \alpha)$. The objective function value of this dual feasible solution reduces to the form below which yields the desired result:

$$y_0 + y_1 m_1 + y_\alpha m_\alpha = \frac{(m_\alpha - m_1^\alpha)(\alpha-1)^{\alpha-1}}{\alpha^\alpha q^{\alpha-1} - \alpha^2 m_1^{\alpha-1}(\alpha-1)^{\alpha-1}}.$$

(b) $\alpha \in (1,2)$: Note that unlike the $\alpha > 2$ case, setting $K = \alpha C^{\alpha-1}$ does not ensure a dual feasible solution for $\alpha \in (1,2)$. To see this, observe that by applying the generalized binomial expansion, the term $\triangle(q)$ reduces to:

$$\triangle(q) = \left(\left(\frac{q}{C}\right)^\alpha - \alpha\left(\frac{q}{C}\right) + 1\right) - \sum_{k=0}^{\infty} \binom{\frac{\alpha}{\alpha-1}}{k} \left(\frac{q}{C}\right)^{\alpha-(\alpha-1)k} (1-\alpha)^k,$$

where $\binom{r}{k}$ is defined as $r(r-1)\ldots(r-k+1)/k!$ for general values of r (not necessarily integer). Expanding the first few terms, gives:

$$\begin{aligned}\triangle(q) &= \left(\frac{q}{C}\right)^\alpha - \alpha\left(\frac{q}{C}\right) + 1 - \left(\frac{q}{C}\right)^\alpha + \alpha\left(\frac{q}{C}\right) - \frac{\alpha}{2}\left(\frac{q}{C}\right)^{2-\alpha} - \sum_{k=3}^{\infty} \binom{\frac{\alpha}{\alpha-1}}{k} \left(\frac{q}{C}\right)^{\alpha-(\alpha-1)k} (1-\alpha)^k, \\ &= -\frac{\alpha}{2}\left(\frac{q}{C}\right)^{2-\alpha} + 1 - \sum_{k=3}^{\infty} \binom{\frac{\alpha}{\alpha-1}}{k} \left(\frac{q}{C}\right)^{\alpha-(\alpha-1)k} (1-\alpha)^k,\end{aligned}$$

where the leading term of the expression with exponent $2-\alpha > 0$ has a negative coefficient. This implies that for large values of q, $\triangle(q)$ becomes negative. To deal with this technical issue, we modify the dual solution by choosing for a strictly positive small $\epsilon > 0$, the value $K = (\alpha+\epsilon)C^{\alpha-1}$. In this case, we need to verify that for q above a certain value (determined later), the following inequality holds:

$$\triangle(q) = \left(\left(\frac{q}{C}\right)^\alpha - (\alpha+\epsilon)\left(\frac{q}{C}\right) + 1\right) - \left(\left(\frac{q}{C}\right)^{\alpha-1} - \alpha - \epsilon + 1\right)^{\alpha/(\alpha-1)} > 0?$$

Note that, by applying the generalized binomial expansion, the term $\triangle(q)$ reduces to:

$$\begin{aligned}\triangle(q) &= \left(\frac{q}{C}\right)^\alpha - (\alpha+\epsilon)\left(\frac{q}{C}\right) + 1 - \sum_{k=0}^{\infty} \binom{\frac{\alpha}{\alpha-1}}{k} \left(\frac{q}{C}\right)^{\alpha-(\alpha-1)k} (1-\alpha-\epsilon)^k, \\ &= \left(\frac{\epsilon}{\alpha-1}\right)\left(\frac{q}{C}\right) + 1 - \sum_{k=2}^{\infty} \binom{\frac{\alpha}{\alpha-1}}{k} \left(\frac{q}{C}\right)^{\alpha-(\alpha-1)k} (1-\alpha-\epsilon)^k,\end{aligned}$$

where the leading term of the expression with exponent 1 has a positive coefficient. The derivative of the function $\triangle(q)$ with respect to q is given by:

$$\frac{d}{dq}\triangle(q) = \frac{\alpha}{C}\left(\left(\frac{q}{C}\right)^{\alpha-1} - 1 - \frac{\epsilon}{\alpha} - \left(\frac{q}{C}\right)^{\alpha-2}\left(\left(\frac{q}{C}\right)^{\alpha-1} - \alpha - \epsilon + 1\right)^{1/(\alpha-1)}\right). \tag{3.31}$$

Furthermore, the second derivative of the function $\triangle(q)$ with respect to q is given by:

$$\frac{d^2}{dq^2}\triangle(q) = \frac{\alpha q^{\alpha-3}}{C^{\alpha-1}} h(q), \tag{3.32}$$

where $h(q)$ is defined as:

$$\begin{aligned} h(q) &:= (\alpha-1)\left(\frac{q}{C}\right) + (2-\alpha)\left(\left(\frac{q}{C}\right)^{\alpha-1} - \alpha - \epsilon + 1\right)^{1/(\alpha-1)} \\ &\quad - \left(\frac{q}{C}\right)^{\alpha-1}\left(\left(\frac{q}{C}\right)^{\alpha-1} - \alpha - \epsilon + 1\right)^{(2-\alpha)/(\alpha-1)}, \\ &> 0, \qquad \forall q > (\alpha+\epsilon-1)^{1/(\alpha-1)}C. \end{aligned} \tag{3.33}$$

The nonnegativity of the second derivative follows from using the strict form of the weighted arithmetic and geometric mean inequality given by $\lambda x_1 + (1-\lambda)x_2 > x_1^{\lambda}x_2^{1-\lambda}$ which is valid for $\lambda \in (0,1)$ and $x_1, x_2 > 0$, $x_1 \neq x_2$ by setting $\lambda = \alpha - 1 \in (0,1)$ with $\alpha \in (1,2)$, $x_1 = q/C$ and $x_2 = ((q/C)^{\alpha-1} - \alpha - \epsilon + 1)^{1/(\alpha-1)}$. To finish the proof, observe $\underline{q} = (\alpha+\epsilon)^{1/(\alpha-1)}C$ is a root of the equation $\triangle(q) = 0$, since:

$$\begin{aligned} \triangle(\underline{q}) &= \left((\alpha+\epsilon)^{\alpha/(\alpha-1)} - (\alpha+\epsilon)^{\alpha/(\alpha-1)} + 1\right) - (\alpha+\epsilon-\alpha-\epsilon+1)^{\alpha/(\alpha-1)}, \\ &= 0. \end{aligned} \tag{3.34}$$

Furthermore, the derivative:

$$\begin{aligned} \frac{d}{dq}\triangle(q)|_{q=\underline{q}} &= \frac{\alpha}{C}\left(\alpha+\epsilon-1-\frac{\epsilon}{\alpha} - (\alpha+\epsilon)^{(\alpha-2)/(\alpha-1)}\left(\alpha+\epsilon-\alpha-\epsilon+1\right)^{1/(\alpha-1)}\right), \\ &= \frac{\alpha+\epsilon}{C}\left(\alpha - 1 - \alpha\left(\alpha+\epsilon\right)^{-1/(\alpha-1)}\right), \\ &< 0, \end{aligned} \tag{3.35}$$

where the first equation is obtained by plugging in $\underline{q}$ into (3.31), the second equation is obtained from straightforward algebraic manipulations and the inequality is obtained by observing that for $0 < \epsilon < (\alpha/(\alpha-1))^{\alpha-1} - \alpha$, the right hand side is negative. Since the function is strictly convex from (3.33) with $\lim_{q\to\infty}\triangle(q) = \infty$ and one of the roots is given by $\underline{q}$ where the derivative is negative, the function is positive for all values of q above the second root $\overline{q}$ to the equation:

$$\triangle(\overline{q}) = 0,$$

which lies in the range $(\underline{q}, \infty)$. The objective function value of this dual feasible solution reduces to the form below which yields the desired result for $\alpha \in (1,2)$:

$$y_0 + y_1 m_1 + y_\alpha m_\alpha = \frac{(m_\alpha - m_1^\alpha)(\alpha-1)^{\alpha-1}}{\alpha^\alpha q^{\alpha-1} - (\alpha+\epsilon)\alpha m_1^{\alpha-1}(\alpha-1)^{\alpha-1}}.$$

□

3.4 Numerical Example

We provide a numerical illustration of the quality of the bounds from Propositions 3.3 and 3.4 respectively. To compute the worst-case expected value, we solve the dual formulation in (3.4) using a semidefinite program (SDP) for rational values of α. Assume that $\alpha = p/q$, where p and q are strictly positive integers. Then, the dual formulation is given as:

$$\begin{aligned} \inf \quad & y_0 + y_1 m_1 + y_{p/q} m_{p/q} \\ \text{s.t.} \quad & y_0 + y_1 w + y_{p/q} w^{p/q} \geq 0, \qquad \forall w \geq 0, \\ & y_0 + q + (y_1 - 1) w + y_{p/q} w^{p/q} \geq 0, \quad \forall w \geq 0. \end{aligned} \tag{3.36}$$

Applying the transformation by defining the variable $d = w^{1/q}$ or equivalently $d = e^q$, we obtain a reformulation of the dual problem as:

$$\begin{aligned} \inf \quad & y_0 + y_1 m_1 + y_{p/q} m_{p/q} \\ \text{s.t.} \quad & y_0 + y_1 d^q + y_{p/q} d^p \geq 0, \qquad \forall d \geq 0, \\ & y_0 + q + (y_1 - 1) d^q + y_{p/q} d^p \geq 0, \quad \forall d \geq 0, \end{aligned} \tag{3.37}$$

