Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Asymmetry and Ambiguity in Newsvendor Models

Karthik Natarajan, Melvyn Sim, Joline Uichanco

MANAGEMENT SCIENCE
Vol. 64, No. 7, July 2018, pp. 3146–3167
ISSN 0025-1909 (print), ISSN 1526-5501 (online)
http://pubsonline.informs.org/journal/mnsc/

Asymmetry and Ambiguity in Newsvendor Models

Karthik Natarajan,[a] Melvyn Sim,[b] Joline Uichanco[c]

[a] Engineering Systems and Design, Singapore University of Technology and Design, Singapore 487372; [b] NUS Business School, National University of Singapore, Singapore 117591; [c] Ross School of Business, University of Michigan, Ann Arbor, Michigan 48109

Contact: karthik_natarajan@sutd.edu.sg (KN); dscsimm@nus.edu.sg, http://orcid.org/0000-0001-9798-2482 (MS); jolineu@umich.edu, http://orcid.org/0000-0003-2110-7088 (JU)

Received: December 3, 2014
Revised: February 14, 2016; November 23, 2016
Accepted: February 5, 2017
Published Online in Articles in Advance: June 27, 2017

https://doi.org/10.1287/mnsc.2017.2773

Abstract. A basic assumption of the classical newsvendor model is that the probability distribution of the random demand is known. But in most realistic settings, only partial distribution information is available or reliably estimated. The distributionally robust newsvendor model is often used in this case where the worst-case expected profit is maximized over the set of distributions satisfying the known information, which is usually the mean and covariance of demands. However, covariance does not capture information on asymmetry of the demand distribution. In this paper, we introduce a measure of distribution asymmetry using second-order partitioned statistics. Semivariance is a special case with a single partition of the univariate demand. With mean, variance, and semivariance information, we show that a three-point distribution achieves the worst-case expected profit and derive a closed-form expression for the distributionally robust order quantity. For multivariate demand, the distributionally robust problem with partitioned statistics is hard to solve, but we develop a computationally tractable lower bound through the solution of a semidefinite program. We demonstrate in numerical experiments that asymmetry information significantly reduces expected profit loss particularly when the true distribution is heavy tailed. In computational experiments on automotive spare parts demand data, we provide evidence that the distributionally robust model that includes partitioned statistics outperforms the model that uses only covariance information.

History: Accepted by Yinyu Ye, optimization.
Funding: The research of the first author was partly supported by the Ministry of Education [MOE Tier 2 Grant MOE2013-T2-2-168, "Distributional Robust Optimization for Consumer Choice in Transportation Systems"].
Supplemental Material: The electronic companion is available at https://doi.org/10.1287/mnsc.2017.2773.

Keywords: inventory production • approximations • heuristics • uncertainty • stochastic models • programming • nonlinear theory

1. Introduction

The newsvendor problem is the foundation of many operations management models ranging from inventory control to revenue management. In the classical version, a firm decides on how many units to order for n products before observing the actual demand. Once demand is realized, it is satisfied as much as possible with the inventory on hand. A per-unit purchase cost c_i is incurred for each unit of product i, and the firm receives a per-unit revenue p_i for each sale of product i, where $p_i > c_i$. We assume without loss of generality that unsold units at the end of the sales period have zero salvage value. Given a joint probability distribution f of the multivariate random demand $\tilde{\mathbf{d}} = (\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_n)$, the firm will choose an order quantity vector $\mathbf{q} = (q_1, q_2, \ldots, q_n)$ that maximizes the total expected profit:

$$\max_{\mathbf{q}\in\mathscr{Q}} E_f\left(\sum_{i=1}^{n} p_i \min(\tilde{d}_i, q_i) - \sum_{i=1}^{n} c_i q_i\right),$$

where $\mathscr{Q}$ is a convex set that defines the feasible order quantities. If the demand is continuously distributed, then the optimal solution to the classical newsvendor problem satisfies the first-order optimality conditions. If $\mathscr{Q} = \mathbb{R}_+^n$, then the newsvendor problem is separable, and the optimal order quantity for product i is the $(p_i - c_i)/p_i$ quantile of the distribution of $\tilde{d}_i$. This ratio is commonly referred to as the critical ratio.

However, in most practical situations, it is difficult to elicit the exact distribution of the random demand. Consider a firm that only knows partial information on the joint demand distribution. If the firm commits to an order quantity based on a misspecified demand distribution, then this order quantity can potentially be suboptimal under the true demand distribution. A distributionally robust approach protects against adverse effects of distribution misspecification by choosing order quantities that perform well under any joint distribution satisfying the partial information. Mathematically, if $\mathscr{F}$ is the set of candidate joint distributions, then the distributionally robust counterpart of the newsvendor problem is

$$\max_{\mathbf{q}\in\mathscr{Q}} \inf_{f\in\mathscr{F}} E_f\left(\sum_{i=1}^{n} p_i \min(\tilde{d}_i, q_i) - \sum_{i=1}^{n} c_i q_i\right). \tag{1}$$

Problem (1) can be interpreted as a game against an adversarial nature who, given order quantities $\mathbf{q}$, chooses the distribution in $\mathcal{F}$ that results in the lowest expected profit. The solution to (1) is the best strategy to take against an adversarial nature.

The choice of the information in the distributionally robust problem is important. A common choice of the uncertainty set $\mathcal{F}$ is the set of all distributions with a fixed mean and covariance matrix. However, knowledge of mean and covariance does not provide information on the asymmetry of the demand distribution. Therefore the distribution set $\mathcal{F}$ includes distributions with various forms of asymmetry, and the resulting solution is conservative since it has to be robust to distributions with distinct asymmetries. On the other hand, if $\mathcal{F}$ is determined from third-order or higher moments, then the distributionally robust problem (1) becomes intractable. In this paper, our aim is to develop partial information sets for the distributionally robust newsvendor problem that are useful in practice and result in a tractable problem.

We motivate the approach proposed in this paper through data from a European automotive manufacturer. The manufacturer has provided daily demand data for 36 spare parts' stock-keeping units (SKUs) over the period of one year. Figure 1 shows the scatter plot matrix displaying correlation between three of these SKUs. One observation from the pairwise scatter plots is that the demands are strongly positively correlated. Since the distributionally robust newsvendor problem (1) with mean and covariance information is NP-hard (Hanasusanto et al. 2014), a possible tractable approach is to only use the marginal demand information (i.e., mean and variance of the demands), which results in n independent bounds of the type studied in Scarf (1958). However, this approach disregards correlation information of the joint demand, which, as Figure 1 shows, can be significant. Another observation from Figure 1 is that the histograms of the marginal demand distributions are asymmetric. In fact, when we fit the demand data of the 36 SKUs to different distribution families, most often the best-fitting distribution (i.e., has the lowest Bayesian information criterion) is asymmetric and heavy tailed; that is, the tail decays at a subexponential rate. The distributional asymmetry information implied by the data is lost by only capturing the mean and covariance. This is because deviations above and below the mean have an equivalent effect in the variance and covariance measures. As a result, under the true demand distribution, the expected profit of the distributionally robust solution might be significantly smaller than the true optimal expected profit.

In this paper, we propose to capture asymmetry through the second-order partitioned statistics, which we define through the first and second moments of the vector of partitioned random demands:

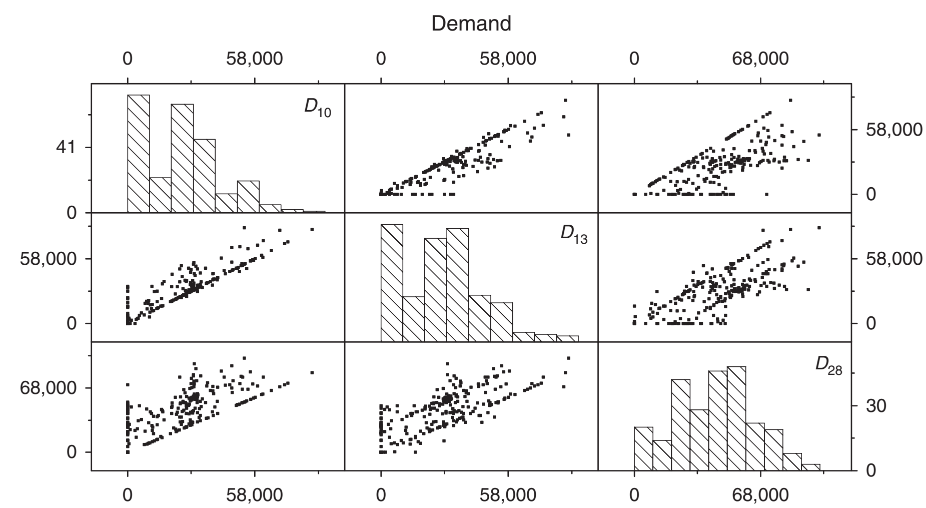
$$\begin{pmatrix} (\tilde{\mathbf{d}} - \mathbf{m})^+ \\ (\mathbf{m} - \tilde{\mathbf{d}})^+ \end{pmatrix},$$

where $\mathbf{m}$ is the mean vector of the random demand $\tilde{\mathbf{d}}$ and $\mathbf{x}^+$ denotes the component-wise maximum of the entry of the vector and 0. The second-order partitioned statistics measures correlations of the partitioned demand, where the partitioning naturally models demand asymmetry. Figure 2 shows the scatter plots displaying correlation of the partitioned demands for the three SKUs. We observe from this figure that there is a strong positive correlation in the partitioned demands, especially between the upper tails.

