

On the heavy-tail behavior of the distributionally robust newsvendor

Bikramjit Das* Anulekha Dhara† Karthik Natarajan‡§

June 2018

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 optimal order quantity is computed by accounting for the worst possible distribution from a set of demand distributions that is characterized by partial information, such as moments. The model is criticized at times for being overly conservative since the worst-case distribution is discrete with a few support points. However, it is the order quantity from the model that is typically of practical relevance. A simple observation shows that the optimal order quantity in Scarf’s model with known first and second moment is also optimal for a heavy-tailed censored student-t distribution with degrees of freedom 2. In this paper, we generalize this “heavy-tail optimality” property of the distributionally robust newsvendor to a more general ambiguity set where information on the first and the n th moment is known, for any real number $n > 1$. We provide a characterization of the optimal order quantity under this ambiguity set by showing that for high critical ratios, the order quantity is optimal for a regularly varying distribution with an approximate power law tail with tail index n . We illustrate the applicability of the model by calibrating the ambiguity set from data and comparing the performance of the order quantities computed via various methods in a dataset.

1 Introduction

Since the pioneering work of Scarf [39], 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

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

†Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan 48109. Email: ad-hara@umich.edu

‡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 and the 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?”.

item to order before the actual demand is observed. The unit purchase cost is c where $c > 0$ and the unit revenue is p where $p > c$. 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. Furthermore, the probability distribution of the demand, denoted by $F(d) := \mathbb{P}(\tilde{d} \leq d)$ is ambiguous and only assumed to lie in a set of possible distributions, denoted by \mathcal{F} . The ambiguity in the demand distribution might arise due to one of several reasons. It might arise as a subjective input when a new product is introduced into the market for which past demand data is unavailable and one is unsure about an exact distribution or it might arise when a set of plausible demand distributions is constructed from past data using moments, structural information or probability distance metrics. All relevant information on the demand distribution that the newsvendor possesses is assumed to be captured in the set \mathcal{F} . The distributionally robust newsvendor then orders the quantity that maximizes the minimum (worst-case) expected profit. Mathematically, this problem is formulated as choosing an order quantity q to maximize the minimum expected profit as follows:

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

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:

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

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

1.1 Scarf's Model

The earliest version of the model is attributed to Scarf [39] who assumed that the mean and the variance of the demand are specified in the set \mathcal{F} , but the exact form of the distribution is unknown. The set of demand distributions is defined as:

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

where $\mathbb{M}(\mathbb{R}_+)$ 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 moments satisfying $m_2 \geq m_1^2 > 0$. Note that when $m_2 = m_1^2$, the demand is deterministic with support at m_1 . In the standard newsvendor model, when the set of distributions is a singleton with a cumulative distribution function F , the optimal order quantity reduces to the well-known critical fractile formula $q^* = F^{-1}(\alpha)$, where $F^{-1}(\cdot)$ is the generalized inverse of the cumulative distribution function. However, in the robust model, the worst-case demand distribution might change with the order quantity. Scarf [39] explicitly characterized the unique two point distribution that attains the worst-case in (1.2) for the set $\mathcal{F}_{1,2}$. Given an order quantity $q > m_2/2m_1$, the worst-case demand

distribution for $\sup_{F \in \mathcal{F}_{1,2}} \mathbb{E}_F[\tilde{d} - q]_+$ is given by the distribution with two support points:

$$\tilde{d}_q^* = \begin{cases} q - \sqrt{q^2 - 2m_1q + m_2}, & \text{w.p. } \frac{1}{2} \left(1 + \frac{q - m_1}{\sqrt{q^2 - 2m_1q + m_2}} \right), \\ q + \sqrt{q^2 - 2m_1q + m_2}, & \text{w.p. } \frac{1}{2} \left(1 - \frac{q - m_1}{\sqrt{q^2 - 2m_1q + m_2}} \right), \end{cases}$$

where the support points and the probabilities are dependent on q (the ‘‘power’’ of the adversary). In the case, 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:

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

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

$$\sup_{F \in \mathcal{F}_{1,2}} \mathbb{E}_F[\tilde{d} - q]_+ = \begin{cases} \frac{1}{2} \left(\sqrt{q^2 - 2m_1q + m_2} - (q - m_1) \right), & \text{if } q > \frac{m_2}{2m_1}, \\ m_1 - \frac{qm_1^2}{m_2}, & \text{if } 0 \leq q \leq \frac{m_2}{2m_1}. \end{cases} \quad (1.3)$$

Plugging in the expression (1.3) into (1.2) and a direct application of calculus provides a closed-form solution for the optimal order quantity as follows:

$$q^* = \begin{cases} m_1 + \frac{\sqrt{m_2 - m_1^2}}{2} \frac{2\alpha - 1}{\sqrt{\alpha(1 - \alpha)}}, & \text{if } \frac{m_2 - m_1^2}{m_2} \leq \alpha < 1, \\ 0, & \text{if } 0 \leq \alpha < \frac{m_2 - m_1^2}{m_2}. \end{cases}$$

While the optimal order quantity is derived in a simple closed-form manner, this model is criticized at times for being too conservative¹. However in some cases, it is known to provide a good approximation. Scarf [39] in his original treatise had observed that for a large range of critical ratios (specifically α in the range $[0.05, 0.95]$), the optimal order quantity for the two moment model is very close to optimal order quantity for a normal approximation of a Poisson distribution, while for higher critical ratios, the model prescribed higher order quantities. Gallego and Moon [22] in a follow-up set of experiments compared the order quantity from Scarf’s model with the optimal order quantity for normally distributed demands and concluded that for a large range of critical ratios, the loss is profit is not significant. On the other hand, Wang, Glynn and Ye [47] found that when the true distribution is exponential, the results from Scarf’s model might be fairly different for particular choices of parameter values. For example, with

¹We cite for example from page 243 in Wang, Glynn and Ye [47]: ‘‘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.’’

$m_1 = 50$, $m_2 = 5000$ and $\alpha = 0.5$, Scarf's model prescribes an optimal order quantity of 50 while the optimal order quantity for the exponential distribution is approximately 34.65. In Figure 1, we provide a comparison of the optimal order quantities for a fixed demand distribution and Scarf's model where only the first 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 significantly higher order quantities for high critical ratios. In this paper, we provide a precise analytical characterization of this numerical insight.

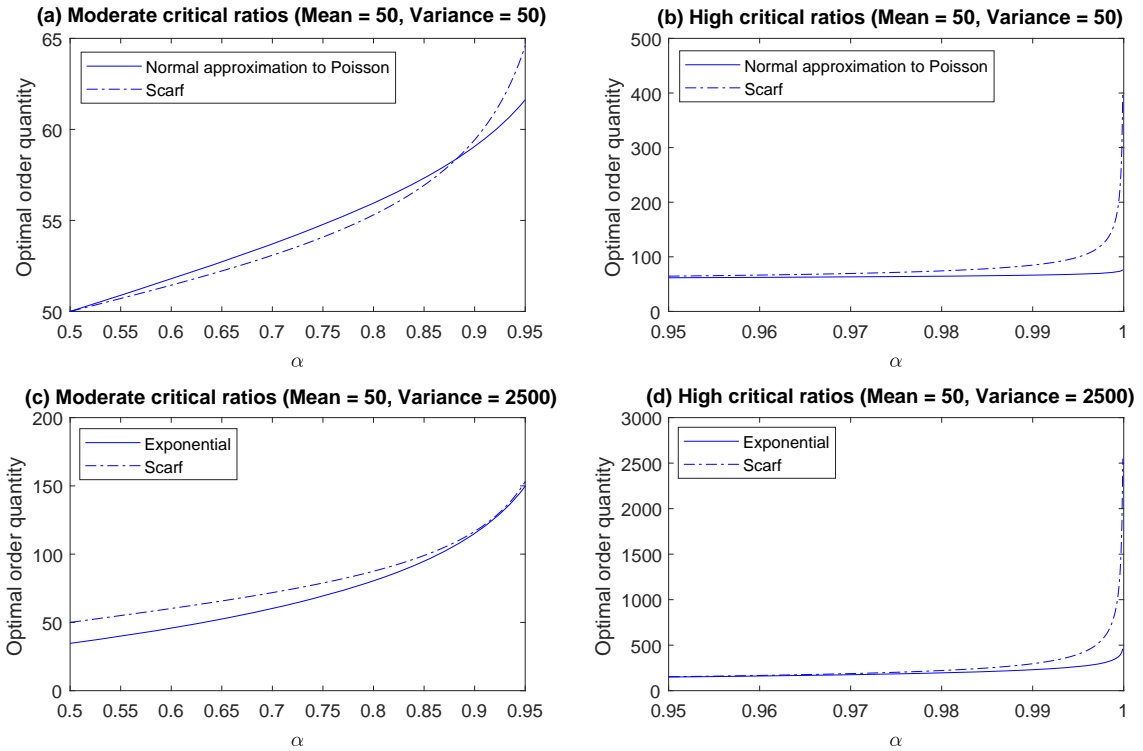


Figure 1: The plots at the top compare the optimal order quantities for a normal approximation to a Poisson demand distribution and Scarf's model with mean demand 50 and variance 50. The plots at the bottom compare the optimal order quantities for an exponential demand distribution and Scarf's model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

The format of the paper and the main contributions are discussed next:

- (a) In Section 2, we provide an overview of the distributionally robust newsvendor problem while analytically characterizing a heavy-tail optimality property that the Scarf's newsvendor model possesses. While this observation has been made in prior research, the result does not seem to be very well-known and has not been analyzed for generalizations of Scarf's model, to the best of our knowledge. We also discuss empirical research that suggests that heavy-tailed distributions might occur in real demand datasets.
- (b) In Section 3, we propose a generalization of the ambiguity set from the first and the second moment to the first and the n th moment. The ambiguity set is simple since it is not overly parameterized

while providing flexibility in allowing for distributions without a finite variance. However, unlike Scarf’s model, there is a technical challenge since the worst-case bound does not appear to be solvable in closed form. Towards this, we derive new upper and lower bounds on the worst-case expected value that is asymptotically optimal for large values of order quantities, by creating appropriate primal and dual feasible solutions to the moment problem.