The constraints in (3.37) are the standard nonnegativity conditions on univariate polynomials over the half-line for which semidefinite representations are available (see Bertsimas and Popescu [3], Lasserre [27], Nesterov [32]). For example, with $\alpha = 3$ ($p = 3, q = 1$), the semidefinite programming formulation is given as:

$$\begin{aligned} \inf_{y_0,y_1,y_3,a_1,b_1,c_1,a_2,b_2,c_2} \quad & y_0 + y_1 m_1 + y_3 m_3 \\ \text{s.t.} \quad & \begin{bmatrix} y_0 & 0 & a_1 & b_1 \\ 0 & y_1 - 2a_1 & -b_1 & c_1 \\ a_1 & -b_1 & -2c_1 & 0 \\ b_1 & c_1 & 0 & y_3 \end{bmatrix} \succeq 0, \\ & \begin{bmatrix} y_0 + q & 0 & a_2 & b_2 \\ 0 & y_1 - 1 - 2a_2 & -b_2 & c_2 \\ a_2 & -b_2 & -2c_2 & 0 \\ b_2 & c_2 & 0 & y_3 \end{bmatrix} \succeq 0. \end{aligned} \tag{3.38}$$

Assume a mean demand of 50. In addition, assume $m_3 = 125150$ and $m_{3/2} = \sqrt{125150}$ respectively where Holder's inequality requires $m_3 \geq m_{3/2}^2 \geq m_1^3$. In Figures 2 and 3, we compare the upper and lower bounds and the worst-case expected value obtained from solving the SDP. The semidefinite programs were solved in Matlab R2017a with SDPT3 version 4.0 (see Toh, Todd and Tutuncu [38, 39]). The value of the worst-case expected values and the bounds for $\alpha = 3$ are smaller than the values for $\alpha = 3/2$ for a given q as should be expected since the former ambiguity set makes stronger assumptions on the existence of moments. The figures also illustrate that the scaling behavior of the bounds as a function of q and provides the range of q from Propositions 3.3 and 3.4 over which the bounds are valid in these instances. We observe that the upper bounds are closer to the exact value in comparison to lower bound. While this suggests that it might be possible to construct stronger closed form lower bounds, especially when $\alpha < 2$, as we show in the next section, the proposed lower and upper bounds are sufficient to provide a characterization of the worst-case value for large values of q using the theory of regularly variation. In Table 1, we evaluate the effect of varying ϵ on the value of the order quantity

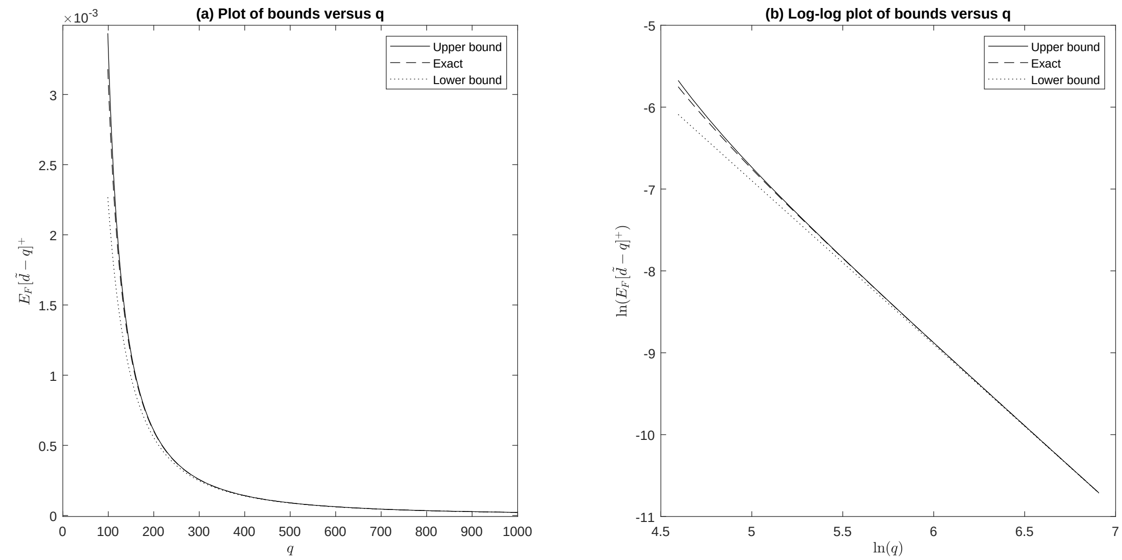

Figure 2: Plot (a) compares the upper and lower bounds with the exact bound obtained from solving a SDP as a function of q while plot (b) provides a log-log plot to characterize the scaling behavior. The mean demand is set to $m_1 = 50$ and the third moment is set to $m_3 = 125150$. The lower bound is valid for $q > 50.013$ and the upper bound is valid for $q > 57.735$.

		Percentage deviation from the worst-case expected value					
ϵ	$\overline{q}(m_1, \alpha, \epsilon)$	$q = 1000$	$q = 500$	$q = 400$	$q = 300$	$q = 200$	$q = 100$
0.07	484.59	0.374	0.035	-	-	-	-
0.10	240.67	0.864	0.809	0.700	0.412	-	-
0.15	110.51	1.692	2.126	2.251	2.348	2.221	-
0.20	65.01	2.532	3.478	3.849	4.361	5.119	5.636

Table 1: Effect of ϵ for $\alpha = 3/2$, $m_1 = 50$, $m_{3/2} = \sqrt{125150}$ on $\overline{q}(m_1, \alpha, \epsilon)$ and the percentage deviation of the upper bound from the worst-case expected value. A '-' indicates that the upper bound is not valid for the specified ϵ for the given q.

$\overline{q}(m_1, \alpha, \epsilon)$ beyond which the upper bound is valid and the percentage deviation of the upper bound over the worst-case expected value for several choices of q. As should be expected, as ϵ is decreased, $\overline{q}(m_1, \alpha, \epsilon)$ increases (the upper bound is valid for a smaller range of q values) and the upper bounds get closer to optimal.

4 Characterization of Heavy-Tail Optimality

In this section, we use the lower and upper bounds to provide a characterization of the tail of the demand distribution F^* for which the distributionally robust newsvendor order quantity remains optimal. To do

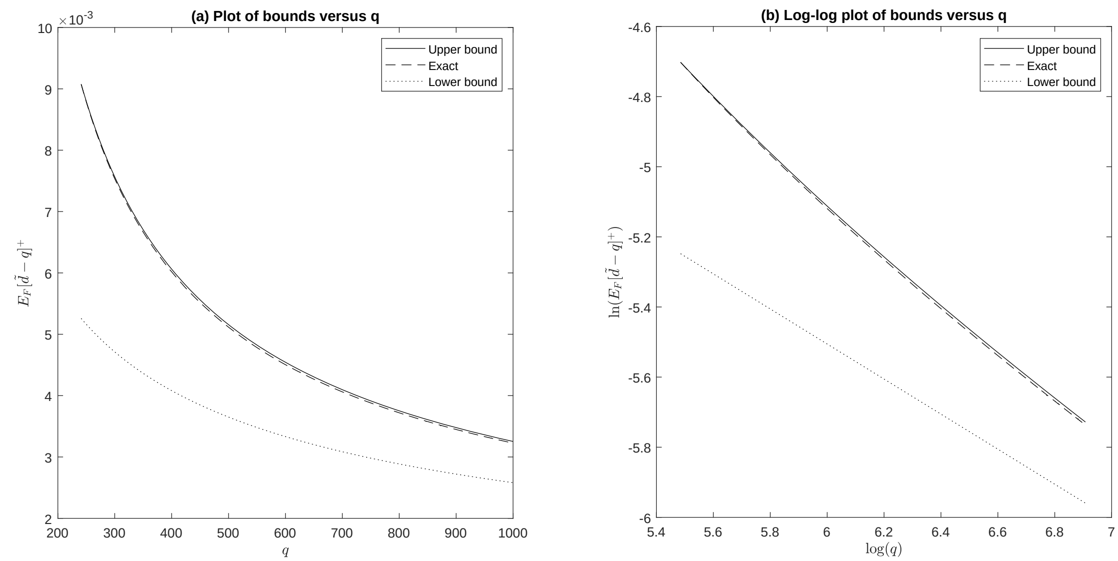

Figure 3: Plot (a) compares the upper and lower bounds with the exact bound obtained from solving a SDP as a function of q while plot (b) provides a log-log plot to characterize the scaling behavior. The mean demand is set to $m_1 = 50$ and the highest moment is set to $m_{3/2} = \sqrt{125150}$ with $\alpha = 3/2$. The bounds are plotted by setting $\epsilon = 0.1$ where the lower bound is valid for $q > 50.034$ and the upper bound is valid for $q > 240.67$.

so, we use the theory of regularly variation which is popular in the study of heavy-tailed distributions (see Bingham, Goldie and Teugels [7], de Haan [15]). The key property of such distributions is that the behaviour at infinity is similar to the behaviour of a power law distribution. A function $u : \Re_+ \to \Re_+$ is said to be regularly varying at infinity with index $\alpha \in \Re$ if for all $t > 0$, we have:

$$\lim_{x\to\infty} \frac{u(tx)}{u(x)} = t^{\alpha}.$$

We express this by $u \in \mathcal{RV}_{\alpha}$. A non-negative random variable $\tilde{d}$ with distribution function F is regularly varying if $\overline{F} := 1 - F \in \mathcal{RV}_{-\alpha}$ for some $\alpha \geq 0$. The distribution function is said to have tail parameter α if $\overline{F} \in \mathcal{RV}_{-\alpha}$. A brief overview of examples and properties of regularly varying distributions is provided in the Appendix. As we see in this section, this is exactly the type of behavior that F^* satisfies.

4.1 From the Ambiguity Set $\mathcal{F}_{1,\alpha}$ to a Regularly Varying Distribution F^*

Propositions 3.3 and 3.4 indicate that in fact the tails of the worst-case expected value are close to a power law (Pareto-like) tail. In this section, we show that there exists a regularly varying random variable $d^* \sim F^*$ which attains the worst-case expected value.