We next outline the main contributions in this paper:

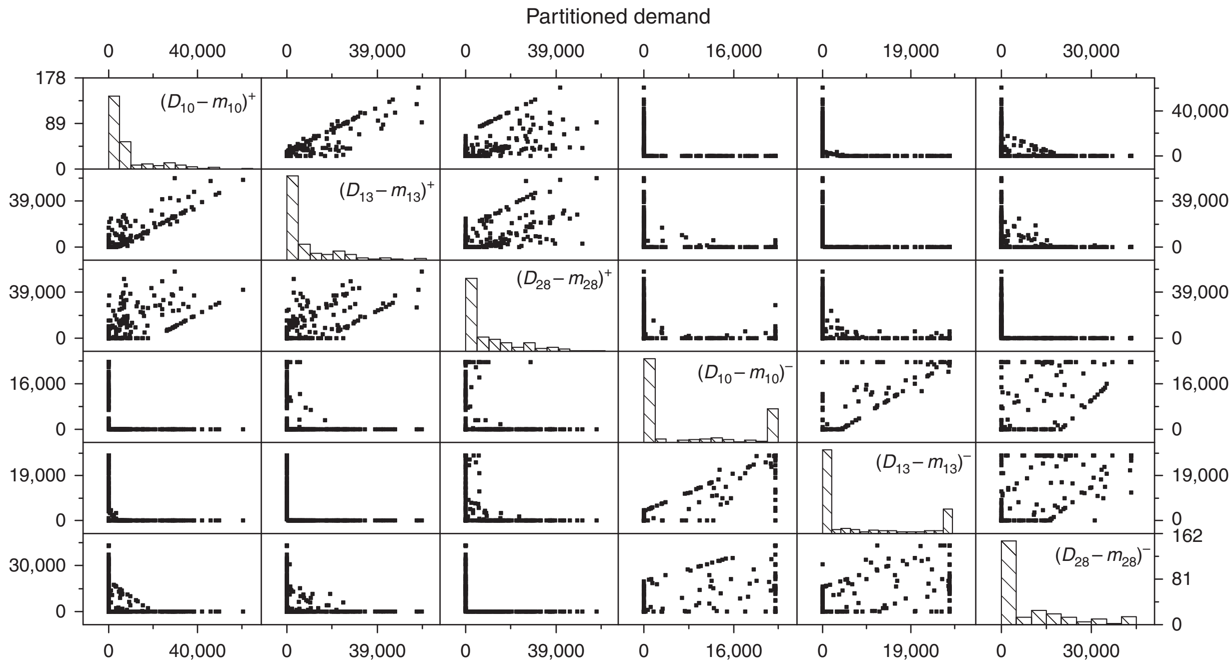
1. *Model of asymmetry through second-order moments of partitioned random demand*. We introduce a measure of

Figure 1. Scatter Plot Matrix Displaying Correlation Between the Demand of SKUs 10, 13, and 28



Note. The univariate histograms (diagonal panels) show that the daily demand distributions are asymmetric.

Figure 2. Scatter Plot Matrix Displaying Correlation Between Partitioned Demand of SKUs 10, 13, and 28



Note. A strongly positive correlation is observed for several of the partitioned demands.

asymmetry for the distributionally robust newsvendor problem (1) using second-order partitioned statistics. When $n = 1$, a special case of the second-order partitioned statistics is semivariance, a well-known risk measure (Markowitz 1959). We demonstrate in extensive numerical simulations that when the true demand distribution is asymmetric and heavy tailed, then compared to having only mean and variance information, the distributionally robust problem with partitioned statistics has a significantly smaller expected profit loss. We believe that this insight is of practical importance since it has been empirically observed in several industries that demand follows heavy-tailed distributions (see Chevalier and Goolsbee 2003, Gaffeo et al. 2008, Bimpikis and Markakis 2016).

2. *Closed-form solution for the single-item newsvendor with asymmetry*. We derive a closed-form expression for the solution to the distributionally robust newsvendor problem with mean, variance, and semivariance (MVS) information (Theorem 2.2). We provide the worst-case expected profit across all distributions satisfying the partial information, and we prove that a three-point distribution achieves this bound (Theorem 2.1). We also show that there is less ambiguity in the expected profit from including semivariance information, since the gap between the worst-case and best-case expected profit is significantly reduced (see Figure 4).

3. *Asymmetry through multiple partitions of random demand*. We generalize semivariance by introducing second-order statistics of multiple partitioned random demand. Increasing the number of partitions provides more information about distribution asymmetry. We derive a lower bound on the corresponding distributionally robust newsvendor problem through a second-order cone program (Theorem 3.1). If partitioned statistics are estimated from data, we provide guidelines on an appropriate number of partitions based on the number of samples, the critical ratio, and whether the distribution is known to be heavy tailed or light tailed.

4. *Semidefinite program lower bound on expected profit with mean and second-order partitioned statistics*. The general distributionally robust newsvendor problem (1) with mean and covariance information is NP-hard (Hanasusanto et al. 2014). Building on a reformulation technique in Natarajan and Teo (2017), we derive a lower bound on the expected profit with mean and second-order partitioned statistics through a semidefinite program (SDP). We show across a range of distributions that the expected value of additional information obtained from the partitioned statistics is significant in comparison to knowing only the covariance.

5. *Computational experiments on real-world data from an automotive spare parts manufacturer*. We conduct computational experiments to compare in a large-scale ($n = 36$) real-world manufacturing setting, the performance of the robust solution with partitioned statistics against that with mean and covariance. We show that if the moments are estimated from data, the robust solution with partitioned statistics achieves a significantly smaller total cost per day.

1.1. Literature Review

There has been a growing interest in approaches to tackle the newsvendor problem when the probability distribution of the demand is ambiguous. In this section, we review the literature on the single-period, single-item, and multi-item distributionally robust newsvendor problems. Given a set of univariate demand distributions $\mathcal{F}$, the single-item distributionally robust newsvendor problem is formulated as

$$\max_{q\in\mathcal{Q}\subseteq\mathbb{R}_+} \inf_{f\in\mathcal{F}} E_f(p\min(\tilde{d}, q) - cq). \tag{2}$$