- (c) In Section 4, we provide a characterization of the optimal order quantity from the distributionally robust newsvendor model by showing that it is optimal for a regularly varying distribution with tail parameter n , using techniques developed to model heavy-tailed distributions. 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. Particularly, it shows that while the worst-case distribution in a distributionally robust newsvendor problem might be discrete with a few support points, the order quantities remain optimal for high critical ratios, for a regularly varying continuous distribution with an infinite n th moment.
- (d) In Section 5, we discuss techniques to construct the ambiguity set using data and illustrate how one might construct estimates for n and the moments. We compare the expected profit of the optimal order quantity from the robust model with alternate approaches such as using a normal approximation, empirical quantiles and Scarf’s method on new car sales data and illustrate the value of solving the distributionally robust newsvendor problem. We finally 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 taken on any value in $[0, \infty)$. Our interest in such ambiguity sets stems from an attempt to provide a characterization of the tail behavior which is of particular interest in solving newsvendor problems with high service levels. There is growing evidence in the literature that a stockout for retailer has significant short-term and long-term effects that needs to be minimized (see Anderson, Fitzsimons and Simester [1]). The high service level regime is the natural domain of interest in this case. We also review in this section, prior empirical research that provides evidence on the existence of heavy-tailed demand distributions.

2.1 Distributionally Robust Newsvendor Models

Shapiro and Kleywegt [43] and Shapiro and Ahmed [42] developed a reformulation of the distributionally robust newsvendor as a classical newsvendor problem through the construction of a new probability demand distribution. The key insight to their reformulation is the observation (see Theorem 2.1 and Section 3.1 on page 532 in Shapiro and Kleywegt [43]) that given a set \mathcal{F} , there exists a non-negative random variable \tilde{d}^* with probability distribution F^* such that the following equality holds for all values

of q :

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

The random variable \tilde{d}^* with the distribution F^* on the right hand side of equation (2.1) corresponds to a random variable that dominates all the random variables \tilde{d} in the set \mathcal{F} in an increasing convex order sense (see Müller and Stoyan [33], Shaked and Shanthikumar [41]). Unlike the extremal distribution on the left hand side of the equation which might vary with q , the random variable \tilde{d}^* on the right hand side has a distribution F^* which is independent of q . This equivalence helps convert the distributionally robust newsvendor problem to the classical newsvendor problem as follows:

$$\min_{q \in \mathbb{R}_+} \left(\mathbb{E}_{F^*}[\tilde{d}^* - q]^+ + (1 - \alpha)q \right), \quad (2.2)$$

where F^* is independent of the critical ratio α . However, the challenge in applying this technique to solve the distributionally robust newsvendor problem is that F^* in most cases does not have an explicit characterization in terms of the original set of distributions \mathcal{F} and might not even lie in this set. However, the equivalence provides an important insight as it identifies a new distribution F^* for which the optimal order quantity from the distributionally robust newsvendor model in (1.2) is optimal, regardless of the parameters of the problem.

In Scarf's model, it is straightforward to construct the distribution F^* as \tilde{d}_q^* is known for each value of q . Taking the derivative of the worst-case bound in (1.3), we obtain the characterization of the distribution F^* as follows:

$$F^*(d) = \mathbb{P}(\tilde{d}^* \leq d) = \begin{cases} \frac{1}{2} \left(1 + \frac{d - m_1}{\sqrt{d^2 - 2m_1d + m_2}} \right), & \text{if } d > \frac{m_2}{2m_1}, \\ 1 - \frac{m_1^2}{m_2}, & \text{if } 0 \leq d \leq \frac{m_2}{2m_1}. \end{cases} \quad (2.3)$$

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

$$\tilde{d}^* = \begin{cases} \tilde{t}_2(m_1, (m_2 - m_1^2)/2), & \text{if } \tilde{t}_2(m_1, (m_2 - m_1^2)/2) > \frac{m_2}{2m_1}, \\ 0, & \text{otherwise,} \end{cases} \quad (2.4)$$

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

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

The distribution of the censored student-t random variable in (2.4) is defined by a discrete probability

mass function:

$$\mathbb{P}(\tilde{d}^* = 0) = \frac{m_2 - m_1^2}{m_2},$$

and a continuous probability density function given by:

$$f^*(d) = \frac{1}{2} \frac{m_2 - m_1^2}{(d^2 - 2m_1d + m_2)^{3/2}}, \quad \forall d > \frac{m_2}{2m_1}.$$

A straightforward calculation indicates that for this random variable, the second moment is infinite, that is:

$$\mathbb{E}_{F^*}[\tilde{d}^{*2}] = \infty.$$

This implies that to recreate the optimal order quantity of the distributionally robust newsvendor model under the assumption of a known mean and variance, we need to solve a standard newsvendor problem with a censored student-t distribution with degrees of freedom 2. Thus the demand distribution in the standard newsvendor model has to possess infinite variance which is an heavy-tail property to recreate the distributionally robust newsvendor solution with a finite variance. To the best of our knowledge, this observation has been made for real-valued random variables in Theorem 1.10.7 on page 57 in Müller and Stoyan [33] with its application to the newsvendor model discussed in Gallego [21]. The representation in (2.4) provides the generalization to non-negative demand random variables. Our interest in this paper is to validate if this insight generalizes to other moment information.

Since the pioneering work of Scarf [39], there have been several generalizations of the distributionally robust newsvendor model to new ambiguity sets. While for some of these ambiguity sets (see Ben-Tal and Hochman [6, 7] and Natarajan, Uichanco and Sim [34]), the problem is solvable in a closed-form-manner, in most cases, numerical optimization techniques are need to solve the problem. Bertsimas and Popescu [3, 4] and Lasserre [28] 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 degree $n \in \mathbb{Z}_+$:

$$\mathcal{F}_{1,2,\dots,n} = \left\{ F \in \mathbb{M}(\mathcal{R}_+) : \int_0^\infty dF(w) = 1, \int_0^\infty w^i dF(w) = m_i, \quad i = 1, 2, \dots, n \right\}.$$

An application of semidefinite programming duality implies that the distributionally robust newsvendor problem is solvable as a semidefinite program. 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 [26], Zuluaga, Peña and Du [48]). Popescu [36] generalized these bounds by incorporating additional structural properties such as symmetry and unimodality to the ambiguity sets. Semidefinite optimization and second order conic optimization methods have been developed to find the worst-case bounds for such problems under structural information (see Popescu [36], Van Parys, Goulart and Kuhn [46], Li, Jiang and Mathieu [30]). 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 [27] recently proposed an ambiguity set, where information on the tail probability of the random variable for a given threshold, the density function at that threshold and the left derivative

of the density function at the threshold is known with an additional assumption that the tail density function is convex. Under this ambiguity set, they showed that the worst-case distribution is either extremely light tailed or heavy-tailed and proposed the use of low dimensional nonlinear optimization methods to compute the bound. In contrast to their approach which models the tail behavior in the ambiguity set, we focus on ambiguity sets with moment information and characterize the tail behavior implied by the model. Ben-Tal et al. [5] 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. Blanchet and Murthy [10] build on this model to show that under the assumption that the reference measure is a distribution such as exponential, Weibull or Pareto distribution, the use of an ambiguity set with a Kullback-Leibler distance measure contains distributions where the tail probabilities decay at a very slow rate for which the worst-case expected costs might be infinite. To overcome this pessimism, they proposed alternative ambiguity sets using the Renyi divergence measure for which the worst-case tails are heavier than the reference measure, but not as heavy as the Kullback-Leibler divergence measure. In contrast, we focus on the moment ambiguity set in this paper. A related recent stream of literature has focussed on solving distributionally robust optimization problems including the newsvendor model using ambiguity sets around a reference measure defined with the Wasserstein metric (see Esfahani and Kuhn [19] and Gao and Kleywegt [23]). These models are solved using convex optimization methods where the worst-case distributions are typically less sensitive to the specification of the support of the demand distribution. To the best of our knowledge, a precise characterization of F^* for such ambiguity sets is currently not known. We next review empirical research that provides evidence on the existence of heavy-tailed distributions in demand data.

2.2 Empirical Evidence of Heavy Tailed Demand

There has been growing evidence in the recent years that heavy-tailed demand distributions can occur in practice and has to be better accounted for in operational settings. Clauset, Shalizi and Newman [12] in their well-cited study on the presence of power law distributions in real world datasets, developed a set of statistical tests to help validate if the data follows a power law. They studied twenty four different datasets across a broad range of disciplines in physics, earth sciences, biology and engineering where prior research in these domains had conjectured that the data followed a power law. Of these in seventeen of the datasets, the statistical tests provided evidence that the power law hypothesis was a reasonable one and could not be firmly ruled out while in the remaining seven datasets, the p-values were too small and with reasonable confidence, the power law could be ruled out. Among these datasets, two of them which are particularly relevant to demand models are: 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 both these datasets, the authors found strong evidence that the power law tail is a reasonable model in comparison to the exponential and stretched exponential distributions but at the same time it was not

possible to rule out other heavy-tail distributions such as the lognormal distribution as a possible fit. In another study, Gaffeo, Scorcu and Vici [20] 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 lesser than 2 is a good fit to the right tail of the demand distribution. Bimpkis and Markakis [8] 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. Using data from a North American retailer over a one year period with 626 products, their statistical tests showed that the exponential and normal distributions were a poor fit to the data while the power law provided a reasonable approximation to the dataset. Building on this observation, they showed that for a class of heavy-tailed stable demand distributions, the benefits from pooling in inventory can be much lower than that predicted for normally distributed demands. Natarajan, Sim and Uichanco [34] used data from an European automotive manufacturer with 36 spare part SKUs over a one year period. In fitting demand distributions to the data over 17 different families, they found that the best-fit 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 from the online book retailer Amazon.com to obtain an estimate on the sales quantity of the books. In their numerical experiments, they identified that the Pareto distribution with a parameter of 1.2 was a reasonable approximation to the demand data. The Internet has particularly fueled the phenomenon of the long tail where niche products gives rise to a large share of the total sales for online retailers, popularly referred to as the long-tail phenomenon (see Anderson [2], Brynjolfsson, Hu and Simester [11]). 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 from this empirical evidence.

3 Bounds with the First and n th Moment

Consider an ambiguity set defined as follows:

$$\mathcal{F}_{1,n} = \left\{ F \in \mathbb{M}(\mathbb{R}_+) : \int_0^\infty dF(w) = 1, \int_0^\infty w dF(w) = m_1, \int_0^\infty w^n dF(w) = m_n \right\}, \quad (3.1)$$

where m_1 and m_n are the first and the n th moment respectively, satisfying $m_n \geq m_1^n > 0$. Note that for $m_n = m_1^n$, the only feasible distribution in the ambiguity set is the point m_1 that occurs with probability 1, which is a trivial case to deal with. Hence, we restrict our attention to the case where $m_n > m_1^n > 0$, from this point onwards. There are a couple of reasons for considering this ambiguity set as we discuss next:

- (a) We allow for any real value $n > 1$ (not necessarily integer) in the definition of $\mathcal{F}_{1,n}$. Clearly, when $n = 2$, this corresponds to the original model of Scarf [39]. Thus it provides a natural generalization of the original ambiguity set of Scarf denote by $\mathcal{F}_{1,2}$, but allows for the possibility of lighter tails

($n > 2$) or heavier tails ($n < 2$) than Scarf's model. In conjunction with the empirical evidence discussed in Section 2.2, assuming the knowledge of a finite mean also seems reasonable in most applications involving real demand data.