Theorem 4.1. *Given the worst-case expected value* $\Pi_{1,\alpha}(q) = \sup_{F\in\mathcal{F}_{1,\alpha}} \mathbb{E}_F[\tilde{d} - q]_+$ *where* $\Pi_{1,\alpha} : \Re_+ \to \Re_+$*, the following holds:*

(a) The worst-case expected value is regularly varying with:

$$\Pi_{1,\alpha} \in \mathcal{RV}_{-\alpha+1}.$$

(b) *There exists a random variable* $\tilde{d}^*$ *with distribution function* $F^* \notin \mathcal{F}_{1,\alpha}$ *such that*

$$\Pi_{1,\alpha}(q) = \sup_{F \in \mathcal{F}_{1,\alpha}} \mathbb{E}_F[\tilde{d} - q]_+ = \mathbb{E}_{F^*}[\tilde{d}^* - q]_+, \quad \forall q \geq 0.$$

For the distribution we have $\overline{F}^* \in \mathcal{RV}_{-\alpha}$ *with* $\mathbb{E}_{F^*}[(\tilde{d}^*)^{\alpha_1}] < \infty$ *for all* $0 \leq \alpha_1 < \alpha$ *and* $\mathbb{E}_{F^*}[(\tilde{d}^*)^{\alpha_1}] = \infty$ *if* $\alpha_1 \geq \alpha$*. Here* F^* *is independent of the choice of* q*.*

Proof. Note that the case $\alpha = 2$ boils down to Scarf's model discussed in Section 1.1. From (1.3), we have for $q \geq m_2/2m_1$,

$$\Pi_{1,2}(q) = \frac{1}{2}\left(\sqrt{q^2 - 2m_1 q + m_2} - (q - m_1)\right).$$

It is easy to check that:

$$\lim_{q\to\infty} \frac{\Pi_{1,2}(tq)}{\Pi_{1,2}(q)} = \lim_{q\to\infty} \frac{\frac{m_2}{2tq} + o(1/q)}{\frac{m_2}{2q} + o(1/q)} = \frac{1}{t}.$$

Hence $\Pi_{1,2} \in \mathcal{RV}_{-1}$ which is as claimed in (a). Moreover, from (2.3)-(2.5), we have $\tilde{d}^* \sim F^*$ satisfying (b) where $\overline{F}^* \in \mathcal{RV}_{-2}$ and is a censored t-distribution with degree of freedom $\nu = 2$ for which $\mathbb{E}_{F^*}[\tilde{d}^{*2}] = \infty$. We now concentrate on $\alpha \neq 2$ for the rest of the proof.

(a) Notice that from Proposition 3.3, for $q > \underline{q}(m_1, m_\alpha, \alpha)$, we have:

$$\Pi_{1,\alpha}(q) \geq C_1 \frac{1}{q^{\alpha-1}}, \tag{4.1}$$

where $C_1 = (m_\alpha - m_1^\alpha)\,\frac{(\alpha-1)^{\alpha-1}}{\alpha^\alpha}$. Furthermore, we have the following upper bounds:

(i) For $\alpha \in (2, \infty)$, using Proposition 3.4, for $q > \overline{q}(m_1, \alpha)$, we have:

$$\Pi_{1,\alpha}(q) \leq C_1 \frac{1}{q^{\alpha-1}}\left(1 - \frac{C_2}{q^{\alpha-1}}\right)^{-1}, \tag{4.2}$$

where $C_2 = m_1^{\alpha-1}\frac{(\alpha-1)^{\alpha-1}}{\alpha^{\alpha-2}}$.

(ii) For $\alpha \in (1, 2)$, fixing small $\epsilon > 0$ using Proposition 3.4, for $q > \overline{q}(m_1, \alpha, \epsilon)$, we have:

$$\Pi_{1,\alpha}(q) \leq C_1 \frac{1}{q^{\alpha-1}}\left(1 - \frac{C_2}{q^{\alpha-1}}\right)^{-1}, \tag{4.3}$$

where $C_2 = m_1^{\alpha-1}\frac{(\alpha-1)^{\alpha-1}}{\alpha^{\alpha-1}}(\alpha + \epsilon)$.

Now if $\alpha \in (2, \infty)$, choose $Q^* = \max(\underline{q}(m_1, m_\alpha, \alpha), \overline{q}(m_1, \alpha))$ and if $\alpha \in (1, 2)$, then fix $\epsilon > 0$ and choose $Q^* = \max(\underline{q}(m_1, m_\alpha, \alpha), \overline{q}(m_1, \alpha, \epsilon))$. Hence combining (4.1), (4.2) and (4.3), for $q > Q^*$ and $tq > Q^*$ we get:

$$t^{-\alpha+1}\left(1 - \frac{C_2}{q^{\alpha-1}}\right) \leq \frac{\Pi_{1,\alpha}(tq)}{\Pi_{1,\alpha}(q)} \leq t^{-\alpha+1}\left(1 - \frac{C_2}{(tq)^{\alpha-1}}\right)^{-1}.$$

Since $1 - C_2/(tq)^{\alpha-1} \to 1$ and $1 - C_2/q^{\alpha-1} \to 1$, as $q \to \infty$, we can infer that for $t > 0$ as Q^* is bounded:

$$\lim_{q\to\infty} \frac{\Pi_{1,\alpha}(tq)}{\Pi_{1,\alpha}(q)} = t^{-\alpha+1}.$$

Hence, $\Pi_{1,\alpha} \in \mathcal{RV}_{-\alpha+1}$.

(b) As a consequence of Theorem 2.1 and Section 3.1, page 32 in Shapiro and Kleywegt [37], we observe for any $q \geq 0$:

$$\Pi_{1,\alpha}(q) = \mathbb{E}_{F^*}[\tilde{d}^* - q]_+.$$

We can write:

$$\mathbb{E}_{F^*}[\tilde{d}^* - q]_+ = \int_q^\infty \mathbb{P}(\tilde{d}^* > w)\,\mathrm{d}w = \int_q^\infty \overline{F}^*(w)\,\mathrm{d}w, \tag{4.4}$$

where $\overline{F}^* = 1 - F^*$. From part (a), we have $\Pi_{1,\alpha} \in \mathcal{RV}_{-\alpha+1}$. Now since $-\alpha+1 < 0$ and $\overline{F}^*$ is non-increasing, using Theorem 6.2 (b) in the Appendix (the converse part of Karamata's Theorem), we have $\overline{F}^* \in \mathcal{RV}_{-\alpha}$. Note that for any $\alpha_1 \geq 0$, and some $C > 0$, we have:

$$\mathbb{E}_{F^*}[(\tilde{d}^*)^{\alpha_1}] = \int_0^C \alpha_1 w^{\alpha_1-1}\overline{F}^*(w)\,\mathrm{d}w + \int_C^\infty \alpha_1 w^{\alpha_1-1}\overline{F}^*(w)\,\mathrm{d}w.$$

The first term in the summand is bounded above by $C^{\alpha_1} < \infty$. The integrand in the second term is a function in $\mathcal{RV}_{\alpha_2}$ where $\alpha_2 = -\alpha + \alpha_1 - 1$. For $\alpha_1 < \alpha$, we have $\alpha_2 < -1$ and using Theorem 6.1, $v(C) = \int_C^\infty \alpha_1 w^{\alpha_1-1}\overline{F}^*(w)\,\mathrm{d}w$ is finite (which is what we need) and regularly varying with $v \in \mathcal{RV}_{\alpha_1-\alpha}$. Hence for $\alpha_1 < \alpha$, we have $\mathbb{E}_{F^*}[(\tilde{d}^*)^{\alpha_1}] < \infty$. Finally, we show that $\mathbb{E}_{F^*}[(\tilde{d}^*)^{\alpha}] = \infty$ which implies that any higher moment will also be infinite. Note that for any $q > 0$ we have

$$\begin{aligned}\Pi_{1,\alpha}(q) - \Pi_{1,\alpha}(2q) &= \mathbb{E}_{F^*}[\tilde{d}^* - q]_+ - \mathbb{E}_{F^*}[\tilde{d}^* - 2q]_+,\\ &= \int_q^{2q} \overline{F}^*(w)\,\mathrm{d}w,\\ &\leq q\overline{F}^*(q),\end{aligned}$$

since $\overline{F}^*$ is non-increasing. Hence, for large enough q satisfying both (4.1) and (4.2) (or (4.3) depending on the value of α), we have

$$\begin{aligned}\overline{F}^*(q) &\geq \frac{1}{q}\left[\Pi_{1,\alpha}(q) - \Pi_{1,\alpha}(2q)\right],\\ &\geq \frac{1}{q}\left[\frac{C_1}{q^{\alpha-1}} - \frac{C_1}{(2q)^{\alpha-1}}\left(1 - \frac{C_2}{(2q)^{\alpha-1}}\right)^{-1}\right],\\ &\geq \frac{1}{q}\left[\frac{C_1}{q^{\alpha-1}} - \frac{C_1}{(2q)^{\alpha-1}} \times \left(1 - \frac{1}{2^{\alpha-1}}\right)^{-1}\right] =: \frac{1}{q^\alpha}C_3, \qquad (\text{for } q^{\alpha-1} > C_2),\end{aligned}$$

where $C_3 = C_1(1 - 1/(2^{\alpha-1} - 1))$. Hence we have for q large enough (e.g., $q > C_2^{1/(\alpha-1)}$):

$$\begin{aligned}
\mathbb{E}_{F^*}[(\tilde{d}^*)^\alpha] &= \int_0^q \alpha w^{\alpha-1}\overline{F}^*(w)\,\mathrm{d}w + \int_q^\infty \alpha w^{\alpha-1}\overline{F}^*(w)\,\mathrm{d}w, \\
&\geq \alpha \int_q^\infty w^{\alpha-1}\overline{F}^*(w)\,\mathrm{d}w, \\
&\geq \alpha \int_q^\infty w^{\alpha-1}\frac{C_3}{w^\alpha}\,\mathrm{d}w = \infty.
\end{aligned}$$

Hence for any $\alpha_1 \geq \alpha$, we also have $\mathbb{E}_{F^*}[(\tilde{d}^*)^{\alpha_1}] = \infty$. □

As a consequence of Theorem 4.1, we can relate the optimal order quantity of the distributionally robust newsvendor, the optimal worst-case newsvendor cost defined in (1.2) and the newly characterized distribution F^*, when the critical ratio approaches 1. While a similar characterization has been previously obtainedin modeling the relationship between Value-at-Risk and Conditional Value-at-Risk in risk management for distributions with regularly varying tails (see Proposition 1 in Hua and Joe [24]), the connection to distributionally robust optimization does not seem to have been made, to the best of our knowledge.