Early research by Scarf (1958) assumed that the only information available on the demand is the mean and the variance for which he developed a closed-form expression for the optimal order quantity. Ben-Tal and Hochman (1976) extended this result to develop a closed-form optimal order quantity when the mean and the mean absolute deviation are known. One of the standard criticisms of these models is that it might be too conservative since the worst-case demand distributions are often two-point distributions, which might be too esoteric for practical applications. Scarf (1958) and Gallego and Moon (1993) numerically validated that the optimal order quantity assuming only mean and variance information is close to the optimal order quantity assuming a normal demand distribution with the same two moments, except for very high critical ratios. However, for demand distributions such as the exponential distribution, Scarf's optimal ordering quantity can be quite far away from the true optimal order quantity (see Zhu et al. 2016). To address the conservatism of these models, several alternatives have been proposed. One approach has been to consider alternate objective functions. Perakis and Roels (2008) and Yue et al. (2006) proposed to choose the order quantity based on the concept of min-max regret, which minimizes the maximum opportunity cost (regret) of not making the optimal decision if the true demand distribution from the set of distributions is indeed known. Yue et al. (2006) characterized the optimal order quantity for the min-max regret objective given the mean and variance of demand while Perakis and Roels (2008) characterized the optimal order quantity under various assumptions on the set of distributions including information on the mean, median, mode, symmetry, unimodality, and variance of demand. Perakis and Roels (2008) showed through extensive simulation experiments that the optimal order quantity for the maximum entropy distribution performs well under the regret criterion, thereby reducing some of the criticism of conservatism in these models. Levi et al. (2012) solved the min-max regret problem with absolute mean spread information, which they found in Levi et al. (2015) to be closely related to the regret objective. Andersson et al. (2013) also performed several numerical experiments under various distributions to compare the optimal order quantities under the maximin objective, the min-max regret objective, and the maximum entropy distribution. Zhu et al. (2013) extended the results of Perakis and Roels (2008) by minimizing the worst-case relative regret, defined as the ratio of the expected cost based on limited information to that based on complete information. Han et al. (2014) modified the newsvendor objective function by combining the expected profit with the standard deviation of profit, thus incorporating risk into the model. Given only the mean and variance of the demand, they developed a closed-form solution for the distributionally robust risk-averse newsvendor and showed that as risk aversion increases, the optimal order quantity deviates from Scarf's optimal order quantity toward the mean demand. Another approach that has been explored in tackling the conservatism of the models is to incorporate additional distributional information. For example, by incorporating higher order moment information on the demand, the single-item distributionally robust newsvendor problem can be reformulated as a semidefinite program (see Bertsimas and Popescu 2002). Although computationally tractable, the optimal order quantity cannot be found in closed form in these cases, and hence it is difficult to derive managerial insights even for a single item. For a fixed order quantity, the worst-case expected profit even with third moment information has a complicated representation involving roots of cubic equations (Jansen et al. 1986, De Schepper and Heijnen 2007, Zuluaga et al. 2009), and finding simple closed-form solutions does not seem easy. Attempts have also been made in the data-driven approach to incorporate robustness that allows for the distribution to deviate from the empirical distribution. Zhu et al. (2016) proposed a likelihood robust optimization approach where the objective is to maximize the worst-case expected profit in the set of distributions that makes the observed data achieve a certain level of likelihood. The optimal order quantity under the likelihood robust optimization approach is found by solving a convex program. Saghafian and Tomlin (2016) have recently developed an approach to combine demand realizations with partial distributional information in the form of bounds on the mean demand or the tail probability. By developing a Bayesian updating mechanism that integrates new demand observations with moment bounds, they showed that it can be beneficial to incorporate moment information into a data-driven approach when deciding order quantities.

Gallego and Moon (1993), building on the results of Scarf (1958), studied the multi-item distributionally robust newsvendor model. Assuming that for each item the mean and variance of the demand is known, they formulated the multi-item problem as a convex optimization problem under budget constraints on the total ordering costs. Their approach is based on applying Scarf's bound for the individual items worst-case expected profit in the objective function. Ben-Tal et al. (2013) proposed an alternative set of distributions for the multi-item newsvendor problem when the demand for each item is assumed to be a discrete random variable. Under their uncertainty set, the probability vector for the demand of each item is assumed to lie in a set around a sample probability distribution where the distance is defined using the concept of ϕ-divergence. The ϕ-divergence measure includes several statistical measures such as chi-squared divergence and Kullback–Leibler divergence as special cases. Under this uncertainty set, they showed that the distributionally robust multi-item newsvendor problem can be solved as a tractable convex program when only information on the marginal distributions is given. However, these models do not capture the correlation of demands. Given information on the correlations of demands, the distributionally robust multi-item newsvendor problem turns out to be much harder than the single-item problem. Building on the results of Delage and Ye (2010) and Bertsimas et al. (2010), Hanasusanto et al. (2014) showed that solving the multi-item newsvendor problem given the mean and covariance of demands is NP-hard. They also extended their model to include risk aversion in the objective function and allow for the demand distribution to be multimodal by assuming a mixture distribution where each distribution in the mixture is specified with mean and covariance information. They developed semidefinite programming formulations to solve this problem. However, it should be noted that the exact formulation of the multi-item newsvendor problem scales exponentially in the number of products. They also proposed heuristically solving the problem using simple quadratic decision rules, which can be viewed as applying Scarf's bound for each item. Wiesemann et al. (2014) recently proposed a class of ambiguity sets that contain all distributions with prescribed conic representable confidence sets and mean values residing on an affine manifold. For the distributionally robust multi-item newsvendor problem, they constructed ambiguity sets from the component-wise semi-mean absolute deviations and pairs of semi-mean absolute deviations and developed conic programming formulations for the distributionally robust newsvendor problem. Numerically, they showed that this approach captures dependence information among demands in a more tractable manner, but the formulation is exponential sized in the number of products (see Gorissen and den Hertog 2013 for a greater discussion on this issue in inventory models). Ardestani-Jaafari and Delage (2016) solved the multi-item distributionally robust newsvendor problem when the information that is available about the distribution includes the support in a budgeted uncertainty set, the mean and upper and lower bounds on the first-order partial moments. By formulating this problem as a linear program, they showed that the distributionally robust problem is tractable and the exact linear program scales polynomially in the number of items. However, their model does not capture all possible correlations among demand. Natarajan and Teo (2017) recently developed semidefinite programming relaxations for the multi-item distributonally robust newsvendor problem assuming that the mean and covariance information is given. The semidefinite program is developed by showing a connection of the multi-item newsvendor problem with the Boolean quadric polytope and using semidefinite relaxations of the Boolean quadric polytope. Their approach provides tighter bounds than applying Scarf's bound to each item and captures dependence information. While the bound is not tight in general, the size of the semidefinite program scales polynomially in the number of items.

A stream of literature that is relevant to this paper develops bounds in a portfolio optimization application using lower first- and second-order partial moments such as the semivariance. Chen et al. (2011) developed bounds on the semivariance using mean and variance information and applied it to distributionally robust portfolio optimization problems. Natarajan et al. (2010) developed bounds on the conditional value at risk of portfolios with only information on the mean and covariance and partitioned information. The objective of these streams of work is to better capture asymmetry information using first- and second-order partial moments.

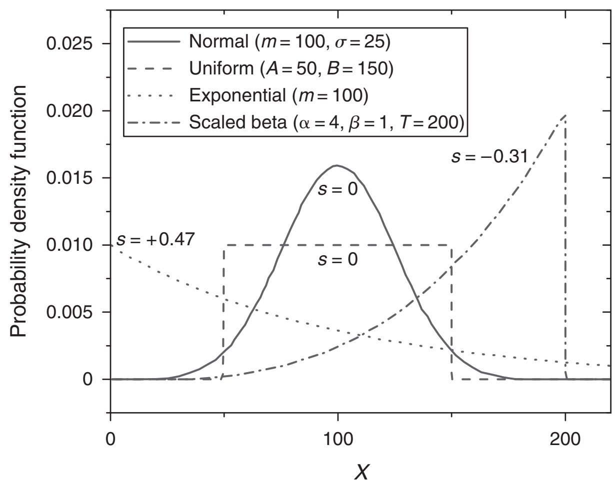
2. Modeling Asymmetry in the Single-Item Newsvendor Problem Using Semivariance

In this section, we introduce the second-order partitioned statistics in the single-item distributionally robust newsvendor problem (2). In this special case, second-order partitioned statistics is equivalent to knowledge of the standard deviation σ and of the normalized semivariance s defined as

$$s = \frac{E_f((\tilde{d}-m)^{+2}) - E_f((m-\tilde{d})^{+2})}{\sigma^2}.$$

We show that the distributionally robust newsvendor problem with mean, variance, and semivariance information admits a solution with a closed-form expression. We also show that semivariance information when available significantly reduces expected profit ambiguity and profit loss under the true distribution.

Figure 3. Some Common Probability Density Functions with Their Parameters and Normalized Semivariances



Note. The normalized semivariance can be interpreted as a measure of distribution asymmetry.

The value of the normalized semivariance s indicates whether the random demand's deviations from the mean are concentrated above or below the mean. The normalized semivariance s takes values between -1 and 1 since

$$\sigma^2 = E_f((\tilde{d} - m)^{+2}) + E_f((m - \tilde{d})^{+2}).$$

Figure 3 shows examples of common probability distributions and the normalized semivariances. Normal and uniform distributions have a normalized semivariance equal to 0. An exponential distribution has a normalized semivariance equal to approximately 0.47, while the normalized semivariance of a beta distribution can be positive or negative depending on the distribution's parameters. We can interpret the normalized semivariance as a weak form of distributional symmetry. All symmetric distributions (e.g., uniform, normal) must have a normalized semivariance of 0. However, the converse is not necessarily true.[1] A positively skewed distribution corresponds to a positive normalized semivariance. Similarly, a negatively skewed distribution corresponds to a negative normalized semivariance. The following proposition provides a necessary and sufficient condition on the mean, variance, and normalized semivariance of a nonnegative random variable.