- (b) It preserves the simplicity of Scarf's [39] moment ambiguity set as it is parameterized by the choice of only three parameters - m_1 , m_n and n , rather than all moments up to n .

On the other hand, to the best of our knowledge, no closed-form representation for the worst-case bound $\sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+$ under this ambiguity set is currently known. Furthermore, this appears to be hard to find in closed-form². Instead, we develop lower and upper bounds on the worst-case expected value that is valid for large values of the order quantity q . Our approach is based on constructing approximately optimal primal-dual solutions that attains the bounds in this regime.

3.1 Lower Bound

To develop the lower bound, we first consider a related ambiguity set that was studied by Grundy [24] with a fixed n th moment only:

$$\mathcal{F}_n = \left\{ F \in \mathbb{M}(\mathbb{R}_+) : \int_0^\infty dF(w) = 1, \int_0^\infty w^n dF(w) = m_n \right\}, \quad (3.2)$$

where $\mathcal{F}_{1,n} \subseteq \mathcal{F}_n$. While Grundy [24] evaluated the worst-case bound for this ambiguity set in an option pricing context, the model remains largely unexplored in the newsvendor context. Grundy [24] characterized the unique two point distribution that attains the bound in $\sup_{F \in \mathcal{F}_n} \mathbb{E}_F[\tilde{d} - q]_+$. Given a value $q > (n-1)m_n^{1/n}/n$, the worst-case demand distribution was characterized as follows:

$$\tilde{d}_q^* = \begin{cases} \frac{qn}{n-1}, & \text{w.p. } \frac{(n-1)^n m_n}{n^n q^n}, \\ 0, & \text{otherwise,} \end{cases} \quad (3.3)$$

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

$$\sup_{F \in \mathcal{F}_n} \mathbb{E}_F[\tilde{d} - q]_+ = \begin{cases} \frac{m_n}{n} \left(\frac{n-1}{nq} \right)^{n-1}, & \text{if } q > \frac{n-1}{n} m_n^{1/n}, \\ m_n^{1/n} - q, & \text{if } 0 \leq q \leq \frac{n-1}{n} m_n^{1/n}. \end{cases}$$

The worst-case distribution in this ambiguity set 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 (3.3) to a three point distribution to make it feasible for the ambiguity set $\mathcal{F}_{1,n}$. This brings us to our first main result that provides a lower bound on the worst-case expected value for large

²The computation of moment bounds under this ambiguity set needs the solution to equations of the form $ax^n + bx + c = 0$. Using a symbolic computation package such as Mathematica fails to provide closed-form solutions for this equation for arbitrary values of n . Note that for general polynomial equations, the Abel-Ruffini impossibility theorem states that there is no algebraic solution for polynomials of degree five or higher with arbitrary coefficients, of which the current equation is a special case.

values of q .

Proposition 3.1. *Given an ambiguity set $\mathcal{F}_{1,n}$, there exists a positive quantity $\underline{q}(m_1, m_n, n)$ which depends on the parameters m_1 , m_n and n , such that:*

$$\sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+ \geq \frac{(m_n - m_1^n)}{n^n q^{n-1}} (n-1)^{n-1}, \quad \forall q \geq \underline{q}(m_1, m_n, n). \quad (3.4)$$

Proof. We derive the lower bound through the construction of a three point feasible distribution. The proof is developed in two steps. In Step 1, we provide a three point distribution by a modification of the two point worst-case distribution in (3.3) 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 for this distribution provides the desired lower bound in (3.4).

Step 1: Consider a three point random variable \tilde{d} with a distribution defined as follows:

$$\tilde{d} = \begin{cases} \frac{qn}{n-1}, & \text{w.p. } \frac{(m_n - m_1^n)}{n^n q^n} (n-1)^n, \\ d, & \text{w.p. } p, \\ 0, & \text{w.p. } 1 - p - \frac{(m_n - m_1^n)}{n^n q^n} (n-1)^n, \end{cases} \quad (3.5)$$

where we choose particular values of d and p as discussed next to ensure feasibility. Our choice of these values is such that one obtains a strictly positive value of d that is lesser than q and a probability p such that the first and n th moment constraints are met. To do so, we start by ensuring that the n -th moment constraint for this distribution is met as follows:

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

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

$$d^n p = m_1^n. \quad (3.6)$$

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

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

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

$$dp = m_1 - \frac{(m_n - m_1^n)}{n^{n-1} q^{n-1}} (n-1)^{n-1}. \quad (3.7)$$

Solving the two simultaneous equations (3.6) and (3.7) gives:

$$d = \frac{m_1^{n/(n-1)}}{\left(m_1 - \frac{(m_n - m_1^n)}{n^{n-1}q^{n-1}}(n-1)^{n-1}\right)^{1/(n-1)}}, \quad (3.8)$$

$$p = \frac{\left(m_1 - \frac{(m_n - m_1^n)}{n^{n-1}q^{n-1}}(n-1)^{n-1}\right)^{n/(n-1)}}{m_1^{n/(n-1)}}. \quad (3.9)$$

We note that for large values of q , the demand realization d in (3.8) is smaller than q as it is given by a strictly decreasing function of q . This condition is satisfied when:

$$q > \left(\left(\frac{m_n - m_1^n}{m_1} \right) \left(\frac{n-1}{n} \right)^{n-1} + m_1^{n-1} \right)^{1/(n-1)}. \quad (3.10)$$

Step 2: In this step, we ensure that (3.5) corresponds to a valid probability measure for the chosen d and p for large values of q . We start by observing that the value of p as given in (3.9) is strictly positive when:

$$q > \left(\frac{m_n - m_1^n}{m_1} \right)^{1/(n-1)} \left(\frac{n-1}{n} \right). \quad (3.11)$$

Clearly, $p < 1$ holds trivially for all q since $m_n > m_1^n$. Similarly, we can verify that the probability of the atom $qn/(n-1)$ is strictly less than 1 when:

$$q > (m_n - m_1^n)^{1/n} \left(\frac{n-1}{n} \right), \quad (3.12)$$

where the non-negativity of the probability holds trivially. Conditions (3.10), (3.11) and (3.12) are thus all valid for q large enough. The final condition that we need to ensure for the validity of the probability distribution for the chosen values of p and d is to verify that the probability of the atom 0 given by $1 - p - (m_n - m_1^n)(n-1)^n/(n^n q^n)$ is strictly positive. Plugging in the value of p , this is equivalent to verifying that for q large enough, the following inequality holds:

$$\left(1 - (m_n - m_1^n) \left(\frac{n-1}{nq} \right)^n \right)^{n-1} - \left(1 - \left(\frac{m_n - m_1^n}{m_1} \right) \left(\frac{n-1}{nq} \right)^{n-1} \right)^n > 0?$$

Let $C = (m_n - m_1^n)((n-1)/n)^{n-1}$. Then this is equivalent to verifying that for q large enough, the following inequality holds:

$$\Delta(q) := \left(1 - C \left(\frac{n-1}{n} \right) \frac{1}{q^n} \right)^{n-1} - \left(1 - \frac{C}{m_1} \frac{1}{q^{n-1}} \right)^n > 0?$$

Under the assumption that $q > \max((C(n-1)/n)^{1/n}, (C/m_1)^{1/(n-1)})$, we can apply the generalized

binomial theorem to both the terms to obtain an equivalent representation:

$$\Delta(q) = \sum_{k=0}^{\infty} \binom{n-1}{k} \left(-C \left(\frac{n-1}{n} \right) \frac{1}{q^n} \right)^k - \sum_{k=0}^{\infty} \binom{n}{k} \left(-\frac{C}{m_1} \frac{1}{q^{n-1}} \right)^k,$$

where $\binom{\alpha}{k}$ is defined as $\alpha(\alpha-1)\dots(\alpha-k+1)/k!$ for general values of α (not necessarily integer). Expanding the first few terms of the expressions yields:

$$\begin{aligned} \Delta(q) &= \left(1 - (n-1)C \left(\frac{n-1}{n} \right) \frac{1}{q^n} + \frac{(n-1)(n-2)}{2} C^2 \left(\frac{n-1}{n} \right)^2 \frac{1}{q^{2n}} - \dots \right) \\ &\quad - \left(1 - \frac{nC}{m_1} \frac{1}{q^{n-1}} + \frac{n(n-1)}{2} \frac{C^2}{m_1^2} \frac{1}{q^{2n-2}} - \dots \right), \\ &= \frac{nC}{m_1} \frac{1}{q^{n-1}} - (n-1)C \left(\frac{n-1}{n} \right) \frac{1}{q^n} + o\left(\frac{1}{q^{2n-2}}\right), \\ &> 0, \text{ for large enough values of } q, \end{aligned}$$

where the last inequality holds as the leading term has a non-negative coefficient. This implies that the distribution is feasible in $\mathcal{F}_{1,n}$ for large values of q . Since by construction, we have only one support point that is strictly greater than q , the expected value of the objective function for this feasible distribution is given as:

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

which provide a lower bound to the expected value. This leads to the desired result. \square

3.2 Upper Bound

To develop the upper bound on the worst-case expected value, we consider the dual formulation for the moment problem. We will show through an appropriate construction of a dual feasible solution in conjunction with the primal feasible distribution, that this bound is approximately optimal for large value of q .