Proposition 4.1. *Consider the ambiguity set $\mathcal{F}_{1,\alpha}$ with $\alpha > 1$. For $\eta \in (0,1)$, let q_η^* be an optimal order quantity to the distributionally robust newsvendor in (1.2) and C_η^* be the optimal cost. Then q_η^* is also optimal to a standard newsvendor problem with the underlying demand distribution F^* described in Theorem 4.1(b) and satisfies the property:*

$$q_\eta^* \sim \frac{\alpha-1}{\alpha}\frac{1}{1-\eta}C_\eta^*, \quad \text{as} \quad \eta \to 1. \tag{4.5}$$

Proof. Note that from Theorem 4.1(b) , we have

$$q_\eta^* = \arg\min_{q\geq 0}\sup_{F\in\mathcal{F}_{1,\alpha}}\Big((1-\eta)q + \mathbb{E}_F[\tilde{d}-q]_+\Big) = \arg\min_{q\geq 0}\Big((1-\eta)q + \mathbb{E}_{F^*}[\tilde{d}-q]_+\Big),$$

and hence is optimal for (2.2) with $F \equiv F^*$ and $\tilde{d}^* \sim F^*$. Moreover, we have $1-\eta = \mathbb{P}(\tilde{d}^* > q_\eta^*)$ from the optimality of the newsvendor solution q_η^* for F^*, and,

$$\begin{aligned}
C_\eta^* &= \min_{q\geq 0}\sup_{F\in\mathcal{F}_{1,\alpha}}\Big((1-\eta)q + \mathbb{E}_F[\tilde{d}-q]_+\Big), \\
&= \min_{q\geq 0}\Big((1-\eta)q + \mathbb{E}_{F^*}[\tilde{d}^*-q]_+\Big), \\
&= (1-\eta)q_\eta^* + \mathbb{E}_{F^*}[\tilde{d}^*-q_\eta^*]_+.
\end{aligned}$$

Since $\overline{F}^* \in \mathcal{RV}_{-\alpha}$, a direct application of Karamata's theorem (cf. [24], page 351) yields:

$$\begin{aligned}
\lim_{\eta\uparrow 1} \frac{C^*_\eta}{(1-\eta)q^*_\eta} &= \lim_{\eta\uparrow 1} \frac{(1-\eta)q^*_\eta + \mathbb{E}_{F^*}[\tilde{d}^* - q^*_\eta]_+}{(1-\eta)q^*_\eta}, \\
&= 1 + \lim_{\eta\uparrow 1} \frac{\int_{q^*_\eta}^\infty \mathbb{P}(\tilde{d}^* > x)\,\mathrm{d}x}{q^*_\eta \mathbb{P}(\tilde{d}^* > q^*_\eta)}, \\
&= \frac{\alpha}{\alpha-1}.
\end{aligned}$$

□

5 Numerical Examples

In this section, we provide numerical examples to compare the performance of a classical newsvendor model where the demand is assumed to be known with the distributionally robust newsvendor model. We consider the following three demand distributions that possess different kinds of tail behavior:

(a) Exponential random variable with mean 50;

(b) Lognormal random variable with parameters $m = \log(50/\sqrt{2})$ and $s = \sqrt{\log(2)}$;

(c) Pareto random variable with shape parameter $\beta = 1+\sqrt{2}$ and scale parameter $x_m = 50\sqrt{2}/(1+\sqrt{2})$.

The exponential distribution is light-tailed where all moments of finite order exist, the lognormal distribution is heavy-tailed where all moments of finite order exist, while the Pareto random variable is a heavy-tailed distributon with finite moments only for $\alpha < \beta$. Among these, only the Pareto distribution is regularly varying. The parameter of the demand distributions are selected such that for all three distributions, the mean is 50 and standard deviation is 50. Hence, Scarf's model would prescribe exactly the same optimal order quantity in all three cases. In contrast, since the moments m_α need not be the same when α is not equal to 2, the order quantities from the distributionally robust newsvendor models would change for other values of α.

To compute the robust optimal order quantities, one approach is to directly use the dual SDP formulations discussed in Section 3.4. For example, for $\alpha = 3$, this would reduce to solving:

$$\begin{aligned}
\min_{q,y_0,y_1,y_3,a_1,b_1,c_1,a_2,b_2,c_2} \quad & (1-\eta)q + y_0 + y_1 m_1 + y_3 m_3 \\
\text{s.t.} \quad & \begin{bmatrix} y_0 & 0 & a_1 & b_1 \\ 0 & y_1 - 2a_1 & -b_1 & c_1 \\ a_1 & -b_1 & -2c_1 & 0 \\ b_1 & c_1 & 0 & y_3 \end{bmatrix} \succeq 0, \\
& \begin{bmatrix} y_0 + q & 0 & a_2 & b_2 \\ 0 & y_1 - 1 - 2a_2 & -b_2 & c_2 \\ a_2 & -b_2 & -2c_2 & 0 \\ b_2 & c_2 & 0 & y_3 \end{bmatrix} \succeq 0, \\
& q \geq 0,
\end{aligned} \tag{5.1}$$

However, a standard reformulation comes at the price that for large values of α, the SDP involves several additional variables, besides q, y_0, y_1, y_α. Since the dual constraints are equivalent to nonnegativity constraints of sparse univariate polynomials, we can use the relative entropy reformulation for signomial optimization which preserves sparsity (see Chandrasekaran and Shah [11]). Specifically for any $\alpha > 1$, the problem is given as follows:

$$\begin{aligned} \min \quad & (1-\eta)q + y_0 + y_1 m_1 + y_\alpha m_\alpha \\ \text{s.t.} \quad & y_0 + y_\alpha w^\alpha \geq -y_1 w, \qquad \forall w \geq 0, \\ & y_0 + q + y_\alpha w^\alpha \geq (1-y_1)w, \qquad \forall w \geq 0 \\ & q \geq 0, \end{aligned} \tag{5.2}$$

where the variables y_0 and y_α must be non-negative for feasibility. Using a change of variable with $w = e^z$ where $z \in \Re$ and through an application of Lagrangian duality (see Lemma 1 on page 1150 in [11]), we can rewrite the problem as a relative entropy optimization problem:

$$\begin{aligned} \min_{q,y_0,y_1,y_\alpha,v_1,v_2,v_3,v_4} \quad & (1-\eta)q + y_0 + y_1 m_1 + y_\alpha m_\alpha \\ \text{s.t.} \quad & v_1 \log\left(\frac{v_1}{e y_0}\right) + v_2 \log\left(\frac{v_2}{e y_\alpha}\right) \leq y_1, \\ & v_3 \log\left(\frac{v_3}{e(y_0+q)}\right) + v_4 \log\left(\frac{v_4}{e y_\alpha}\right) \leq y_1 - 1, \\ & (\alpha-1)v_2 = v_1, \\ & (\alpha-1)v_4 = v_3, \\ & q, v_1, v_2, v_3, v_4 \geq 0, \end{aligned} \tag{5.3}$$

which is a convex optimization problem in the variables $q, y_0, y_1, y_\alpha, v_1, v_2, v_3, v_4$. Such a relative entropy formulation can be solved using an off the shelf convex optimization solver such as MOSEK.

5.1 Value of Incorporating Moments Beyond Scarf's Model

We compute the optimal order quantities for the distributionally robust model assuming the highest order moment is given for $\alpha = 4/3, 3/2, 7/4, 2, 3, 5$ and 8 respectively in cases (a) and (b), while for case (c), we consider $\alpha = 4/3, 3/2, 7/4$ and 2 only. Note that for the Pareto random variable in case (c), the moments are finite only for $\alpha < 1 + \sqrt{2} \approx 2.4142$. We estimate the optimal order quantities for critical ratios η in the range $[0.97, 0.99998]$.

In Figures 4, 5 and 6, we provide the log-log plots of the distributionally robust order quantities and the optimal order quantities for the exponential, lognormal and Pareto distributions respectively. From the figures, we observe that as α increases (higher moment information is assumed), the robust order quantity gets closer to the true optimal order quantity for higher order quantities (near the extreme right of the x-axis). For the lognormal demand distribution, in the specified range of critical ratios, we observe that the optimal order quantity for $\alpha = 8$ still exceeds $\alpha = 5$, but as the critical ratio increases further, this result is reversed (see Table 2). Note that Scarf's model would prescribe the same optimal order quantity for all three cases and does not capture the tail behavior. This clearly indicates the value

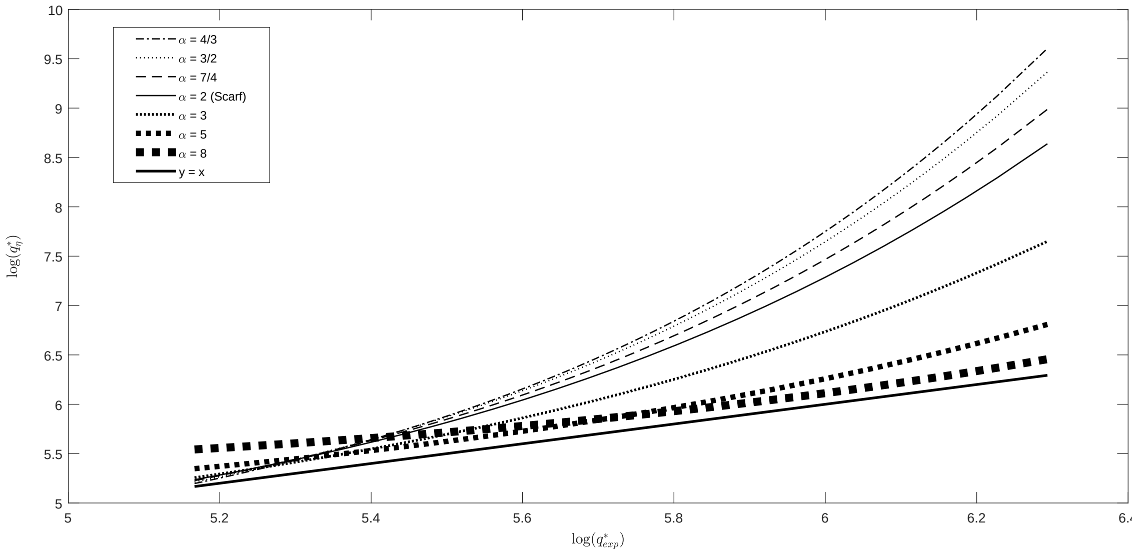

Figure 4: The log-log plot compares the optimal order quantities for the distributionally robust newsvendor with $\alpha = 4/3, 3/2, 7/4, 2, 3, 5, 8$ with the optimal order quantity for the exponential distribution for $\eta \in [0.97, 0.99998]$. As the figure illustrates for larger critical ratios, the knowledge of higher moment information makes the robust model less conservative and closer to the $y = x$ line.