Proposition 2.1. *Consider a nonnegative random variable with mean $m > 0$, standard deviation $\sigma > 0$, and normalized semivariance s. Then,*

$$\frac{\sigma^2 - m^2}{\sigma^2 + m^2} \le s < 1. \tag{3}$$

Moreover, given any triplet (m, σ, s) that satisfies condition (3), there exists a nonnegative distribution with these moments. If the lower bound on s is tight, then this distribution is unique.

Proof. See Section A.1 in the electronic companion. □

Consider the newsvendor model where the exact demand distribution is unknown, but the mean m, standard deviation σ, and normalized semivariance s are known. Given this information, the worst-case expected newsvendor profit with the mean, variance, and semivariance information for the order quantity q is

$$\begin{aligned}\underline{\Pi}^{\mathrm{MVS}}(q) = \inf_f\, & p\mathrm{E}_f(\min\{\tilde{d}, q\}) - cq\\ \text{s.t. } & \mathrm{E}_f(\tilde{d}) = m, \quad \mathrm{E}_f((\tilde{d} - m)^2) = \sigma^2,\\ & \mathrm{E}_f((\tilde{d} - m)^{+2}) - \mathrm{E}_f((m - \tilde{d})^{+2}) = s\sigma^2,\\ & \mathrm{E}_f(1) = 1, \quad f(d) \ge 0,\ \forall\, d \ge 0.\end{aligned} \tag{4}$$

Given only mean, variance, and semivariance information, the distributionally robust newsvendor problem chooses an order quantity q that maximizes $\underline{\Pi}^{\mathrm{MVS}}(q)$. Note that the moment problem (4) is a semi-infinite linear program. A classical result by Isii (1962) states that if the moment vector is an interior point of the set of feasible moment vectors, then strong duality holds in a moment problem, and the primal moment problem and its dual achieve the same optimal value. Based on Proposition 2.1, this is achieved when $m > 0$, $\sigma > 0$, and $s \in ((\sigma^2 - m^2)/(\sigma^2 + m^2), 1)$.

2.1. Tractability

In the following theorem (Theorem 2.1), we develop a closed-form expression for $\underline{\Pi}^{\mathrm{MVS}}(q)$. The proof of the theorem, which we provide in the electronic companion, consists of constructing a dual feasible solution and a corresponding primal feasible distribution that achieves the same objective value.

Theorem 2.1. *Consider the set of all nonnegative demand distributions with mean $m > 0$, standard deviation $\sigma > 0$, and normalized semivariance $s \in ((\sigma^2 - m^2)/(\sigma^2 + m^2), 1)$. For an order quantity q, a lower bound on the expected profit is*

$$\underline{\Pi}^{\mathrm{MVS}}(q) = (p - c)q - \frac{p(1-s)\sigma^2}{2m^2}q, \quad \text{if } q \in \left[0, \frac{m}{2}\right],$$

$$\underline{\Pi}^{\mathrm{MVS}}(q) = (p - c)q - \frac{p(1-s)\sigma^2}{8(m-q)},$$
$$\text{if } q \in \left[\frac{m}{2}, m - \frac{\sigma}{2}\sqrt{\frac{1-s}{1+s}}\right],$$

$$\underline{\Pi}^{\mathrm{MVS}}(q) = p\left(\frac{(1-s)}{2}q + \frac{(1+s)}{2}m - \frac{\sigma}{2}\sqrt{1-s^2}\right) - cq,$$
$$\text{if } q \in \left[m - \frac{\sigma}{2}\sqrt{\frac{1-s}{1+s}}, m + \frac{\sigma}{2}\sqrt{\frac{1+s}{1-s}}\right],$$

$$\underline{\Pi}^{\mathrm{MVS}}(q) = pm - cq - \frac{p(1+s)\sigma^2}{8(q-m)},$$
$$\text{if } q \in \left[m + \frac{\sigma}{2}\sqrt{\frac{1+s}{1-s}}, m + \frac{m(1+s)}{2(1-s)}\right],$$

$$\underline{\Pi}^{\mathrm{MVS}}(q)=\frac{p}{2}\left(m+bq-\sqrt{(bq-m)^2-(1-b)^2m^2+\frac{(1+s)\sigma^2 b}{2}}\right)-cq,\quad \textit{if } q\in\left[m+\frac{m(1+s)}{2(1-s)},\infty\right),$$

where

$$b=1-\frac{(1-s)\sigma^2}{2m^2}.$$

Moreover, among the set of distributions, there exists a distribution with at most three support points that achieves this bound.

Proof. See Section A.2 in the electronic companion. □

Note that $\underline{\Pi}^{\mathrm{MVS}}(q)$ is a concave function of q. This follows from observing that for any realization of the demand, the profit is a concave function of q. Linearity of expectations and the infimum operator maintains the concavity of the function $\underline{\Pi}^{\mathrm{MVS}}(q)$. Using Theorem 2.1, we can find the optimal order quantity q^*_{MVS} that maximizes the worst-case expected profit given mean, variance, and semivariance information. The following theorem gives the closed-form expression for the solution that is increasing in the critical ratio (profit margin) $(p-c)/p$.

Theorem 2.2. *A solution to the distributionally robust newsvendor model (1), where $\mathcal{F}$ is the set of all nonnegative demand distributions with mean $m>0$, standard deviation $\sigma>0$, and normalized semivariance $s\in((\sigma^2-m^2)/(\sigma^2+m^2),1)$ is*

$$q^*_{\mathrm{MVS}}=0,\quad \textit{if } \frac{p-c}{p}\in\left[0,\frac{(1-s)\sigma^2}{2m^2}\right],$$

$$q^*_{\mathrm{MVS}}=m-\frac{\sigma}{2}\sqrt{\frac{(1-s)p}{2(p-c)}},\quad \textit{if } \frac{p-c}{p}\in\left[\frac{(1-s)\sigma^2}{2m^2},\frac{1}{2}(1+s)\right],$$

$$q^*_{\mathrm{MVS}}=m+\frac{\sigma}{2}\sqrt{\frac{(1+s)p}{2c}},\quad \textit{if } \frac{p-c}{p}\in\left[\frac{1}{2}(1+s),1-\frac{1}{2}\frac{(1-s)^2}{(1+s)}\frac{\sigma^2}{m^2}\right],$$

$$q^*_{\mathrm{MVS}}=\frac{m}{b}+\frac{(pb-2c)}{2b}\sqrt{\frac{(1+s)\sigma^2 b-2(1-b)^2m^2}{2c(pb-c)}},\quad \textit{if } \frac{p-c}{p}\in\left[1-\frac{1}{2}\frac{(1-s)^2}{(1+s)}\frac{\sigma^2}{m^2},1\right],$$