Proposition 3.2. *Consider the ambiguity set $\mathcal{F}_{1,n}$.*

- (a) *When $n \in (2, \infty)$, there exists a positive quantity $\bar{q}(m_1, m_n, n)$ which depends on the parameters m_1, m_n and n , such that:*

$$\sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+ \leq \frac{(m_n - m_1^n)}{n^n q^{n-1} - n^2 m_1^{n-1} (n-1)^{n-1}} (n-1)^{n-1}, \quad \forall q \geq \bar{q}(m_1, m_n, n). \quad (3.13)$$

- (b) *When $n \in (1, 2)$, for any $\epsilon > 0$, there exists a positive quantity $\bar{q}(m_1, m_n, n, \epsilon)$ which depends on*

the parameters m_1 , m_n , n and ϵ such that:

$$\sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+ \leq \frac{(m_n - m_1^n)}{n^n q^{n-1} - n^2 m_1^{n-1} (n-1)^{n-1} - \epsilon} (n-1)^{n-1}, \quad \forall q \geq \bar{q}(m_1, m_n, n, \epsilon). \quad (3.14)$$

Proof. The primal formulation for the moment problem is given as:

$$\begin{aligned} & \sup \int_0^\infty [w - q]_+ dF(w) \\ & \text{s.t.} \int_0^\infty dF(w) = 1, \\ & \int_0^\infty w dF(w) = m_1, \\ & \int_0^\infty w^n dF(w) = m_n, \\ & F \in \mathbb{M}(\mathfrak{R}_+). \end{aligned} \quad (3.15)$$

The dual formulation is given as:

$$\begin{aligned} & \inf \quad y_0 + y_1 m_1 + y_n m_n \\ & \text{s.t.} \quad y_0 + y_1 d + y_n d^n \geq 0, \quad \forall d \geq 0, \\ & \quad \quad y_0 + y_1 d + y_n d^n \geq d - q, \quad \forall d \geq 0, \end{aligned} \quad (3.16)$$

where y_0 is the dual variable for the constraint that the total probability is equal to 1 and y_1 and y_n are the dual variables for the first and the n th moment constraints respectively. We derive the upper bound by constructing a dual feasible solution as follows. Define y_0 , y_1 and y_n as:

$$\begin{aligned} y_0 &= \frac{(n-1)m_1^n(n-1)^{n-1}}{n^n(q^{n-1} - K)}, \\ y_1 &= \frac{-nm_1^{n-1}(n-1)^{n-1}}{n^n(q^{n-1} - K)}, \\ y_n &= \frac{(n-1)^{n-1}}{n^n(q^{n-1} - K)}, \end{aligned} \quad (3.17)$$

where we choose a strictly positive K in a manner to be specified later in the proof. We first verify that this forms a dual feasible solution for a large enough q , specifically $q > K^{1/(n-1)}$, by checking each of the dual constraints. Observe that the dual feasibility constraints in (3.16) are equivalent to the following conditions:

$$\min_{d \geq 0} (y_0 + y_1 d + y_n d^n) \geq 0 \quad \text{and} \quad \min_{d \geq 0} (y_0 + q + (y_1 - 1)d + y_n d^n) \geq 0 \quad (3.18)$$

Since the values of the dual variables in (3.17) satisfy $y_n > 0$ and $y_1 < 0$ for the choice of K , the minimum value in the first dual constraint is obtained at $d^* = (-y_1/(ny_n))^{1/(n-1)}$. Substituting this

value, the first dual feasibility constraint is equivalent to verifying the condition:

$$\begin{aligned} y_0 &\geq -y_1 \left(\frac{-y_1}{ny_n} \right)^{1/(n-1)} - y_n \left(\frac{-y_1}{ny_n} \right)^{n/(n-1)}, \\ &= \frac{(-y_1)^{n/(n-1)}}{(ny_n)^{1/(n-1)}} \left(\frac{n-1}{n} \right). \end{aligned}$$

The choice of dual variables in (3.17) satisfy this condition at equality since:

$$\begin{aligned} y_0 - \frac{(-y_1)^{n/(n-1)}}{(ny_n)^{1/(n-1)}} \left(\frac{n-1}{n} \right) &= \frac{(n-1)m_1^n(n-1)^{n-1}}{n^n(q^{n-1}-K)} - \left(\frac{n-1}{n} \right) \frac{(nm_1^{n-1}(n-1)^{n-1})^{n/(n-1)}}{(n(n-1)^{n-1})^{1/(n-1)}n^n(q^{n-1}-K)}, \\ &= 0. \end{aligned}$$

Furthermore as $y_n > 0$, the minimum value in the second dual constraint is obtained at $d^* = ((1-y_1)/(ny_n))^{1/(n-1)}$. Substituting this in, the second dual feasibility constraint is equivalent to verifying if the following condition holds:

$$\delta(q) := y_0 + q - \frac{(1-y_1)^{n/(n-1)}}{(ny_n)^{1/(n-1)}} \left(\frac{n-1}{n} \right) \geq 0? \quad (3.19)$$

The choice of dual variables in (3.17) leads to the following expression:

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

This is equivalent to verifying that for q large enough, the following inequality holds:

$$(m_1^n(n-1)^n + n^nq(q^{n-1}-K))^{n-1} - (n^{n-1}(q^{n-1}-K) + m_1^{n-1}(n-1)^{n-1})^n > 0?$$

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

$$\Delta(q) := \left(1 + \frac{C^n}{q^n} - \frac{K}{q^{n-1}} \right)^{n-1} - \left(1 + \frac{(C^{n-1}-K)}{q^{n-1}} \right)^n > 0?$$

Applying the generalized binomial theorem to both the terms leads to the equivalent expression:

$$\begin{aligned} \Delta(q) &= \sum_{k=0}^{\infty} \binom{n-1}{k} \left(\frac{C^n}{q^n} - \frac{K}{q^{n-1}} \right)^k - \sum_{k=0}^{\infty} \binom{n}{k} \left(\frac{C^{n-1}-K}{q^{n-1}} \right)^k \\ &= \left(1 + (n-1) \left(\frac{C^n}{q^n} - \frac{K}{q^{n-1}} \right) + \frac{(n-1)(n-2)}{2} \left(\frac{C^n}{q^n} - \frac{K}{q^{n-1}} \right)^2 + \dots \right) \\ &\quad - \left(1 + n \frac{(C^{n-1}-K)}{q^{n-1}} + \frac{n(n-1)}{2} \left(\frac{C^{n-1}-K}{q^{n-1}} \right)^2 + \dots \right). \end{aligned}$$

We need to verify that $\Delta(q)$ is strictly positive for large values of q . To do so, we consider two cases:

(a) $n \in (2, \infty)$: In this case, the term $\Delta(q)$ reduces to:

$$\begin{aligned}\Delta(q) &= \frac{K - nC^{n-1}}{q^{n-1}} + \frac{(n-1)C^n}{q^n} + o\left(\frac{1}{q^{2n-2}}\right), \\ &= \frac{(n-1)C^n}{q^n} + o\left(\frac{1}{q^{2n-2}}\right),\end{aligned}$$

where the constant K is set equal to $nC^{n-1} = n(m_1(n-1)/n)^{n-1}$. This guarantees that $\Delta(q) > 0$ for large enough values of q as the leading term has a non-negative coefficient. The objective function value of this dual feasible solution reduces to:

$$y_0 + y_1 m_1 + y_n m_n = \frac{(m_n - m_1^n)(n-1)^{n-1}}{n^n q^{n-1} - n^2 m_1^{n-1} (n-1)^{n-1}},$$

which yields the desired result for $n > 2$.

(b) $n \in (1, 2)$: In this case, the term $\Delta(q)$ reduces to:

$$\Delta(q) = \frac{K - nC^{n-1}}{q^{n-1}} + \frac{(n-1)((n-2)K^2 - n(C^{n-1} - K)^2)}{2q^{2(n-1)}} + o\left(\frac{1}{q^n}\right).$$

Note that unlike the $n > 2$ case, setting $K = nC^{n-1}$ does not guarantee a dual feasible solution as this gives:

$$\Delta(q) = -\frac{n(n-1)C^{2(n-1)}}{2q^{2(n-1)}} + o\left(\frac{1}{q^n}\right),$$

where the leading term has a negative coefficient for large values of q . To modify the solution, for any strictly positive $\epsilon > 0$, we choose $K = nC^{n-1} + \epsilon/n^n$ which implies the term $\Delta(q)$ is given as:

$$\Delta(q) = \frac{\epsilon}{n^n q^{n-1}} + \frac{(n-1)((n-2)K^2 - n(C^{n-1} - K)^2)}{2q^{2(n-1)}} + o\left(\frac{1}{q^n}\right).$$

This is strictly positive for large enough values of q given the parameters and $\epsilon > 0$ as the leading term has a positive coefficient. The objective function value of this dual feasible solution reduces to:

$$y_0 + y_1 m_1 + y_n m_n = \frac{(m_n - m_1^n)(n-1)^{n-1}}{n^n q^{n-1} - n^2 m_1^{n-1} (n-1)^{n-1} - \epsilon},$$

which yields the desired result. \square

3.3 Numerical Example

We provide a numerical illustration of the quality of the bounds developed in Propositions 3.1 and 3.2 respectively. To compute the worst-case bound, we solve the dual formulation in (3.16) using a semidefinite program (SDP) for rational values of n . Assume that $n = p/q$, where p and q are strictly

positive integers. Then, the dual formulation is given as:

$$\begin{aligned}
& \inf && y_0 + y_1 m_1 + y_{p/q} m_{p/q} \\
& \text{s.t.} && y_0 + y_1 d + y_{p/q} d^{p/q} \geq 0, && \forall d \geq 0, \\
& && y_0 + q + (y_1 - 1)d + y_{p/q} d^{p/q} \geq 0, && \forall d \geq 0,
\end{aligned} \tag{3.20}$$

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

$$\begin{aligned}
& \inf && y_0 + y_1 m_1 + y_{p/q} m_{p/q} \\
& \text{s.t.} && y_0 + y_1 e^q + y_{p/q} e^p \geq 0, && \forall e \geq 0, \\
& && y_0 + q + (y_1 - 1)e^q + y_{p/q} e^p \geq 0, && \forall e \geq 0,
\end{aligned} \tag{3.21}$$

The constraints in (3.21) are the standard nonnegativity conditions on univariate polynomials over the half-line for which semidefinite representations are available (see Bertsimas and Popescu [3], Lasserre [28], Nesterov [35]). For example, with $n = 3$ ($p = 3, q = 1$), the semidefinite programming formulation is given as:

$$\begin{aligned}
& \inf_{y_0, y_1, y_3, a_1, b_1, c_1, a_2, b_2, c_2} && y_0 + y_1 m_1 + y_3 m_3 \\
& \text{s.t.} && \begin{bmatrix} y_0 & 0 & a_1 & b_1 \\ 0 & y_1 - 2a_1 & -b_1 & c_1 \\ a_1 & -b_1 & -2c_1 & 0 \\ b_1 & c_1 & 0 & y_3 \end{bmatrix} \succeq 0, \\
& && \begin{bmatrix} y_0 + q & 0 & a_2 & b_2 \\ 0 & y_1 - 1 - 2a_2 & -b_2 & c_2 \\ a_2 & -b_2 & -2c_2 & 0 \\ b_2 & c_2 & 0 & y_3 \end{bmatrix} \succeq 0,
\end{aligned} \tag{3.22}$$

while for $n = 3/2$ ($p = 3, q = 2$), the semidefinite programming formulation is given as:

$$\begin{aligned}
& \inf_{y_0, y_1, y_{3/2}, a_1, b_1, c_1, a_2, b_2, c_2} && y_0 + y_1 m_1 + y_{3/2} m_{3/2} \\
& \text{s.t.} && \begin{bmatrix} y_0 & 0 & a_1 & b_1 \\ 0 & -2a_1 & -b_1 & c_2 \\ a_1 & -b_1 & y_1 - 2c_2 & 0 \\ b_1 & c_2 & 0 & y_{3/2} \end{bmatrix} \succeq 0, \\
& && \begin{bmatrix} y_0 + q & 0 & a_1 & b_1 \\ 0 & -2a_1 & -b_1 & c_2 \\ a_1 & -b_1 & y_1 - 1 - 2c_2 & 0 \\ b_1 & c_2 & 0 & y_{3/2} \end{bmatrix} \succeq 0,
\end{aligned} \tag{3.23}$$

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 [44, 45]). The value of the bounds for the case $n = 3$ are much smaller than the bounds for $n = 3/2$. This can be explained by observing that when the value of n is larger, the existence of moments of higher order are implied in the ambiguity set, implying that the worst-case bounds will typically be less conservative. The figures also illustrate that the scaling behavior of the bounds as a function of q is asymptotically optimal.