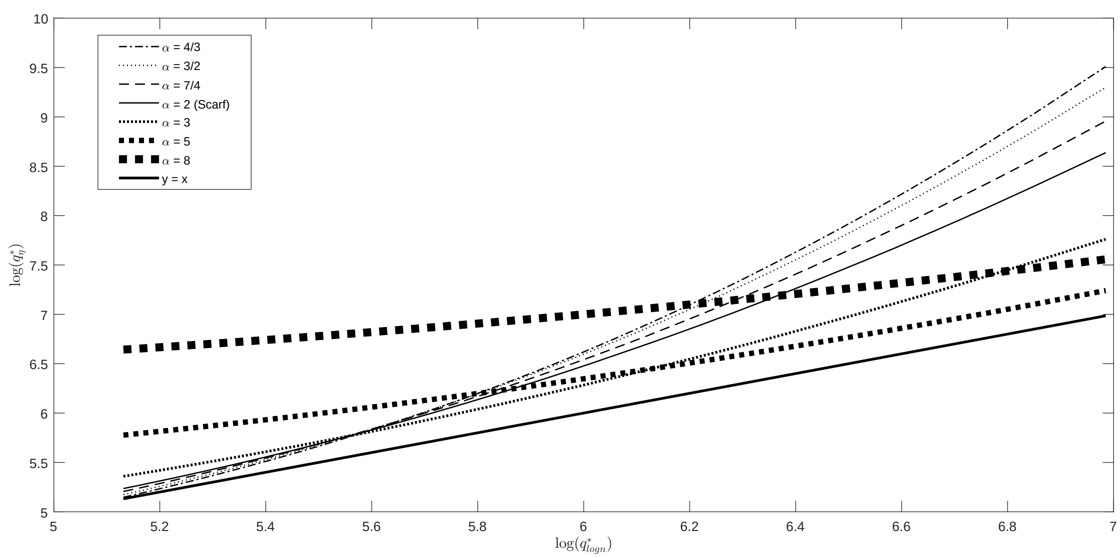

Figure 5: The log-log plot compares the optimal order quantities for the distributionally robust newsvendor with $\alpha = 4/3, 3/2, 7/4, 2, 3, 5, 8$ with the optimal order quantity for the lognormal distribution for $\eta \in [0.97, 0.99998]$. As the figure illustrates for larger critical ratios, the knowledge of higher moment information makes the robust model less conservative and closer to the $y = x$ line. Only for $\alpha = 8$, the line is above the $\alpha = 5$ line for the chosen critical ratios, but the slope indicates that for even higher critical ratios, the robust order quantities for $\alpha = 8$ will get closer to the $y = x$ in comparison to the $\alpha = 5$. This is verified in Table 2.

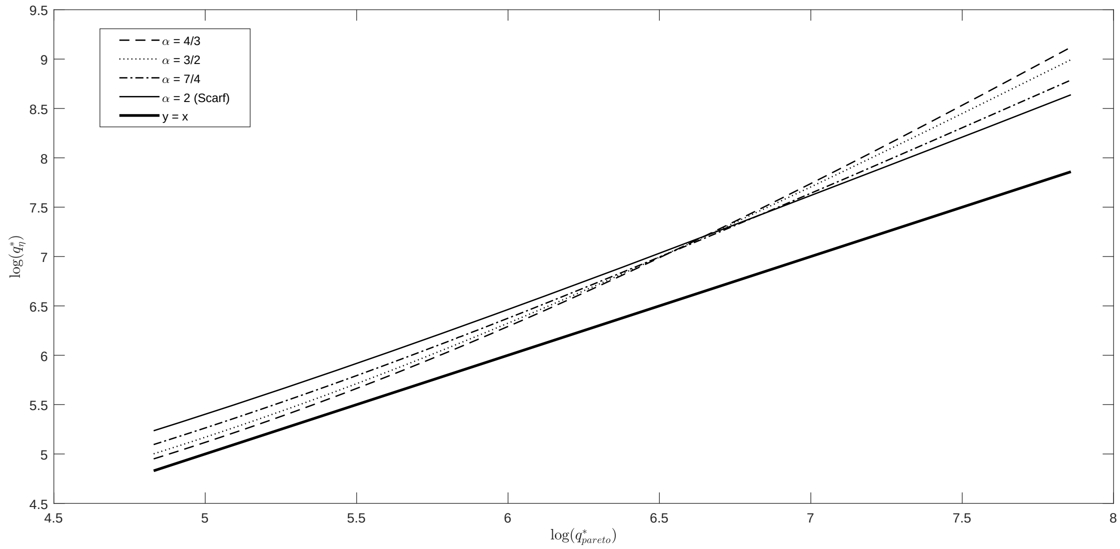

Figure 6: The log-log plot compares the optimal order quantities for the distributionally robust newsvendor with $\alpha = 4/3, 3/2, 7/4, 2$ with the optimal order quantity for the Pareto distribution for $\eta \in [0.97, 0.99998]$. As the figure illustrates for larger critical ratios, the knowledge of higher moment information makes the robust model less conservative and closer to the $y = x$ line.

of having additional moment information in better approximating the tail behavior of the optimal order quantity for a given distribution.

η	q^*_{logn}	q^*_η $(\alpha = 5)$	q^*_η $(\alpha = 8)$
0.99998	1080.46	1389.84	1913.45
0.999998	1643.65	2177.66	2549.77
0.9999998	2405.78	3371.15	3134.68
0.99999998	3418.24	4414.89	4292.02

Table 2: Comparison of optimal order quantities for lognormal with the distributionally robust model for $\alpha = 5$ and $\alpha = 8$.

5.2 Scaling Behavior of Optimal Order Quantities and Optimal Costs

We next validate the scaling behavior of the optimal order quantity and the optimal cost for the distributionally robust model as discussed in Proposition 4.1, illustrating the regularly varying structure and compare it with the corresponding behavior of the optimal solution and the optimal costs for the three distributions in (a)-(c). The scaling constant as the critical ratio approaches 1 for the robust model is provided in Table 3 for the specified values of α. In Figures 7, 8 and 9, we plot these values for the range of critical ratios in $[0.97, 0.99998]$. In the case of the exponential distribution and the lognormal distribution, a simple calculation indicates that these ratios converge to 1 as the critical ratio approaches 1, while the distributionally robust newsvendor model shows a different scaling behavior. In the case of the Pareto distribution, the ratio of $(\beta - 1)/\beta$ is exactly valid for all critical ratios η as shown in the figure.

α	4/3	3/2	7/4	2	3	5	8
$\lim_{\eta\to 1}\dfrac{(1-\eta)q^*_\eta}{C^*_\eta} = \dfrac{\alpha-1}{\alpha}$	1/4	1/3	3/7	1/2	2/3	4/5	7/8

Table 3: Scaling behavior of the optimal order quantity and the optimal cost for the distributionally robust model.

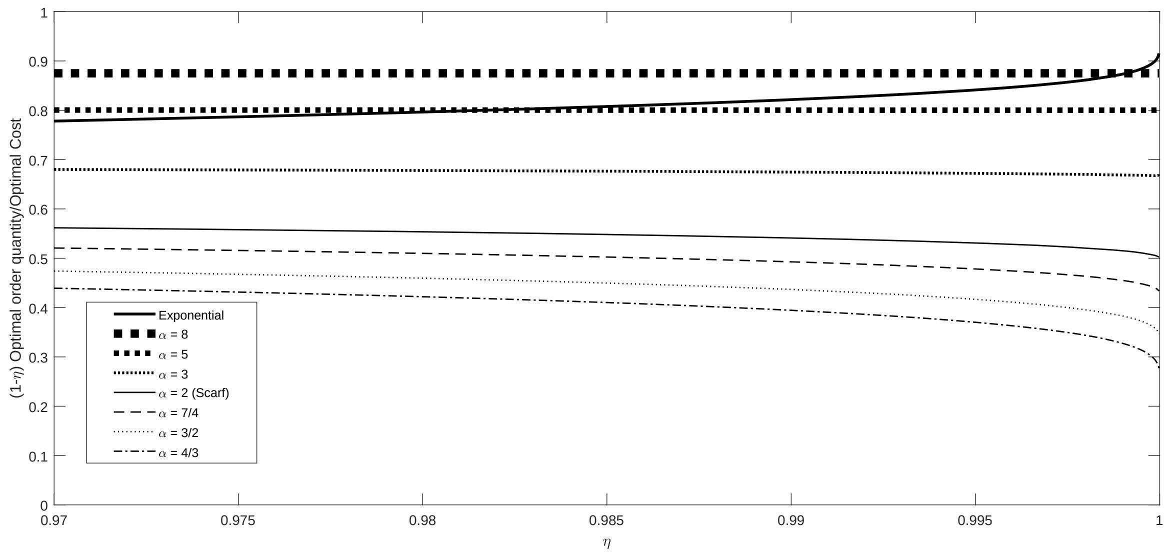

Figure 7: The plot provides the ratios $(1-\eta)q^*_\eta/C^*_\eta$ for the distributionally robust newsvendor model (with different values of α) and the corresponding values for the exponential distribution (which is the dashed line, which tends to 1 as η tends to 1). The figure illustrates the difference in the scaling behavior of the two models with the limit value given by the numbers in Table 3.