where

$$b=1-\frac{(1-s)\sigma^2}{2m^2}.$$

Proof. See Section A.3 in the electronic companion. □

2.2. Conservatism

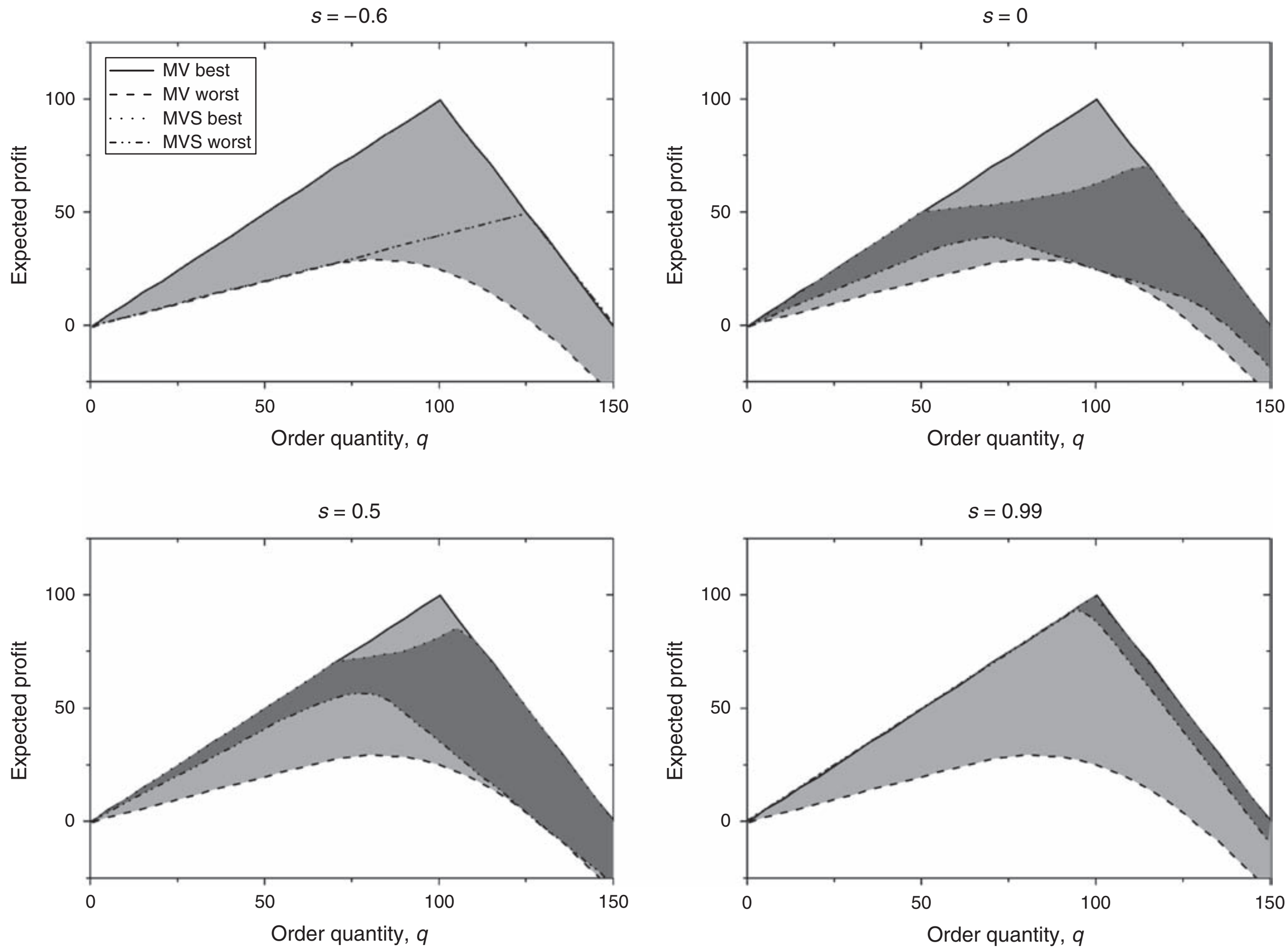
We next demonstrate that semivariance is informative since it significantly reduces ambiguity in a distributionally robust model. We demonstrate this by comparing the best-case and the worst-case expected profit bounds with mean-variance (MV) and MVS information sets. With the MV information set, the worst-case expected profit is the Scarf (1958) bound, while the best-case expected profit is $p\min(m,q)-cq$, which is obtained by applying Jensen's inequality (De Schepper and Heijnen 2007). With the MVS information set, the worst-case expected profit is given in Theorem 2.1, while the best-case expected profit can be found for a given value of q by solving a second-order cone program (see Section A.5 in the electronic companion).

Figure 4 plots the different bounds on the expected profit as a function of q. The four panels of the figure correspond to four different values of the normalized semivariance s. The normalized semivariance of the top-left panel is the smallest possible value for which the model is still well defined (by Proposition 2.1). We observe from Figure 4 that when there is only mean and variance information (MV), the gap between the best-case and worst-case expected profits can be large. This is because with only mean and variance information, there is a relatively large set of candidate distributions under which the expected profit has a large range. For example, with an order quantity $q=100$, the range of possible expected profits is from \$20 to \$100. Since the MVS feasible distribution set is a subset of the MV feasible distribution set, then it follows that the MVS expected profit bounds must be contained within the MV bounds. Moreover, observe from Figure 4 that the MVS bounds are influenced by the value of normalized semivariance s. Note that $s=0$ does not mean that there is no asymmetry information but rather implies that demand is equally distributed above and below the mean. Hence, when $s=0$, the MVS expected profit bounds are closer together than the MV expected profit bounds that do not assume any asymmetry information. Also note that the difference between the MVS expected profit bounds becomes smaller as s approaches its upper and lower limits. This is because at the limits of s, the MVS feasible distribution set admits fewer distributions. For instance, at $s=-0.6$, the feasible set consists of a single distribution (see Proposition 2.1), which explains why the MVS upper and lower bounds are equal. Hence, semivariance is informative for the distributionally robust newsvendor problem since it results in a significantly smaller range of expected profit values.

2.3. Expected Value of Semivariance Information

We next analyze the resulting loss in expected profit of ordering the distributionally robust solution instead of the solution for a true distribution. The distributionally robust model prescribes the solution q^*_{MVS}

Figure 4. Worst-Case and Best-Case Bounds on the Expected Profit with Mean-Variance and Mean-Variance-Semivariance Information



Notes. The gap between the best-case and worst-case bounds is large with only mean and variance information. Including semivariance information can significantly reduce this gap. Bounds were computed for parameters $p = 3$, $c = 2$, $m = 100$, and $\sigma = 50$.

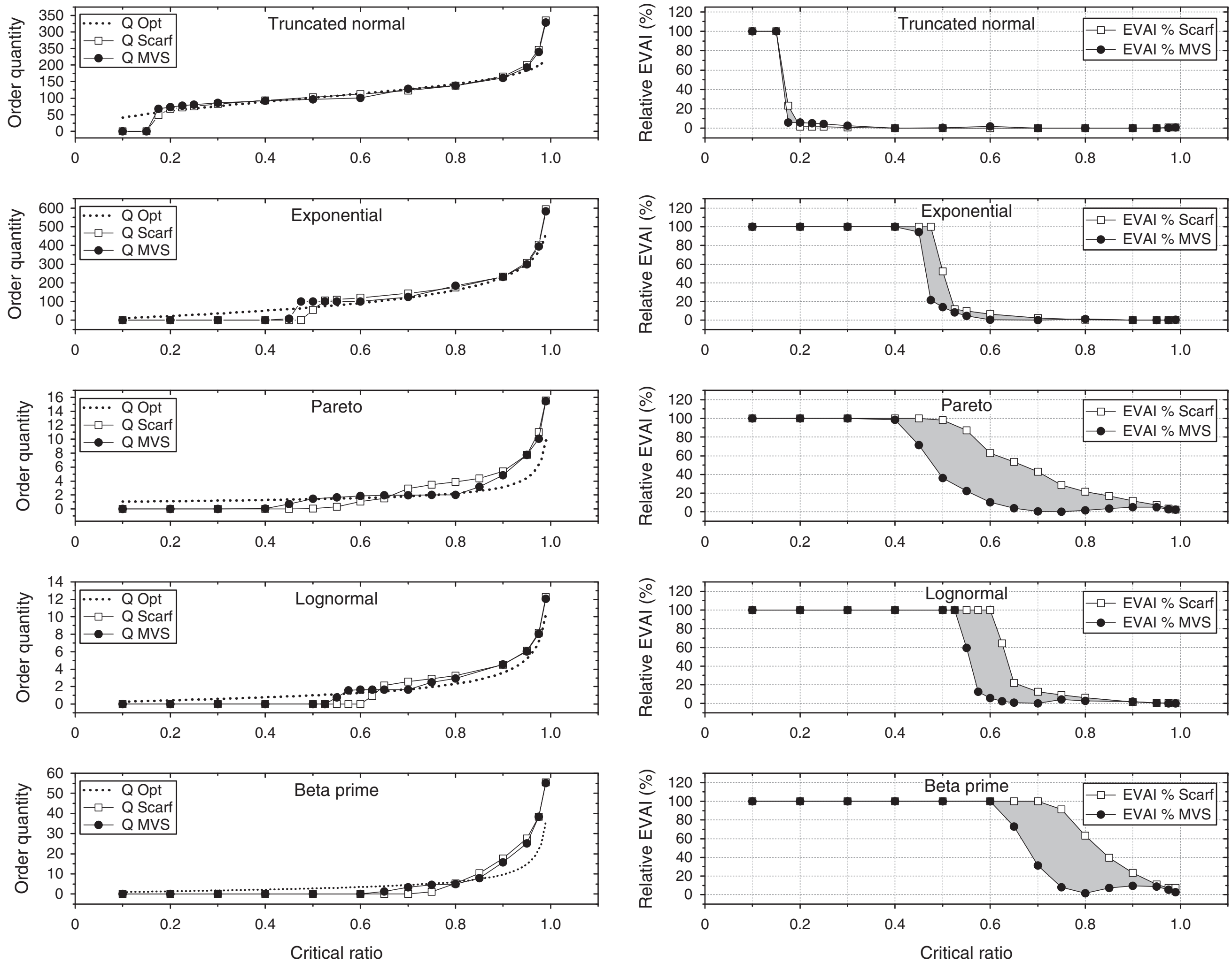
that maximizes the worst-case expected profit $\underline{\Pi}^{\mathrm{MVS}}(q)$ given the partial information. If Π is the expected profit function under the true (unknown) demand distribution and q^* its maximizer, then the profit loss of ordering a robust solution is $\Pi(q^*) - \Pi(q^*_{\mathrm{MVS}})$. This difference is also called the *expected value of additional information* (EVAI), since it represents the maximum willingness to pay to obtain full information about the distribution. If the EVAI is small under a partial information model, then there is little incentive to collect additional information since there is only a small expected profit gain from knowing the true distribution. Note that the gap between the peaks of the best-case and worst-case curves in Figure 4 provides an upper bound on the partial information model's EVAI. Clearly, from Figure 4, if there is only mean and variance information, the upper bound on EVAI is potentially high while if there is additional semivariance information, then this significantly reduces the EVAI upper bound.