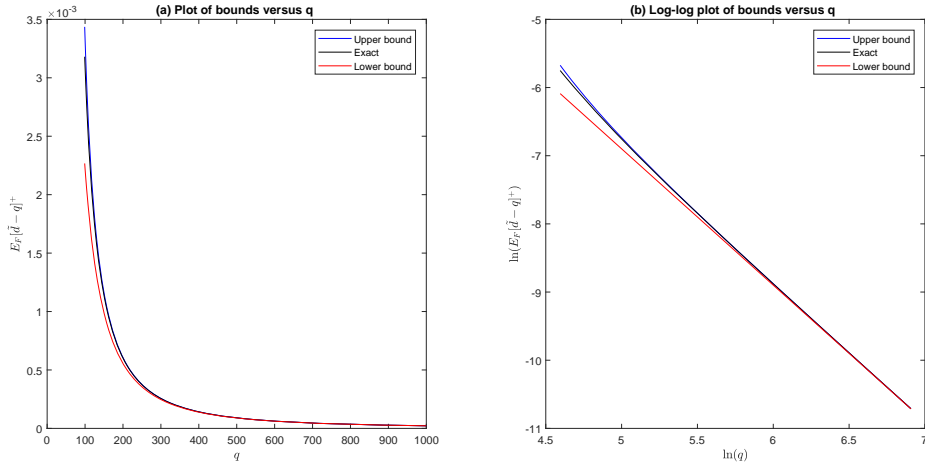


Figure 2: Plot (a) compares the upper and lower bounds with the exact bound obtained from solving a SDP as a function of q while plot (b) provides a log-log plot to characterize the scaling behavior. The mean demand is set to $m_1 = 50$ and the third moment is set to $m_3 = 125150$.

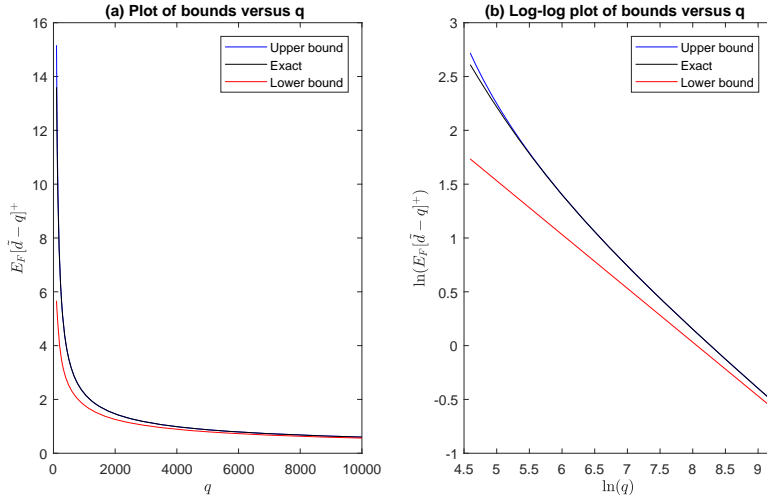


Figure 3: Plot (a) compares the upper and lower bounds with the exact bound obtained from solving a SDP as a function of q while plot (b) provides a log-log plot to characterize the scaling behavior. The mean demand is set to $m_1 = 50$ and the highest moment is set to $m_{3/2} = 500$ with $n = 3/2$. The bounds are plotted by setting $\epsilon = 0.2$.

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 so, we make use of the notion of regularly varying distributions that is popularly used to characterize heavy-tailed distributions (see Bingham, Goldie and Teugels [9], de Haan [16]). The key property of such distributions is that the behaviour at infinity is similar to the behaviour of a power law distribution. As we see in this section, this is exactly the type of behavior that F^* satisfies.

4.1 Regularly Varying Distributions

We first review the popular paradigm of distributions with regularly varying tails that has been used to characterize non-negative heavy-tailed distributions. A function $u : \mathfrak{R}_+ \rightarrow \mathfrak{R}_+$ is said to be regularly varying at infinity with index $\beta \in \mathfrak{R}$ if for all $t > 0$, we have:

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

We express this by $u \in \mathcal{RV}_\beta$. A non-negative random variable \tilde{d} with cumulative distribution function F is regularly varying if $\bar{F} := 1 - F \in \mathcal{RV}_{-\beta}$ for some $\beta \geq 0$. The distribution function is said to have tail parameter β if $\bar{F} \in \mathcal{RV}_{-\beta}$. Two classical examples of regularly varying random variables that are particularly relevant in our context are:

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

$$g(d) = \frac{\beta x_m^\beta}{d^{\beta+1}}, \quad \forall d \geq x_m, \quad (4.1)$$

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

$$\bar{F}(d) := \int_d^\infty g(x) dx = x_m^\beta d^{-\beta},$$

and hence clearly $\bar{F} \in \mathcal{RV}_{-\beta}$. Note that in Grundy's model discussed in Section 3.1, we obtain a characterization of the distribution F^* as follows:

$$F^*(d) = \mathbb{P}(\tilde{d}^* \leq d) = \begin{cases} 1 - m_n \left(\frac{n-1}{nd} \right)^n, & \text{if } d > \frac{n-1}{n} m_n^{1/n}, \\ 0 & \text{if } 0 \leq d \leq \frac{n-1}{n} m_n^{1/n}. \end{cases} \quad (4.2)$$

This defines a Pareto random variable as follows:

$$\tilde{d}^* = \text{Pareto} \left(\left(\frac{(n-1)m_n^{1/n}}{n}, n \right) \right), \quad (4.3)$$

where $\bar{F}^* \in \mathcal{RV}_{-n}$.

- (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 discussed in Section 2.1, we have $\bar{F}^* \in \mathcal{RV}_{-2}$.

Regularly varying functions have a rich theory (see Bingham, Goldie and Teugels [9], de Haan [16]) 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 [18], Resnick [37] 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 [29] shows the effect of integration on regularly varying functions. We state the special case relating to regularly varying distributions with at least first moment finite (see Resnick [37, 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 4.1 (Karamata's Theorem). *Suppose $u : \mathfrak{R}_+ \rightarrow \mathfrak{R}_+$ satisfies $u \in \mathcal{RV}_{-\beta}$ for some $\beta > 1$. Then $\int_x^\infty u(t) dt$ is finite, $\int_x^\infty u(t) dt \in \mathcal{RV}_{-\beta+1}$ and:*

$$\lim_{x \rightarrow \infty} \frac{xu(x)}{\int_x^\infty u(t) dt} = \beta - 1.$$

The next result provides the reverse implication to Karamata's theorem and shows what happens when a regularly varying function is differentiated; see Landau [31], Bingham, Goldie and Teugels [9, Theorem 1.6.1], de Haan [16, p. 23], Resnick [37, Theorem 0.7] for different formulations and proofs.

Theorem 4.2. *Suppose $u : \mathfrak{R}_+ \rightarrow \mathfrak{R}_+$ is locally integrable in $[0, \infty)$ and define:*

$$U(x) := \int_x^\infty u(t) dt.$$

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

$$\lim_{x \rightarrow \infty} \frac{xu(x)}{U(x)} = -\beta, \tag{4.4}$$

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

(b) *If $U \in \mathcal{RV}_{-\beta}$ for $\beta > 0$ and u is monotone, then (4.4) holds and $u \in \mathcal{RV}_{-\beta-1}$.*

4.2 Regularly Varying F^* from the Ambiguity Set $\mathcal{F}_{1,n}$

Propositions 3.1 and 3.2 indicate that in fact the tails of the worst-case expected value are close to a power-law (Pareto-like) tail with index $n - 1$. In this section, we show that there exists a random variable $d^* \sim F^*$ which attains the worst-case expected value for large values of q using the theory of regularly varying functions.

Theorem 4.3. *Given the ambiguity set $\mathcal{F}_{1,n}$, the following holds:*

(a) The following function satisfies:

$$E(q) := \sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+ \in \mathcal{RV}_{-(n-1)}.$$

(b) There exists a distribution function F^* (not in $\mathcal{F}_{1,n}$), such that if $\tilde{d}^* \sim F^*$, for $q \geq 0$ we have:

$$\mathbb{E}_{F^*}[\tilde{d}^* - q]_+ = \sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+.$$

Furthermore, $\bar{F}^* \in \mathcal{RV}_{-n}$ with $\mathbb{E}_{F^*}[(\tilde{d}^*)^k] < \infty$ for all $0 \leq k < n$ and $\mathbb{E}_{F^*}[(\tilde{d}^*)^k] = \infty$ if $k \geq n$.

Proof. (a) Notice that from Proposition 3.1, for q large enough, we have:

$$E(q) := \sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+ \geq C_1 \frac{1}{q^{n-1}}, \quad (4.5)$$

where C_1 is a constant given the parameters m_1 , m_n and n . Similarly, using Proposition 3.2, for large enough q we have:

$$E(q) \leq C_1 \frac{1}{q^{n-1}} \left(1 - \frac{C_2}{q^{n-1}}\right)^{-1}, \quad (4.6)$$

where C_2 is a constant given the parameters m_1 , m_n , n and ϵ . Hence combining (4.5) and (4.6), for large x and fixed $t > 0$, we get:

$$t^{-(n-1)} \left(1 - \frac{C_2}{x^{n-1}}\right) \leq \frac{E(tx)}{E(x)} \leq t^{-(n-1)} \left(1 - \frac{C_2}{(tx)^{n-1}}\right)^{-1}.$$

Since $1 - C_2/(tx)^{n-1} \rightarrow 1$ and $1 - C_2/t^{n-1} \rightarrow 1$, as $x \rightarrow \infty$, we can infer that:

$$\lim_{x \rightarrow \infty} \frac{E(tx)}{E(x)} = t^{-(n-1)}.$$

Hence $E(q) = \sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+ \in \mathcal{RV}_{-(n-1)}$.