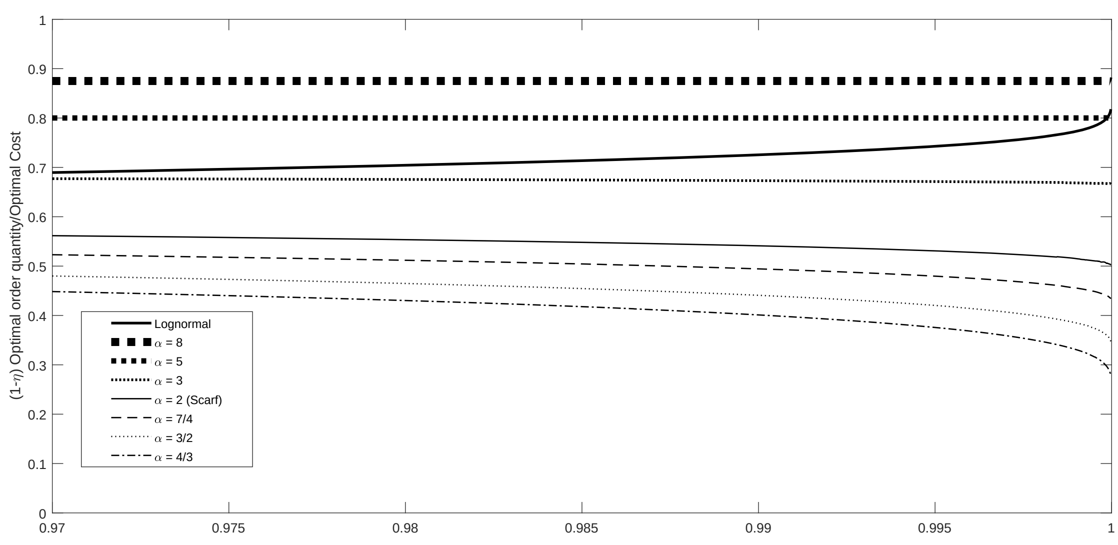

Figure 8: The plot provides the ratios $(1-\eta)q^*_\eta/C^*_\eta$ for the distributionally robust newsvendor model (with different values of α) and the corresponding values for the lognormal distribution (which is the dashed line which tends to 1 as η tends to 1). The figure illustrates the difference in the scaling behavior of the two models with the limit value given by the numbers in Table 3.

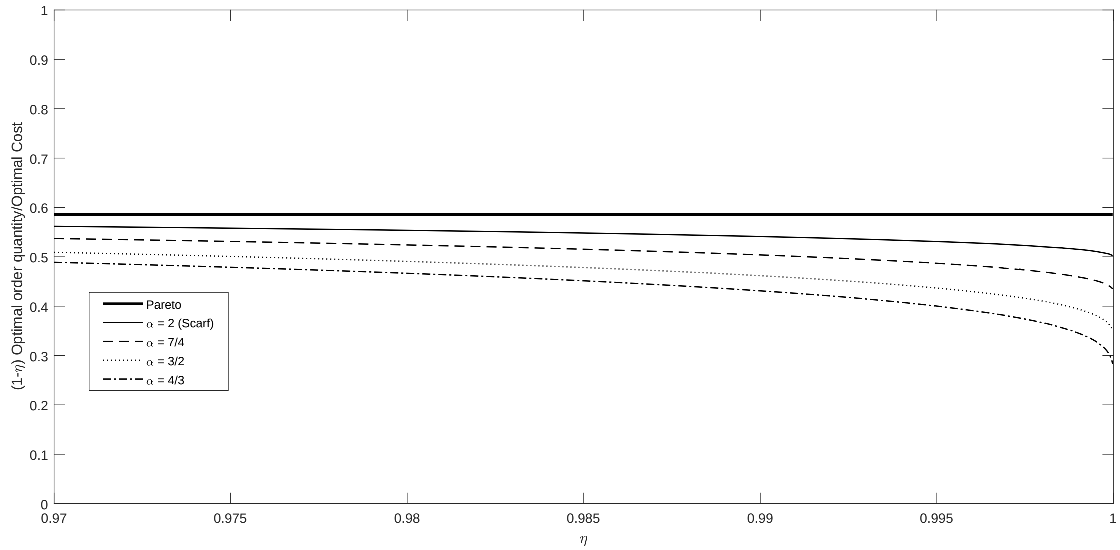

Figure 9: The plot provides the ratios $(1-\eta)q^*_\eta/C^*_\eta$ for the distributionally robust newsvendor model (with different values of α) where the moments are obtained from a Pareto distribution. The Pareto distribution is regularly varying and in this case, the ratio is exactly $\sqrt{2}/(1+\sqrt{2}) \approx 0.5858$.

5.3 Robustness to Contamination

In this section, we compare the performance of the two optimal order quantities - the solution to the classical newsvendor problem and the distributionally robust newsvendor problem. To compare the performance, we use a mixture of two distributions, F_0 and F_1, where F_0 is the original distribution and F_1 is a heavy-tailed (contamination) distribution. The mixture distribution is parameterized by $\lambda \in [0,1]$:

$$F_\lambda \equiv (1-\lambda)F_0 + \lambda F_1, \quad 0 \leq \lambda \leq 1.$$

Such a contamination technique has been proposed in Dupačová [14] to analyze the stability of optimal solutions in stochastic programs when the true distribution is contaminated by another distribution. For the choice of the distribution F_0, we use each of the distributions in (a)-(c). A natural choice for the contamination distribution is $F_1 = F^*$ where F^* is the distribution described in Theorem 4.1(b) and Proposition 4.1 for a chosen α. When $\lambda = 0$, the optimal order quantity is the solution to the newsvendor problem with the corresponding distribution in (a)-(c). When $\lambda = 1$, the optimal order quantity is the solution to the distributionally robust newsvendor for the given α. Denote the corresponding optimal order quantities by q^*_0 and q^*_1 respectively. The order quantities satisfy:

$$C_0(q^*_0) \quad \leq \quad C_0(q^*_1) \quad \text{and} \quad C_1(q^*_0) \quad \geq \quad C_1(q^*_1), \tag{5.4}$$

where C_0 is the newsvendor cost under distribution F_0 and C_1 is the newsvendor cost under distribution F_1. Then, a natural question is the what is the value of λ^*, beyond which the robust order quantity outperforms the classical solution under contamination. In Figures 10, 11 and 12, we plot the λ^* values for each of the exponential, lognormal and Pareto distributions with the contaminating distribution given by the regularly varying distribution for the corresponding α value. The figures illustrate that for high service levels, with even a small amount of contamination, the distributionally robust models will outperform the standard newsvendor solution.

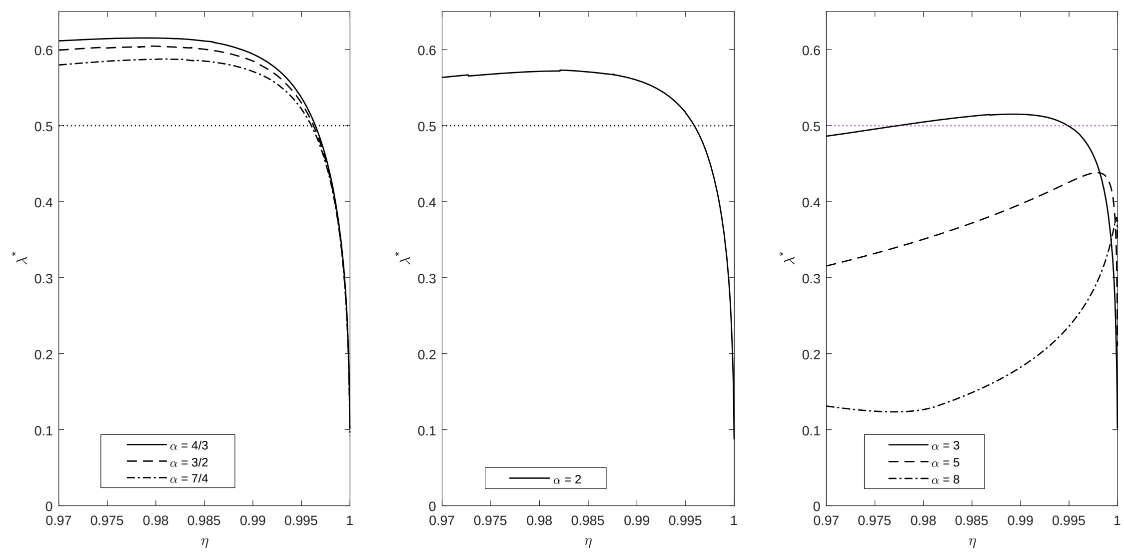

Figure 10: The plot provide the λ^* values for the case where the original distribution is exponential. In each of the cases, we see that as the critical ratio approaches 1, the value of λ^* rapidly drops to 0. This indicates that a small amount of contamination is sufficient for the robust solution to outperform the classical solution for high service levels.

6 Conclusion

The goal of this paper was to characterize properties of the optimal order quantities in a newsvendor model under a robust framework of distributional ambiguity with moment constraints. Building on the observation that the optimal order quantity in Scarf's model is also optimal for a censored student-t distribution with parameter 2, we show that by assuming knowledge of the first and αth moment, the optimal order quantity is also optimal for a regularly varying distribution with tail index α. This provides a characterization of a new distribution, which does not lie in the original ambiguity set, but for which the order quantity from a robust model, continues to remain optimal. We provide numerical evidence to illustrate these results and its applicability.