We next find the actual EVAI of the mean-variance model and the mean-variance-semivariance model for different demand distributions and critical ratios. In these experiments, we use truncated normal, exponential, lognormal, beta prime, and Pareto distributions with parameters given in Table 1. Note that except for the truncated normal distribution, all other distributions are fairly positively skewed. Figure 5 displays the percentage difference of the robust solutions to the

Table 1. Demand Distributions Used in Single-Item Experiments

Distribution	Parameters	Normalized semivariance
Truncated normal	$m = 100, \sigma = 50, a = 0, b = \infty$	$s = 0.05$
Exponential	$m = 100$	$s = 0.47$
Lognormal	$m = 0, \sigma = 1$	$s = 0.70$
Beta prime	$\alpha = 5, \beta = 2.01$	$s = 0.88$
Pareto	$x_m = 1, k = 2.01$	$s = 0.89$

Figure 5. Difference of Robust Order Quantities to Optimal Order Quantity (Left Panels) and Expected Profit Loss of Robust Order Quantities (Right Panels) for Various Critical Ratios and Demand Distributions in Table 1



optimal order quantity (left panels of the figure) and the relative EVAI or expected profit loss of the robust models (right panels).

Under a truncated normal distribution, the robust solutions are both close to the optimal order quantity, except for very high and very low critical ratios. However, the expected profit loss of both robust models under the truncated normal distribution is relatively small (less than 3%) for all critical ratios greater than 0.2. This observation has also been made by Gallego and Moon (1993), who observed that the Scarf (1958) solution has a small profit loss under a normal distribution. Under an exponential distribution, the expected profit loss of the Scarf (1958) solution can be much higher for even moderate values of critical ratio. This also corresponds to the observation of Zhu et al. (2016), who noted that the Scarf (1958) solution may not perform as well for asymmetric distributions. However, we note that for high values of critical ratios, the expected profit loss of both robust models is still very small (less than 3%). The remaining three distributions—Pareto, lognormal, and beta prime—are all heavy tailed. We observe for all these heavy-tailed distributions that the EVAI can be significantly reduced by the semivariance information as illustrated by the shaded area in Figure 5. This implies that when the demand distribution is heavy tailed, then there is significant value in additional semivariance information, since it can bring the expected profit of the robust solution significantly closer to the optimal expected profit particularly for moderate to high critical ratios.

2.4. Rule-of-Thumb Ordering Heuristic

Scarf (1958) proved that the optimal solution to the distributionally robust newsvendor problem with only mean and variance information is

$$q^*_{\mathrm{MV}} = \begin{cases} 0, & \text{if } \dfrac{p-c}{p} < \dfrac{\sigma^2}{m^2+\sigma^2}, \\[2ex] m + \dfrac{\sigma}{2}\dfrac{p-2c}{\sqrt{c(p-c)}}, & \text{if } \dfrac{p-c}{p} > \dfrac{\sigma^2}{m^2+\sigma^2}. \end{cases}$$

Table 2. Rule-of-Thumb Ordering Heuristic for the Distributionally Robust Newsvendor Problem with Different Partial Information Models

	MV	MVS
Order below m	if $(p-c)/p < \frac{1}{2}$	if $(p-c)/p < \frac{1}{2}(1+s)$
Order above m	if $(p-c)/p > \frac{1}{2}$	if $(p-c)/p > \frac{1}{2}(1+s)$

The closed-form expression of Scarf (1958) and the optimal solution in Theorem 2.2 help provide some simple rules of thumb for determining order quantities with only partial information. We summarize these rules in Table 2. When only mean and variance are specified, the Scarf (1958) solution is to order above the mean if the profit margin is above $\frac{1}{2}$ and to order below the mean otherwise. Recall that in the newsvendor problem, the optimal order quantity is equal to the $(p-c)/p$ quantile (critical quantile). Hence, the implication of the Scarf (1958) solution is that the mean is equal to the median in the worst-case distribution. If, in addition, there is semivariance information specified, then in the worst-case distribution, the mean is shifted corresponding to the value of the asymmetry information. In particular, the mean of the worst-case distribution coincides with the $\frac{1}{2}(1+s)$ distribution quantile. Recall that s is positive (negative) for a positively (negatively) skewed distribution.

3. Modeling Asymmetry Using Multiple Demand Partitions

Semivariance is a special case of a second-order statistic of a partitioned demand distribution. In particular, semivariance captures the second-order statistics of the positive and negative parts of the demand minus its mean. We next generalize this idea by modeling second-order statistics of multiple demand partitions. Furthermore, in this section, we explore the effect of incorporating additional demand partitions on expected profit ambiguity and on estimation errors.

Assume that the probability density function f of the random demand $\tilde{d}$ has a support that is partitioned into $t+1$ intervals labeled $B_0, B_1, \ldots, B_t$, where $B_i = [b_i, b_{i+1})$ for $i = 0, \ldots, t-1$ and $B_t = [b_t, \infty)$. Since demand is nonnegative, we set $b_0 = 0$. The intervals are assumed to be set exogenously. We assume that there is only partial information on the demand distribution in the form of the mean demand and the second moments of the partitioned demand:

$$\begin{aligned} m &= E_f(\tilde{d}), \\ \sigma_i^2 &= E_f(\tilde{d}^2 \cdot 1_{[\tilde{d}\in B_i]}), \quad \forall\, i = 0, \ldots, t. \end{aligned} \tag{5}$$

Consider the distributionally robust (2) newsvendor problem, where the decision maker only has information in the form of the t-partitioned moments (5) of the random demand $\tilde{d}$. Given an order quantity q, the worst-case expected newsvendor profit is the optimal value to the following moment problem:

$$\begin{aligned} \Pi^{t\text{-part}}(q) = \inf_f\; & p\mathrm{E}_f(\min\{\tilde{d}, q\}) - cq \\ \text{s.t.}\quad & E_f(1) = 1, \\ & E_f(\tilde{d}) = m, \\ & E_f(\tilde{d}^2 \cdot 1_{[\tilde{d}\in B_i]}) = \sigma_i^2, \quad \forall\, i = 0, \ldots, t, \\ & f(x) \geq 0, \quad \forall\, x \geq 0. \end{aligned} \tag{6}$$

Note with 0-partitioned information, $\Pi^{0\text{-part}}$ reduces to the Scarf (1958) bound, which only assumes mean and variance information. The information set introduced in the beginning of Section 2 is a special case a 1-partitioned model using intervals $B_0 = [0, m)$ and $B_1 = [m, \infty)$. For more than one partition, partitioning can be done in different ways. For example, with $t > 1$ partitions, we can define equal-length partitions:

$$\begin{aligned} B_i &= [im, (i+1)m), \quad \text{for } i \leq t-1, \\ B_t &= [tm, \infty), \end{aligned} \tag{7}$$

or partitions with decreasing length:

$$\begin{aligned} B_1 &= [0, m), \\ B_i &= \left[\sum_{k=1}^{i-1} \frac{m}{k}, \sum_{k=1}^{i} \frac{m}{k}\right), \quad \text{for } i = 2, \ldots, t-1, \\ B_t &= \left[\sum_{k=1}^{t-1} \frac{m}{k}, \infty\right). \end{aligned} \tag{8}$$

3.1. Tractability

In the special case of semivariance information ($t = 1$), we have shown in Theorem 2.2 that we can derive a closed-form expression for the optimal solution. The key in the derivation is that we were able to find primal and dual solutions achieving the same optimal value. In fact, strong duality for moment problems holds when the moments lie in the interior feasible moment cone. For $t > 1$, it is difficult to find a closed-form expression for the optimal solution. However, as we prove in the following theorem, the dual can be formulated as a second-order cone program (SOCP) and provides the worst-case expected profit. As an SOCP, the dual can be solved tractably using commercial off-the-shelf solvers such as SeDuMi and CVX.