(b) As a consequence of Theorem 2.1 and Section 3.1, page 32 in Shapiro and Kleywegt [43], we observe that given the ambiguity set $\mathcal{F}_{1,n}$, there exists a non-negative random variable \tilde{d}^* following a distribution F^* such that, for any $q \geq 0$,

$$E(q) = \sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+ = \mathbb{E}_{F^*}[\tilde{d}^* - q]_+.$$

We can write:

$$\mathbb{E}_{F^*}[\tilde{d}^* - q]_+ = \int_q^\infty \Pr(\tilde{d}^* > w) dw = \int_q^\infty \bar{F}^*(w) dw, \quad (4.7)$$

where $\bar{F}^* = 1 - F^*$. From part (a), we have for $q \rightarrow \infty$,

$$\int_q^\infty \bar{F}^*(w) dw = \mathbb{E}_{F^*}[d^* - q]_+ = E(q) \in \mathcal{RV}_{-(n-1)}.$$

Now since $-(n-1) < 0$ and \bar{F}^* is non-increasing, using Theorem 4.2 (b) (the converse part of Karamata's Theorem), we have $\bar{F}^* \in \mathcal{RV}_{-n}$. Note that for any $k \geq 0$, and some $C > 0$, we have:

$$\mathbb{E}_{F^*}[(\tilde{d}^*)^k] = \int_0^C t^{k-1} \bar{F}^*(t) dt + \int_C^\infty t^{k-1} \bar{F}^*(t) dt.$$

The first sum in the summand is bounded above by C^k which is finite. The integrand in the second term $t^{k-1} \bar{F}^*(t) \in \mathcal{RV}_\beta$ where $\beta = -(n-k) - 1$. For $k < n$, we have $\beta < -1$ and using Theorem 4.1, $\int_C^\infty t^{k-1} \bar{F}^*(t) dt$ is finite (which is what we need) and regularly varying with index $(n-k)$. Hence for $k < n$, we have $\mathbb{E}_{F^*}[(\tilde{d}^*)^k] < \infty$. Finally, we show that $\mathbb{E}_{F^*}[(\tilde{d}^*)^n] = \infty$ which implies that any higher moment will also be infinite. Note that for any $q > 0$ we have

$$\begin{aligned} E(q) - E(2q) &= \mathbb{E}_{F^*}[d^* - q]_+ - \mathbb{E}_{F^*}[d^* - 2q]_+, \\ &= \int_q^{2q} \bar{F}^*(y) dy, \\ &\leq q \bar{F}^*(q), \end{aligned}$$

since \bar{F}^* is non-increasing. Hence, for large enough q satisfying both (4.5) and (4.6) we have

$$\begin{aligned} \bar{F}^*(q) &\geq \frac{1}{q} [E(q) - E(2q)], \\ &\geq \frac{1}{q} \left[\frac{C_1}{q^{n-1}} - \frac{C_1}{(2q)^{n-1}} \left(1 - \frac{C_2}{(2q)^{n-1}} \right)^{-1} \right], \\ &\geq \frac{1}{q} \left[\frac{C_1}{q^{n-1}} - \frac{C_1}{(2q)^{n-1}} \times \left(1 - \frac{1}{2^{n-1}} \right)^{-1} \right] \quad (\text{for } q^{n-1} > C_2), \\ &= \frac{1}{q^n} C_3, \end{aligned}$$

where $C_3 = C_1(1 - 1/(2^{n-1} - 1))$. Hence we have for q large enough:

$$\begin{aligned} \mathbb{E}_{F^*}[(\tilde{d}^*)^n] &= \int_0^q t^{n-1} \bar{F}^*(t) dt + \int_q^\infty t^{n-1} \bar{F}^*(t) dt, \\ &\geq \int_q^\infty t^{n-1} \bar{F}^*(t) dt, \\ &\geq \int_q^\infty t^{n-1} \frac{C_3}{tn} dt = C_3 \int_q^\infty \frac{1}{t} dt = \infty. \end{aligned}$$

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

Theorem 4.3 shows that the distribution F^* is regularly varying with index n and the n th moment

of this distribution is infinite, while all moments strictly below n are finite. From the discussion in Section 2.1, this implies that optimal order quantity for the distributionally robust newsvendor problem with an ambiguity set $\mathcal{F}_{1,n}$ is also optimal for a heavy-tailed regularly varying distribution F^* with tail index n , for high critical ratios. Observe that this distribution does not lie in the set $\mathcal{F}_{1,n}$, since the n th moment is infinite, yet the optimal q^* from the robust model remains optimal for this distribution for high-critical ratios.

4.3 Numerical Example

We provide a numerical illustration of the behavior of the optimal order quantity from the distributionally robust model and compare it with the optimal order quantities for other distributions in the ambiguity set. To find the optimal order quantities, we solve the dual formulations for $n = 3$, from before as follows:

$$\begin{aligned}
& \inf_{y_0, y_1, y_3, a_1, b_1, c_1, a_2, b_2, c_2, q} && y_0 + y_1 m_1 + y_3 m_3 + (1 - \alpha)q \\
& \text{s.t.} && \begin{bmatrix} y_0 & 0 & a_1 & b_1 \\ 0 & y_1 - 2a_1 & -b_1 & c_1 \\ a_1 & -b_1 & -2c_1 & 0 \\ b_1 & c_1 & 0 & y_3 \end{bmatrix} \succeq 0, \\
& && \begin{bmatrix} y_0 + q & 0 & a_2 & b_2 \\ 0 & y_1 - 1 - 2a_2 & -b_2 & c_2 \\ a_2 & -b_2 & -2c_2 & 0 \\ b_2 & c_2 & 0 & y_3 \end{bmatrix} \succeq 0, \\
& && q \geq 0,
\end{aligned} \tag{4.8}$$

while for $n = 3/2$, the dual formulation is given as follows:

$$\begin{aligned}
& \inf_{y_0, y_1, y_{3/2}, a_1, b_1, c_1, a_2, b_2, c_2} && y_0 + y_1 m_1 + y_{3/2} m_{3/2} + (1 - \alpha)q \\
& \text{s.t.} && \begin{bmatrix} y_0 & 0 & a_1 & b_1 \\ 0 & -2a_1 & -b_1 & c_2 \\ a_1 & -b_1 & y_1 - 2c_2 & 0 \\ b_1 & c_2 & 0 & y_{3/2} \end{bmatrix} \succeq 0, \\
& && \begin{bmatrix} y_0 + q & 0 & a_1 & b_1 \\ 0 & -2a_1 & -b_1 & c_2 \\ a_1 & -b_1 & y_1 - 1 - 2c_2 & 0 \\ b_1 & c_2 & 0 & y_{3/2} \end{bmatrix} \succeq 0, \\
& && q \geq 0.
\end{aligned} \tag{4.9}$$

In Figure 4, we compare the order quantities from the robust model with other distributions in the set including the exponential, Pareto and lognormal distributions. As the figure illustrates, the distributionally robust newsvendor optimal order quantity possesses a heavy-tail optimality property that is

characterized by the class of regularly varying distributions. In comparison, the asymptotic behavior of the optimal order quantities for light-tailed distributions such as the exponential distribution or even the heavy-tailed lognormal distribution (see Figure 4d) that does not lie in the class of regularly varying distributions is different.

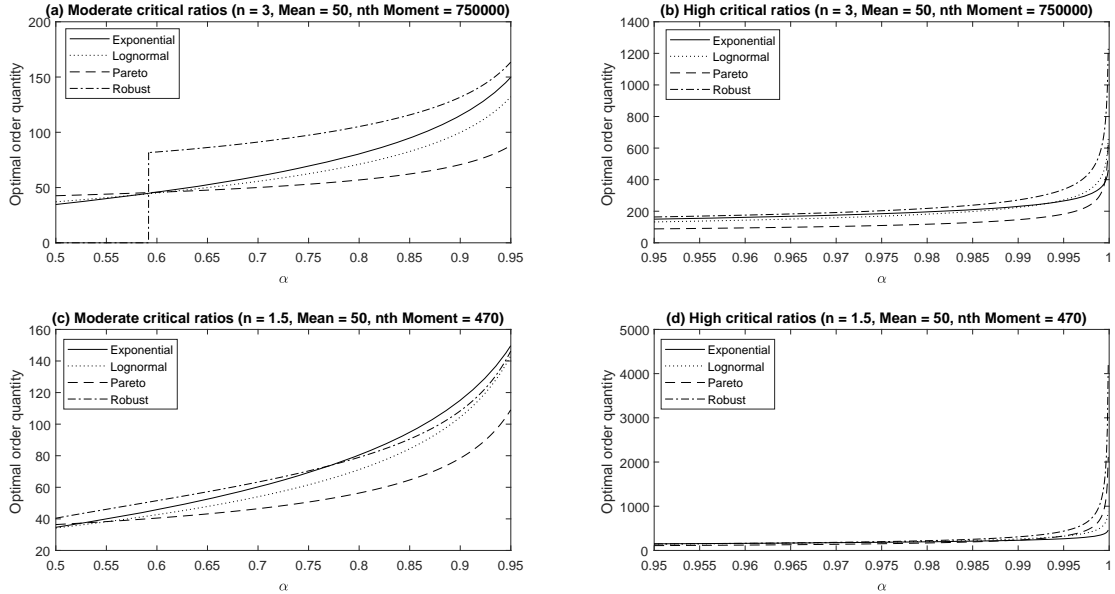


Figure 4: The plots at the top compare the optimal order quantities for the exponential (light-tailed), lognormal (heavy-tailed), Pareto (regularly varying) and the distributionally robust model for moderate and high values of critical ratios with mean demand 50 and third moment 750000. The plots at the bottom compare the optimal order quantities with mean demand 50 and the 3/2th moment 470.

5 Computational Experiments

In this section, we discuss how the model can be applied in practice by discussing the calibration of the ambiguity set from data. We also compare the performance of the order quantities from different methods including our proposed robust approach in a monthly car sales dataset.