Several interesting questions still remain to be answered. It would be interesting to see if there exists an analytical characterization of other aspects of the distribution F^* beyond the regularly varying tail property discussed here. Secondly, it would also be of use to characterize the distribution F^* for other types of ambiguity sets that include distributions around a nominal distribution using other probability metrics. We believe this will help better understand as to when solutions from a robust models will do well. Lastly, implications of these results for multidimensional newsvendor problems need to be studied. We leave this for future research.

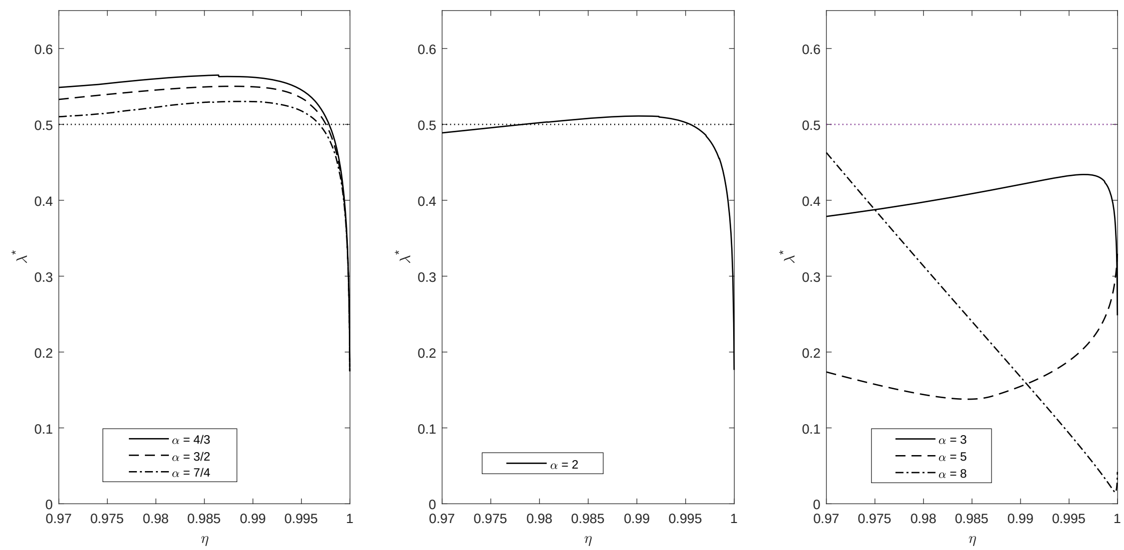

Figure 11: The plot provide the λ^* values for the case where the original distribution is lognormal. In each of the cases, we see that as the critical ratio approaches 1, the value of λ^* rapidly drops to 0. This indicates that a small amount of contamination is sufficient for the robust solution to outperform the classical solution for high service levels.

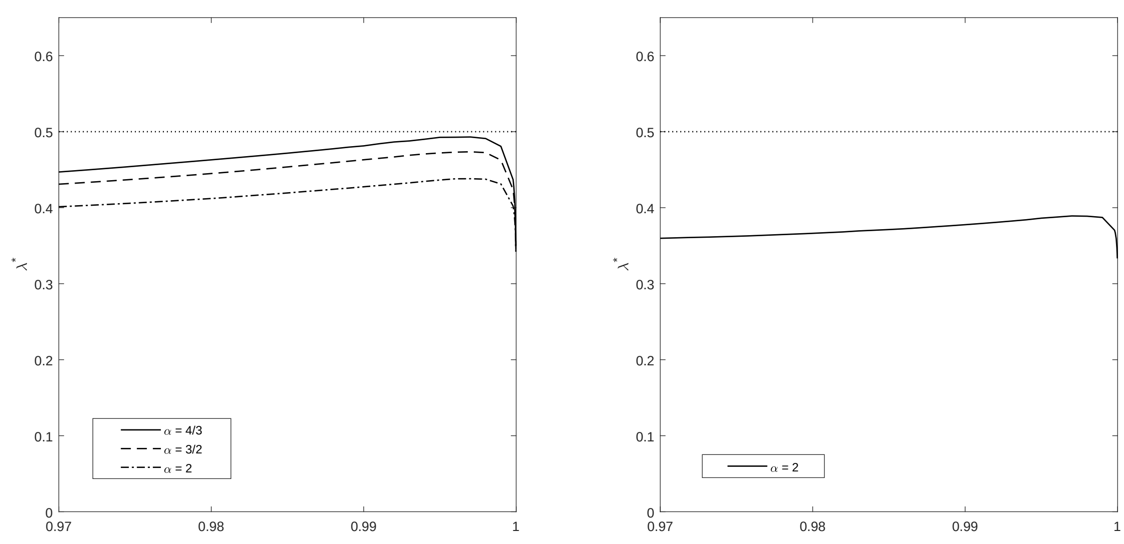

Figure 12: The plot provide the λ^* values for the case where the original distribution is Pareto. In each of the cases, we see that as the critical ratio approaches 1, the value of λ^* rapidly drops to 0. This indicates that a small amount of contamination is sufficient for the robust solution to outperform the classical solution for high service levels.

Acknowledgements

The authors would like to thank the Area Editor Dan Adelman, the Associate Editor and two anonymous reviewers for their very useful inputs and suggestions on improving the paper. The authors would also like to thank Guillermo Gallego (HKUST) for providing useful references on this topic and participants at the Distributionally Robust Optimization Workshop at Banff, 2018 for valuable inputs. The authors are also thankful to He Simai (Shanghai University of Finance and Economics) for sharing some of his ideas on this topic.

Appendix

Implausibility of a closed form solution for general α

A standard approach to try and solve the moment problem analytically is through the dual formulation by reducing the problem to a single variable optimization problem (see Scarf [41]). We illustrate why such an approach does not give a simple closed form solution for the ambiguity set $\mathcal{F}_{1,\alpha}$. Observe that $y_\alpha > 0$ for dual feasibility in (3.4) which implies that the function $g(w) = y_0 + y_1 w + y_\alpha w^\alpha$ is convex for $w \geq 0$. The optimality condition for moment problems implies that the function $g(w)$ touches the piecewise linear convex function $h(w) = [w-q]^+$ at two points, one which is below q and the other which is above q. Let us denote these points by w_1 and w_2. The case where $w_1 = 0$ is the easy case and discussed in Section 3.1. We focus here on the case where $w_1 > 0$. At optimality, the function value of $g(w)$ and its gradient $g'(w)$ matches the function value of $h(w)$ and its gradient $h'(w)$ at both w_1 and w_2. This gives us four equations in five variables y_0, y_1, y_α, w_1 and w_2 as follows:

$$\text{(a) } y_0 + y_1 w_1 + y_\alpha w_1^\alpha = 0 \text{ and (b) } y_0 + y_1 w_2 + y_\alpha w_2^\alpha = w_2 - q,$$

$$\text{(c) } y_1 + \alpha y_\alpha w_1^{\alpha-1} = 0 \text{ and (d) } y_1 + \alpha y_\alpha w_2^{\alpha-1} = 1,$$

Solving for these equations, we can express the variables in terms of y_1 as:

$$\begin{aligned}
y_0 &= q\left(\frac{(-y_1)^{\alpha/(\alpha-1)}}{(1-y_1)^{\alpha/(\alpha-1)} - (-y_1)^{\alpha/(\alpha-1)}}\right), \\
y_\alpha &= \frac{(\alpha-1)^{\alpha-1}}{\alpha^\alpha q^{\alpha-1}}\left((1-y_1)^{\alpha/(\alpha-1)} - (-y_1)^{\alpha/(\alpha-1)}\right)^{\alpha-1}, \\
w_1 &= \frac{\alpha q}{\alpha-1}\left(\frac{(-y_1)^{\alpha/(\alpha-1)}}{(1-y_1)^{\alpha/(\alpha-1)} - (-y_1)^{\alpha/(\alpha-1)}}\right) \\
w_2 &= \frac{\alpha q}{\alpha-1}\left(\frac{(1-y_1)^{\alpha/(\alpha-1)}}{(1-y_1)^{\alpha/(\alpha-1)} - (-y_1)^{\alpha/(\alpha-1)}}\right)
\end{aligned}$$

Plugging the variables back into the dual objective in (3.4) reduces the problem to the minimization problem over a single variable y_1:

$$\min_{y_1} q\left(\frac{(-y_1)^{\alpha/(\alpha-1)}}{(1-y_1)^{\alpha/(\alpha-1)} - (-y_1)^{\alpha/(\alpha-1)}}\right) + m_1 y_1 + \frac{m_\alpha(\alpha-1)^{\alpha-1}}{\alpha^\alpha q^{\alpha-1}}\left((1-y_1)^{\alpha/(\alpha-1)} - (-y_1)^{\alpha/(\alpha-1)}\right)^{\alpha-1},$$

which can be solved by setting the derivative to be zero. Unfortunately for general rational values of α, this problem appears to be difficult to solve in closed form using simple functions such as radicals (involving only taking roots and the four basic arithmetic operations). Using a symbolic computation package such as Mathematica fails to provide closed form solutions for this problem for example for $\alpha = 5$. However in Scarf's model with $\alpha = 2$, this reduces to:

$$\min_{y_1} \frac{qy_1^2}{(1-2y_1)} + m_1 y_1 + \frac{m_2(1-2y_1)}{4q},$$

which is solvable in closed form by setting the derivative to be zero. This gives:

$$y_1 = \frac{1}{2}\left(1 - \frac{q}{\sqrt{q^2 - 2m_1 q + m_2}}\right)$$

and the corresponding first term in the closed form bound in (1.3). Observe that Scarf's solution involves only radicals.