Theorem 3.1. *Consider the set of all nonnegative demand distributions with mean m, and partitioned variances $\{\sigma_i^2\}_{i=0}^t$ defined as*

$$\sigma_i^2 = E_f(\tilde{d}^2 \cdot 1_{[\tilde{d}\in B_t]}), \quad i = 0, 1, \ldots, t,$$

where $B_i = [b_i, b_{i+1}]$ for $i = 0, 1, \ldots, t-1$ and $B_t = [b_t, \infty)$. For an order quantity q, the worst-case expected profit is the optimal value to the following second-order cone program:

$$\Pi^{m\text{-Part}}(q) = \max_{t, r, \mathbf{y}, \boldsymbol{\tau}, \mathbf{v}} \left\{ pm - t - mr - \sum_{i=0}^{t} \sigma_i^2 y_i - cq \right\}$$

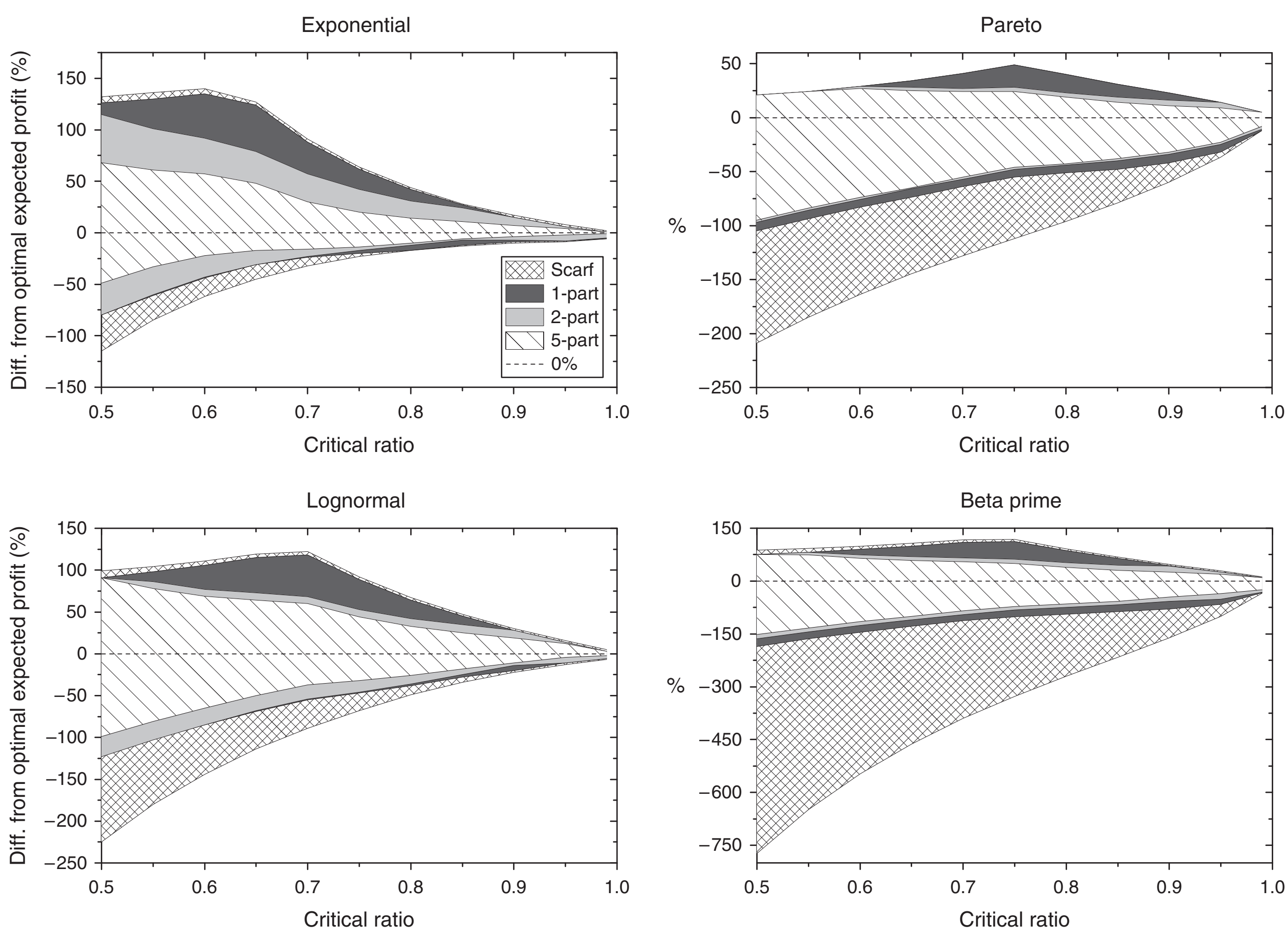
$$
\begin{aligned}
\text{s.t.} \quad & \left\| \begin{matrix} t + \tau_i(b_i b_{i+1} - 1) - y_i \\ r - \tau_i(b_i + b_{i+1}) \end{matrix} \right\|_2 \\
& \quad \le t + \tau_i(b_i b_{i+1} + 1) + y_i, \quad \forall\, i = 0, \ldots, t-1, \\
& \left\| \begin{matrix} t + pq + \upsilon_i(b_i b_{i+1} - 1) - y_i \\ r - p - \upsilon_i(b_i + b_{i+1}) \end{matrix} \right\|_2 \\
& \quad \le t + pq + \upsilon_i(b_i b_{i+1} + 1) + y_i, \quad \forall\, i = 0, \ldots, t-1, \\
& \left\| \begin{matrix} t + \tau_t b_t - y_t \\ r - \tau_t \end{matrix} \right\|_2 \le t + \tau_t b_t + y_t, \\
& \left\| \begin{matrix} t + pq + \upsilon_t b_t - y_t \\ r - p - \upsilon_t \end{matrix} \right\|_2 \le t + pq + \upsilon_t b_t + y_t, \\
& \tau_i \ge 0, \quad \upsilon_i \ge 0, \quad \forall\, i = 0, \ldots, t.
\end{aligned}
$$

Proof. See Section A.4 in the electronic companion. □

Theorem 3.1 gives the worst-case expected profit of q with the partial information (5). We can find the solution to the distributionally robust newsvendor problem with information (5) by choosing an order quantity q that maximizes this lower bound, which can be solved tractably as an SOCP. Additionally, in Section A.5 in the electronic companion, for a fixed quantity q, we derive an SOCP whose optimal value provides the best-case expected profit with (5).

Suppose that we misspecify a demand distribution under which q^* is the optimal order quantity. If we order q^* but the true demand distribution belongs to the uncertainty set assumed by the partial information models, then $[\underline{\Pi}(q^*), \bar{\Pi}(q^*)]$ gives the range of possible expected profits, where $\underline{\Pi}$ is the expected profit for the worst-case distribution and $\bar{\Pi}$ is the expected profit for the best-case distribution. If this range is large, then there is a high degree of ambiguity in the partial information model since the worst-case expected profit is significantly different from the best-case expected profit. Figure 6 shows the relative difference of the range of possible expected profits to the optimal expected profit as a function of the critical ratio under the Scarf, 1-partitioned, 2-partitioned, and

Figure 6. Feasible Expected Profit Regions of the Optimal Newsvendor Quantity with Different Partial Information Models with Demand Partitioning



Notes. The dashed line represents the expected profit under the true distribution. The largest range in expected profits occurs when there are no partitions, and each partition further reduces the range. Asymmetry information significantly reduces the expected profit ambiguity for heavy-tailed distributions (lognormal, beta prime, Pareto).

5-partitioned models with decreasing length partitions as in (8). Each of the four panels in the figure is for a different underlying demand distribution (exponential, Pareto, lognormal, beta prime with parameters specified in Table 1). Since Scarf uses only mean and variance information, the range of possible expected profits with this information is the largest. With more partitions, there is less profit ambiguity since the range of possible expected profits becomes narrower. Note, however, that asymmetry information has a significant effect in reducing expected profit ambiguity when the true demand distribution is heavy tailed (Pareto, lognormal, beta prime).