5.1 Data and Methodology

Given a dataset of independent and identically distributed observations, we split the data into a training set and a test set, each consisting of roughly 50% of the total number of available data points. The training set is used to find moments of the dataset which include m_1 and m_2 for Scarf’s model and m_1 and m_n for our robust model. Note that given any finite number of observations, sample moments of all orders will exist in the training dataset and hence estimating n , the highest order moment which exists is a challenging problem. To estimate n , the approach we adopt is to try and figure out whether the data is heavy-tailed or not, by estimating a tail parameter β and then computing m_n for an appropriately chosen $n < \beta$. We compare optimal order quantities using four different methods: (1) an empirical estimate

using an appropriate quantile of the training data, (2) a normal quantile assuming a normal demand distribution, (3) Scarf's model, and (4) our proposed robust model. Given that the training dataset for demand is x_1, \dots, x_N and the test dataset is y_1, \dots, y_M we perform the following computations. We use the empirical moment estimators for the k th moment:

$$m_k = \frac{1}{N} \sum_{i=1}^N x_i^k,$$

where k is an appropriately chosen positive real number, given by $k = 1, 2$ or n . Based on these estimates, the order quantities are computed for different values of critical ratios $\alpha = 1 - c/p$. In our analysis, we then compute and plot the average profits in the test data for different values of α in the range $[0.65, 0.995]$. The order quantities are computed as follows:

1. EMPIRICAL QUANTILE: The optimal order quantity $q_E(\alpha)$ is obtained by computing the empirical α -quantile of the training data.
2. NORMAL MODEL: The optimal order quantity for the normal demand distribution is given as:

$$q_\Phi(\alpha) = m_1 + \sqrt{(m_2 - m_1^2)}\Phi^{-1}(\alpha),$$

where Φ denotes the standard normal cdf.

3. SCARF'S MODEL: The optimal order quantity is obtained from:

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

4. ROBUST MOMENT MODEL: For the robust moment model, we first estimate the tail parameter for the training set. We use linearity of mean excess plots to ascertain if the data is heavy-tailed (see Das and Ghosh [14]). For i.i.d random variables X_1, \dots, X_N and a threshold u , the empirical mean excess function is defined as

$$M_E(u) := \frac{\sum_{i=1}^N (X_i - u)\mathbf{1}_{[X_i > u]}}{\sum \mathbf{1}_{[X_i > u]}}.$$

It is known from the Pickands-Balkema deHaan theorem (see Theorem 5.2 on page 149 in McNeil, Frey and Embrechts [32]) that for distributions in the maximum domain of attraction of a generalized Pareto distribution (with finite mean), the plot $\{(X_{(k)}, M_E(X_{(k)})) : 1 < k \leq n\}$ for small values of k appears linear. Moreover, a positive slope indicates that the distribution is heavy-tailed, a zero or negative slope indicates light-tails (see Davison and Smith [15] and Das and Ghosh [14]).

The next step is to find the tail parameter when the data is heavy-tailed. In our computational

experiments, we compute the tail parameter even when there is an indication that the data might not be heavy-tailed, to judge the performance of the robust model. For i.i.d random variables X_1, \dots, X_N from a distribution function F where $\bar{F} \in \mathcal{RV}_{-\beta}$, we use the popular Hill estimator (see Hill [25]) which is defined as:

$$H_{k,N} := \frac{1}{k} \sum_{i=1}^k \log \frac{X_{(i)}}{X_{(k+1)}}$$

where $X_{(i)}$ denotes the i th largest value among X_1, \dots, X_n . Classical results in extreme value theory have shown that $H_{k,N} \rightarrow 1/\beta$ in probability as $k, N, N/k \rightarrow \infty$ (see de Haan and Resnick [17] and Resnick [38]). However, the choice of $k < N$ is not always an easy task, and has been an active area of research (see Scarrott and MacDonald [40]). For our datasets, we use a typical choice of $k = 0.4N$. It is possible to also to test whether the data is heavy-tailed or not, by plotting the Hill estimate for various values of k and checking if it stabilizes for small values of k . With the estimated value $\hat{\beta} = H_{k,N}^{-1}$, we proceed to compute m_n for the robust model by choosing a reasonably large $n < \beta$. This is because, if $\bar{F} \in \mathcal{RV}_{-\beta}$, then $\mathbb{E}[X^\gamma] < \infty$ for any $\gamma < \beta$, but $\mathbb{E}[X^\beta]$ may not necessarily exist. Note that in our examples, we chose n to be a rational number, to solve the SDP. We thus make use of the Hill estimator to identify roughly the highest order moment that exists from the sample training data. The exact optimal order quantity $q_{RE}(\alpha)$ is computed using a semidefinite program as described in the previous sections.

Given the order quantities, obtained using the four different methods, we compute the average profit in the test dataset as follows:

$$\text{AverageProfit}(q) = \frac{1}{M} \sum_{i=1}^M p \min(q, y_i) - cq = \frac{1}{M} \sum_{i=1}^M \min(q, y_i) - (1 - \alpha)q,$$

where we take $p = 1$ and $\alpha = 1 - c$. We do a similar computation for the training dataset.

5.2 Norway New Car Sales in the Period 2007-2017

The dataset that we consider provides sales of new cars in Norway in the period from 2007-2017. This dataset was obtained from <https://www.kaggle.com/dmi3kno/newcarsalesnorway> and was published by Opplysningsrådet for Veitrafikken (OFV) (<http://www.ofvas.no/>). The dataset consists of monthly car sales by make (e.g., Volkswagen, Toyota, etc.) from January 2007 to January 2017 in Norway. We make use of the sales data as a proxy for demand for the car makes. We concentrate on a few makes of cars which have reasonable number of data points (the maximum possible is 121). Specifically, we consider three different car makes - Volvo, Jaguar and Jeep, for which, we first investigate the mean excess plot over a threshold in Figure 5.

We observe from Figure 5, that the mean excess plots over a threshold slope in a negative direction for Volvo, indicating the possibility of light tails and are positively sloped for Jaguar and Jeep, indicating the possibility of more heavy-tailed behavior. We then proceed to estimate the tail parameter for each

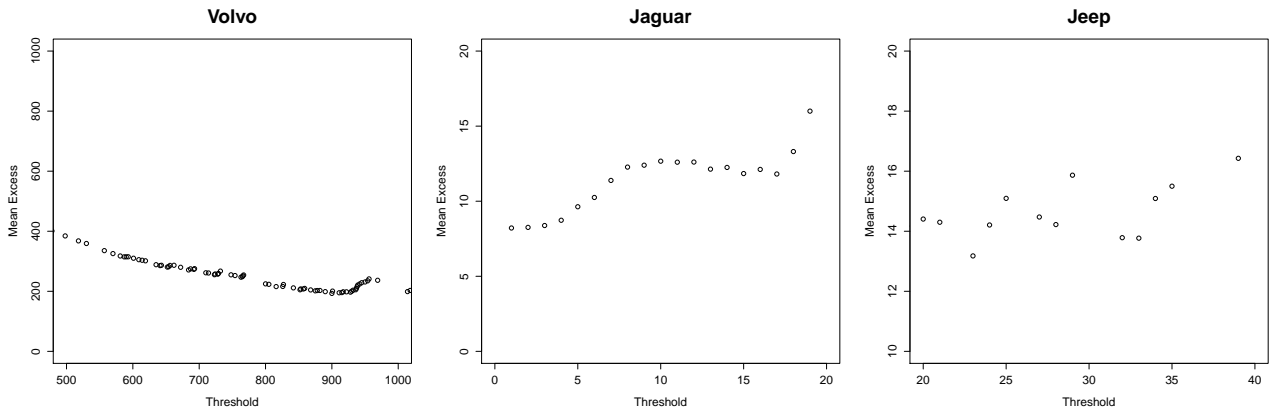


Figure 5: Mean excess plot over a threshold (the top 10-12 points are truncated) for the makers: Volvo, Chevrolet, Jaguar, and Jeep for monthly sales from January 2007 to January 2017.

of the datasets. The number of months for which sales are observed for Volvo, Jaguar and Jeep are respectively 121, 117 and 109, of which roughly half are used for the training set. The Hill estimator for the tail parameter β uses $k = \lfloor 0.4N \rfloor$ (where N is the size of the training set) to get the following estimates:

$$\hat{\beta}(\text{Volvo}) = 5.02, \quad \hat{\beta}(\text{Jaguar}) = 2.07, \quad \hat{\beta}(\text{Jeep}) = 1.67.$$

The three left hand plots in Figure 6 provide more detailed Hill estimate plots for various values of k in the range $[10, N)$. We see more stability in the plots for the cases of Jaguar and Jeep in comparison to the Volvo, providing a case for a heavy-tailed underlying distribution for these two car makes. For these two datasets, the tail parameters are also close to two or lesser, whereas Volvo gives a parameter estimate of 5.02, which is much higher. To implement the robust moment model, we respectively make the following choices of $n < \beta$ as proposed by our computation of $\hat{\beta}$:

$$n(\text{Volvo}) = 5, \quad n(\text{Jaguar}) = 2, \quad n(\text{Jeep}) = 5/3.$$

In Figure 6, we compare the average profit in the test set for the four different methods in the center plot. We observe that for Volvo, the average profits look reasonably close. In case of the Jaguar and Jeep car makes, we have a much stronger indication of a heavy-tailed underlying distribution (right plots in Figure 5) as well as the Hill plots (left plot, third and fourth row in Figure 6). We observe that for the Jaguar make, the order quantity obtained from the robust model is equivalent to Scarf's model since we chose $n = 2$ and this outperforms the other two models. In the case of Jeep sales with the heaviest tails, we observe that the robust model with $n = 5/3$ outperforms Scarf's model with $n = 2$, for a large range of critical ratios. This helps illustrate the value of calibrating n carefully for this example and using it in a prescriptive manner to make ordering decisions. We observe that for extremely high values of the critical ratio, the solution from the empirical quantile and the normal model do well in the test set. This can be partly explained by noting that the data set is comparatively small and particularly for very high critical ratios, the optimal order quantity from the normal model