Regularly Varying Distributions

Two examples of regularly varying random variables that are particularly relevant in our context are:

(a) Pareto(x_m, α): This random variable is defined with two parameters - a scale parameter $x_m > 0$ and a shape parameter $\alpha > 0$ with probability density function given as follows:

$$g(w) = \frac{\alpha x_m^\alpha}{w^{\alpha+1}}, \ \forall w \geq x_m,$$

Then for $w \geq x_m$, we have:

$$\overline{F}(w) := \int_w^\infty g(x)\mathrm{d}x = x_m^\alpha w^{-\alpha},$$

and hence clearly $\overline{F} \in \mathcal{RV}_{-\alpha}$. Note that in Grundy's model discussed in Section 3.2, we obtain a characterization of the distribution F^* as follows:

$$F^*(w) = \mathbb{P}(\tilde{d}^* \leq w) = \begin{cases} 1 - m_\alpha\left(\frac{\alpha-1}{\alpha w}\right)^\alpha, & \text{if } w > \frac{\alpha-1}{\alpha} m_\alpha^{1/\alpha}, \\ 0 & \text{if } 0 \leq w \leq \frac{\alpha-1}{\alpha} m_\alpha^{1/\alpha}. \end{cases} \tag{6.1}$$

This defines a Pareto random variable with $\overline{F}^* \in \mathcal{RV}_{-\alpha}$:

$$\tilde{d}^* = \text{Pareto}\left(\frac{(\alpha-1)m_\alpha^{1/\alpha}}{\alpha}, \alpha\right). \tag{6.2}$$

(b) $\tilde{t}_\nu(\mu, \sigma^2)$: The t-location scale random variable is defined with three parameters - a location parameter $\mu > 0$ and a scale parameter $\sigma > 0$ and degree of freedom parameter ν. This distribution is regularly varying at infinity with index ν. Note that in Scarf's model, we have $\overline{F}^* \in \mathcal{RV}_{-2}$.

Regularly varying functions have a rich theory (see Bingham, Goldie and Teugels [7], de Haan [15]) and has found many applications in the study of power law distributions and extreme risk behavior in insurance, finance, telecommunication, social networks (see Embrechts, Klüppelberg, Mikosch [16], Resnick [33] for details). The class of regularly varying functions admits certain nice properties with respect to summation, composition, taking quotients, integrating and differentiating which helps in understanding tail behavior of regularly varying random variables, their moments and other functionals. The following result below attributed to Karamata [28] shows the effect of integration on regularly varying functions. We state a special case relating to regularly varying distributions with at least first moment finite (see Resnick [33, Theorem 0.6(a)]). In the following theorems, one can think of U as the distribution tail and u as the density in the context of distribution functions.

Theorem 6.1 (Karamata's Theorem)**.** *Suppose $u : \Re_+ \to \Re_+$ satisfies $u \in \mathcal{RV}_{-\alpha}$ for some $\alpha > 1$. Then, (i) $U(x) = \int_x^\infty u(t)\,\mathrm{d}t$ is finite, (ii) $U \in \mathcal{RV}_{-\alpha+1}$, and (iii) the following limit holds:*

$$\lim_{x\to\infty} \frac{xu(x)}{U(x)} = \alpha - 1.$$

The next result provides the reverse implication to Karamata's theorem and characterizes what happens when a regularly varying function is differentiated; see Landau [29], Bingham, Goldie and Teugels [7, Theorem 1.6.1], de Haan [15, p. 23], Resnick [33, Theorem 0.7] for different formulations and proofs.

Theorem 6.2. *Suppose $u : \Re_+ \to \Re_+$ is locally integrable in $[0, \infty)$ and define:*

$$U(x) := \int_x^\infty u(t)\,\mathrm{d}t.$$

(a) If for $\alpha > 0$, we have functions u and U satisfying:

$$\lim_{x\to\infty} \frac{xu(x)}{U(x)} = -\alpha,$$

then $U \in \mathcal{RV}_{-\alpha}$.

(b) If $U \in \mathcal{RV}_{-\alpha}$ for $\alpha > 0$ and u is monotone, then the limit in (a) holds and $u \in \mathcal{RV}_{-\alpha-1}$.

References

[1] E. T. Anderson, G. J. Fitzsimons, and D. Simester, *Measuring and mitigating the costs of stockouts*, Management Science, 52, 11, 1751-1763, 2006.

[2] D. Bertsimas, V. Gupta and N. Kallus, *Robust sample average approximation*, Mathematical Programming Series A, 171, 217-282, 2018.

[3] D. Bertsimas and I. Popescu, *Optimal inequalities in probability theory: A convex optimization approach*, SIAM Journal on Optimization, 15, 3, 780-804, 2005.

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

[5] A. Ben-Tal and E. Hochman, *More bounds on the expectation of a convex function of a random variable*, Journal of Applied Probability, 9, 803-812, 1972.

[6] K. Bimpkis and M. G. Markakis, *Inventory pooling under heavy-tailed demands*, Management Science, 62, 6, 1800-1813, 2016.

[7] N. H. Bingham, C. M. Goldie and J. L. Teugels, *Regular variation*, Cambridge University Press, 1989.

[8] J. Blanchet, F. He and K. Murthy, *On distributionally robust extreme value analysis*, Extremes, 23, 317–347, 2020.

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

[10] A. Clauset, C. R. Shalizi and M. E. J. Newman, *Power law distributions in empirical data*, SIAM Review, 51, 4, 661-703, 2009.

[11] V. Chandrasekaran and P. Shah, *Relative entropy relaxations for signomial optimization*, SIAM Journal on Optimization 26, 2, 1147–1173, 2016.

[12] L. Chen, S. He and S. Zhang, *Tight bounds for some risk measures, with applications to robust portfolio selection*, Operations Research, 59, 4, 847-865, 2011.

[13] J. Chevalier and A. Goolsbee, *Measuring prices and price competition online: Amazon.com and Barnesand-Noble.com*, Quantitative Marketing and Economics, 1, 203-222, 2003.

[14] J. Dupačová, *Stability in stochastic programming with recourse: Contaminated distributions*, Mathematical Programming Study, 27, 133-144, 1986.

[15] L. de Haan, *On regular variation and its application to the weak convergence of sample extremes*, Mathematisch Centrum Amsterdam, 1970.

[16] P. Embrechts, C. Klüppelberg and T. Mikosch, *Modelling extreme events for insurance and finance*, Springer-Verlag, Berlin, 1997.

[17] P. M. Esfahani and D. Kuhn, *Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations*, Mathematical Programming, 171, 1-2, 115-166, 2018.

[18] I. Fedotenkov, *A bootstrap method to test for the existence of finite moments*, Journal of Nonparametric Statistics, 25, 2, 315-322, 2013.

[19] E. Gaffeo, A. E. Scorcu and L. Vici, *Demand distribution dynamics in creative industries: The market for books in Italy*, Information Economics and Policy, 20, 257-268, 2008.

[20] G. Gallego, *New bounds and heuristics for (Q,r) policies*, Management Science, 44, 2, 219–233, 1998.

[21] G. Gallego and I. Moon, *The distribution free newsboy problem: review and extensions*, The Journal of the Operational Research Society, 44, 8, 825-834, 1993.

[22] R. Gao and A. J. Kleywegt, *Distributionally robust stochastic optimization with Wasserstein distance*, Working Paper, 2017

[23] B. D. Grundy, *Option prices and the underlying asset's return distribution*, The Journal of Finance, 46, 3, 1045-1069, 1991.

[24] L. Hua and H. Joe, *Second order regular variation and conditional tail expectations of multiple risks*, Insurance: Mathematics and Economics, 49, 537-546, 2011.

[25] K. Jansen, J. Haezendonck, J. and M. J. Goovaerts, *Upper bounds on stop-loss premiums in case of known moments up to the fourth order*, Insurance: Mathematics and Economics 5, 315–334, 1986.

[26] H. Lam and C. Mottet, *Tail analysis without parametric models: A worst-case perspective*, Operations Research, 2017.

[27] J. B. Lasserre, *Bounds on measures satisfying moment conditions*, The Annals of Applied Probability, 12, 3, 1114-1137, 2002.

[28] J. Karamata, *Sur un mode de croissance régulière des fonctions*, Mathematica (Cluj), 4, 38 -53, 1930.

[29] E. Landau, *Darstellung und Begrundung Einiger Neueren Ergebnisse der Funktionentheorie*, Springer-Verlag, Berlin, 1916.

[30] A. Müller, D. Stoyan, *Comparison methods for stochastic models and risks*. Wiley, Chichester 2002.

[31] K. Natarajan, M. Sim and J. Uichanco, *Asymmetry and ambiguity in newsvendor models*, Management Science, 64, 7, 3146-3167, 2018.

[32] Y. Nesterov, *Squared functional systems and optimization problems*, in H. Frenk, K. Roos, T. Terlaky and S. Zhang (eds.), High Performance Optimization, Kluwer-Dordrecht, 405-440, 2000.

[33] S. I. Resnick, *Extreme values, regular variation and point processes*, Reprint of the 1987 original, Springer, 2008

[34] H. Scarf, *A min-max solution of an inventory problem* in Studies in The Mathematical Theory of Inventory and Production. (K. Arrow, S. Karlin and H. Scarf, Eds.) 201-209, Stanford University Press, California, 1958.

[35] M. Shaked and G. Shanthikumar, *Stochastic orders*, Springer Series in Statistics, New York, 2007.

[36] A. Shapiro and S. Ahmed, *On a class of minimax stochastic programs*, SIAM Journal on Optimization, 14, 4, 1237-1249, 2004.

[37] A. Shapiro and A. Kleywegt, *Minimax analysis of stochastic problems*, Optimization Methods and Software, 17, 523-542, 2002.

[38] K. C. Toh, M. J. Todd, and R. H. Tutuncu, *SDPT3 — a matlab software package for 750 semidefinite programming*, Optimization Methods and Software, 11, 545–581, 1999.

[39] K. C. Toh, M. J. Todd, and R. H. Tutuncu, *Solving semidefinite-quadratic-linear programs using SDPT3*, Mathematical Programming, 95, 189–217, 2003.

[40] Z. Wang, P. W. Glynn and Y. Ye, *Likelihood robust optimization for data-driven problems*, Computational Management Science, 13, 241-261, 2015.

[41] L. F. Zuluaga, J. F. Pena and D. Du, *Third-order extensions of Lo's semiparametric bound for European call options*, European Journal of Operational Research, 198, 2, 557-570, 2009.