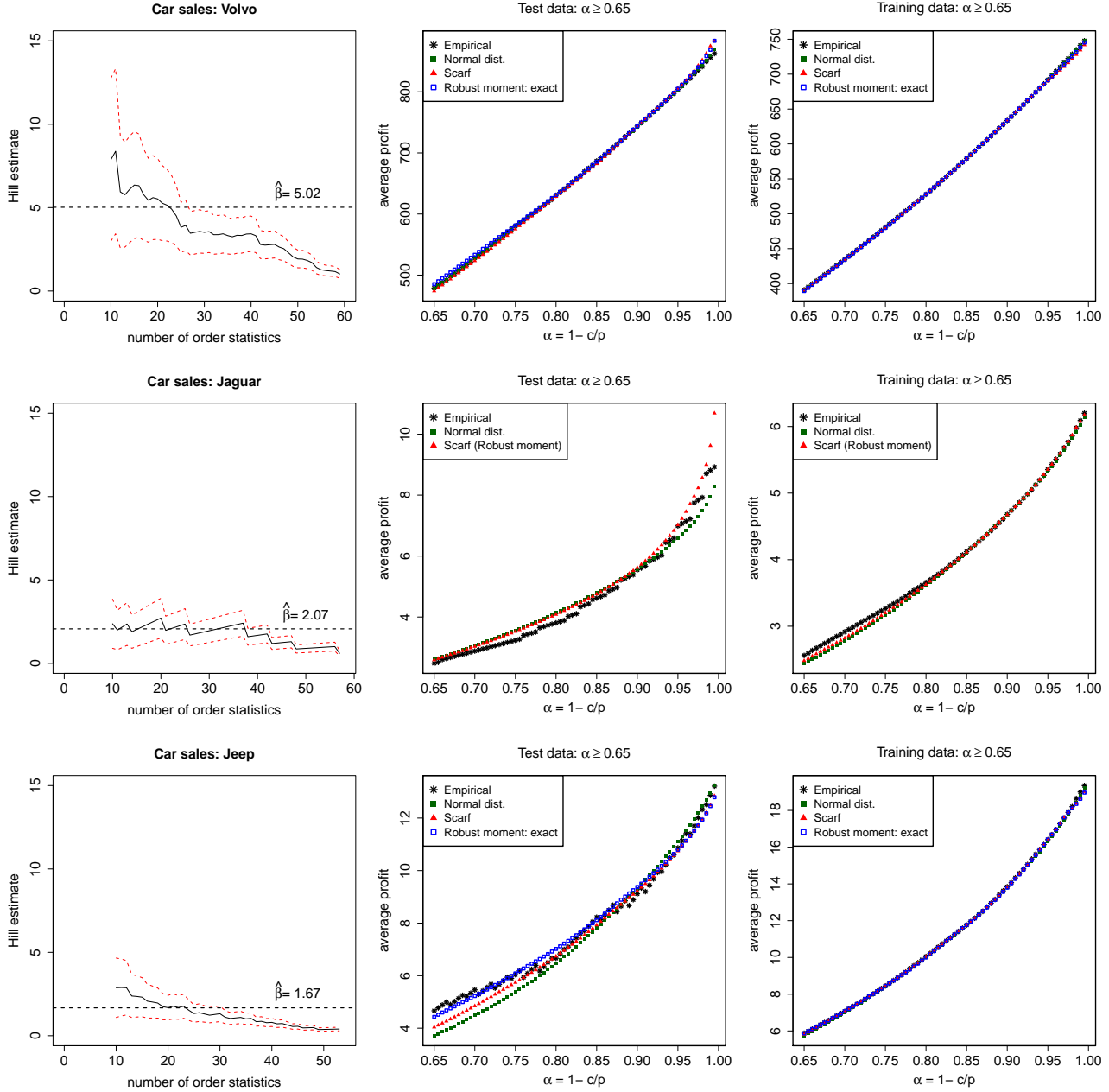


Figure 6: From top to bottom row for monthly sales of Volvo, Jaguar and Jeep we plot respectively left to right the following: (left) Hill estimator of the training set for $10 \leq k < N$ (left), (center) average profit from the test data set using optimal order quantities from the empirical estimate (stars), normal distribution (filled square), Scarf's model (triangle), and the robust model (hollow square), (right) average profit in the training dataset.

will be much lower than the robust model. Assuming that the training set and tests are similar, one needs significantly more samples in the test set to observe the out-performance for such high critical ratios. In the training set, as we would expect, the empirical solution performs the best, but the loss in profit from the other models does not appear to be significant.

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 n -th moment, the optimal order quantity is also optimal for a regularly varying with tail index n . This provides a characterization of a new distribution, which do not lie in the original ambiguity set, but for which the order quantity from a robust model, continues to remain optimal. We also discuss calibrating the ambiguity set, using results from heavy-tailed distributions and then comparing the optimal order quantities across different methods. The analysis indicates that the optimal order quantities obtained under the robust formulation works quite well, especially for cases where the underlying demand distribution is quite heavy-tailed (one example is a case where variance does not exist).

Several interesting questions still remain to be answered. While our results provide a characterization of the distribution F^* in the tails with moment information, it would be interesting to see if there is a more precise analytical characterization of the center of the distribution F^* . Secondly, it would be interesting to characterize the distribution F^* for other types of ambiguity sets that include distributions around a nominal distribution using probability metrics such as ϕ -divergence or the Wasserstein metric. We believe this will help managers better understand as to the types of data under which, the solutions from other robust models will do. Lastly, implications of these results for multidimensional newsvendor problems need to be studied. We leave this for future research.

Acknowledgements

The authors would 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 their valuable inputs on this research.

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] C. Andersson, *The long tail*, Wired Magazine, 12, 10, 170-177.
- [3] D. Bertsimas and I. Popescu, *On the relation between option and stock prices: A convex optimization approach*, Operation Research, 50, 2, 358-374, 2002.

- [4] D. Bertsimas and I. Popescu, *Optimal inequalities in probability theory: A convex optimization approach*, SIAM Journal on Optimization, 15, 3, 780-804, 2005.
- [5] 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.
- [6] 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.
- [7] A. Ben-Tal and E. Hochman, *Stochastic programs with incomplete information*, Operations Research, 24, 2, 336-347, 1976.
- [8] K. Bimpkis and M. G. Markakis, *Inventory pooling under heavy-tailed demands*, Management Science, 62, 6, 1800-1813, 2016.
- [9] N. H. Bingham, C. M. Goldie and J. L. Teugels, *Regular variation*, Cambridge University Press, 1989.
- [10] J. Blanchet and K. Murthy, *On distributionally robust extreme value analysis*, Working Paper, 2016.
- [11] E. Brynjolfsson, Y. Hu and D. Simester, *Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales*, Management Science, 57, 8, 1373-1386, 2011.
- [12] A. Clauset, C. R. Shalizi and M. E. J. Newman, *Power law distributions in empirical data*, SIAM Review, 51, 4, 661-703, 2009.
- [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] B. Das and S. Ghosh, *Weak limits of exploratory plots in the analysis of extremes*, Bernoulli, 19(1), 308-342, 2013.
- [15] A. Davison, R. Smith, *Models for exceedances over high thresholds*, Journal of the Royal Statistical Society. Series B, 52 (3), 393-442, 1990.
- [16] L. de Haan, *On regular variation and its application to the weak convergence of sample extremes*, Mathematisch Centrum Amsterdam, 1970.
- [17] L. de Haan and S.I. Resnick, *On asymptotic normality of the Hill estimator*, Stochastic Models, 14, 849-867, 1970.
- [18] P. Embrechts, C. Klüppelberg and T. Mikosch, *Modelling extreme events for insurance and finance*, Springer-Verlag, Berlin, 1997.
- [19] P. M. Esfahani and D. Kuhn, *Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations*, Mathematical Programming, 2017.
- [20] 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.
- [21] G. Gallego, *New bounds and heuristics for (Q,r) policies*, Management Science, 44, 2, 219233, 1998.
- [22] 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.
- [23] R. Gao and A. J. Kleywegt, *Distributionally robust stochastic optimization with Wasserstein distance*, Working Paper, 2017
- [24] B. D. Grundy, *Option prices and the underlying asset's return distribution*, The Journal of Finance, 46, 3, 1045-1069, 1991.
- [25] B. M. Hill, *A simple general approach to inference about the tail of a distribution*, The Annals of Statistics, 3, 1163-1174, 1975.
- [26] 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, 315334, 1986.

- [27] H. Lam and C. Mottet, *Tail analysis without parametric models: A worst-case perspective*, Operations Research, 2017.
- [28] J. B. Lasserre, *Bounds on measures satisfying moment conditions*, The Annals of Applied Probability, 12, 3, 1114-1137, 2002.
- [29] J. Karamata, *Sur un mode de croissance régulière des fonctions*, Mathematica (Cluj), 4, 38 -53, 1930.
- [30] B. Li, R. Jiang and J. L. Mathieu, *Ambiguous risk constraints with moment and unimodality information*, Working Paper, 2017.
- [31] E. Landau, *Darstellung und Begründung Einiger Neueren Ergebnisse der Funktionentheorie*, Springer-Verlag, Berlin, 1916.
- [32] A. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management Concepts, Techniques and Tools - Revised Edition*, Princeton University Press, 2015.
- [33] A. Müller, D. Stoyan, *Comparison methods for stochastic models and risks*. Wiley, Chichester 2002.
- [34] K. Natarajan, M. Sim and J. Uichanco, *Asymmetry and ambiguity in newsvendor models*, To appear in Management Science, 2018.
- [35] 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.
- [36] I. Popescu, *A semidefinite programming approach to optimal-moment bounds for convex classes of distributions*, Mathematics of Operations Research, 30, 3, 632-657, 2005.
- [37] S. I. Resnick, *Extreme values, regular variation and point processes*, Reprint of the 1987 original, Springer, 2008
- [38] S. I. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer., Springer, 2007.
- [39] 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.
- [40] C. Scarrott and A. MacDonald, *A review of extreme value threshold estimation and uncertainty quantification*, Revstat Statistical Journal, 10(1), 33–60, 2012.
- [41] M. Shaked and G. Shanthikumar, *Stochastic orders*, Springer Series in Statistics, New York, 2007.
- [42] A. Shapiro and S. Ahmed, *On a class of minimax stochastic programs*, SIAM Journal on Optimization, 14, 4, 1237-1249, 2004.
- [43] A. Shapiro and A. Kleywegt, *Minimax analysis of stochastic problems*, Optimization Methods and Software, 17, 523-542, 2002.
- [44] K. C. Toh, M. J. Todd, and R. H. Tutuncu, *SDPT3 a matlab software package for 750 semidefinite programming*, Optimization Methods and Software, 11, 545581, 1999.
- [45] K. C. Toh, M. J. Todd, and R. H. Tutuncu, *Solving semidefinite-quadratic-linear programs using SDPT3*, Mathematical Programming, 95, 189217, 2003.
- [46] B. Van Parys, P. J. Gouart and D. Kuhn, *Generalized Gauss inequalities via semidefinite programming*, Mathematical Programming, 156, 1-2, 271-302, 2016
- [47] Z. Wang, P. W. Glynn and Y. Yu, *Likelihood robust optimization for data-driven problems*, Computational Management Science, 13, 241-261, 2015.
- [48] L. F. Zuluaga, J. F. Peña 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.