3.2. Data-Driven Experiments Using Multiple Demand Partitions

In practice, information is often estimated from demand data. Hence, the choice of the number of partitions should be guided by the potential effect of estimation errors of the partitioned moments on the performance of the distributionally robust model. In this section we demonstrate through simulation experiments that the following factors are important to take into consideration when choosing the number of partitions in a data-driven approach: (i) whether the distribution is light tailed or heavy tailed, (ii) the critical quantile, and (iii) the sample size. We use decreasing length partitions for the numerical experiments. For light-tailed distributions (e.g., normal, exponential), our experiments indicate that the robust solutions and the sample average approximation (SAA) solution have comparable errors, except when the critical ratio is very high and the sample size is small in which case SAA has large relative errors. For small sample sizes, increasing the number of partitions also does not seem to help for moderate values of the critical ratio since the estimation errors of the partitioned moments tend to dominate. On the other hand, for heavy-tailed distributions (e.g., Pareto, lognormal, beta prime) for moderate values of the critical ratio, the partitioned models clearly outperforms Scarf's model.

Suppose that we sample from the true demand distribution with a sample size $N = 5$, 10, 20, 40, 80. Using only this sample, we estimate the mean and t-partitioned second-order statistics ($t = 0, 1, 2$), and we set the order quantity based on the solution to the distributionally robust problem given the estimated moments. Figure 7 shows the box plots of the relative errors (profit loss) of the computed distributionally robust solutions when the demand is normally distributed with parameters specified in Table 1. For comparison, the figure also shows box plots of relative errors by the SAA method, which takes the critical quantile of the sample. Each of the four panels is for a specific critical ratio, and each box plot represents a heuristic with a sample size N. The normal distribution is a light-tailed distribution. In Section 2.3, we observed that the Scarf solution has a small expected profit loss under a light-tailed distribution. Hence, in Figure 7, we observe that when there is a sufficiently large sample size ($N = 40, 80$) to accurately estimate the partitioned moments, the relative errors of all of the robust models are relatively small. On the other hand, if the sample size is small ($N = 5, 10, 20$), the relative errors of the 1-partitioned and 2-partitioned models are higher than Scarf (1958). This is because of the high estimation errors of the partitioned moments for a small sample size. Note also that sample average approximation has comparable performance to the robust methods, except for a high critical ratio (0.99) with a small sample size. This is because SAA estimates a very high quantile using only the small sample. By contrast, the robust methods in this case are near optimal with relative errors of less than 5% even for a small sample size.

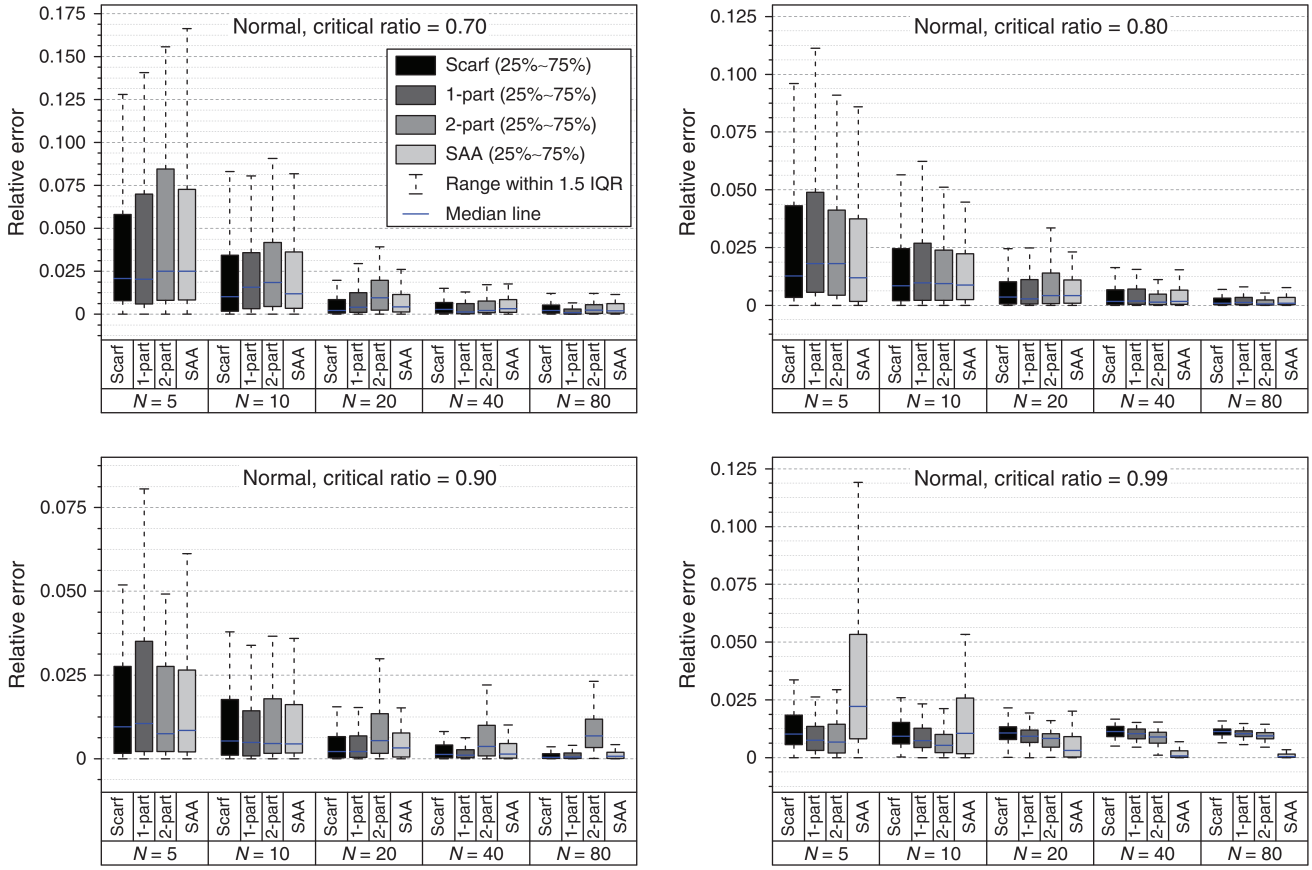
Figure 8 shows results of the same data-driven experiment when the sample is drawn from a heavy-tailed beta prime distribution with parameters specified in Table 1. In Section 2.3, we observed that additional semivariance information (i.e., the 1-part model) significantly reduces the expected profit loss when the demand distribution is heavy tailed. Hence, in Figure 8, we see that for critical ratio 0.7, 0.8, 0.9, the Scarf solution results in larger relative errors than the solutions from the 1-partition and 2-partition methods when the sample size is 10 or greater. When the sample size is equal to 5, then the robust methods have comparable relative errors because of the estimation errors of the 1-partition and 2-partition methods. When the critical ratio is 0.99, the optimal order quantity lies in a very high quantile. However, since the demand distribution is heavy tailed, estimating second-order statistics in the region of the optimal order quantity results in large errors when the sample size is small. Hence, for a very high critical ratio, the estimation errors have a significant effect on the relative errors especially if there are many demand partitions.

4. Modeling Asymmetry in the Multi-Item Newsvendor Problem Through Partitioned Statistics

In the previous two sections, we found that for a single-item newsvendor, having second-order partial statistics results in a computationally tractable distributionally robust problem while also resulting in a low expected profit ambiguity and low expected profit loss. In this section, we extend this insight to the distributionally robust multi-item newsvendor problem (1).

We start by considering the multi-item distributionally robust problem with only mean and covariance information (i.e., no asymmetry information). If the

Figure 7. (Color online) Box Plots Displaying the Errors of the Distributionally Robust Solutions with Partitioning and of SAA When the Models Are Trained with a Sample of Size N Drawn from a Light-Tailed Distribution (Normal)



Notes. SAA has significantly larger small-sample-size errors when the critical ratio is close to 1. For small sample sizes, 1-part and 2-part methods have larger errors than Scarf because of moment estimation errors. With a sufficiently large sample size, all the methods have comparably small errors. IQR, interquartile range.

mean $\mathbf{m} \in \mathbb{R}^n$ and covariance matrix $\boldsymbol{\Sigma} \in \mathcal{S}_+^n$ are available, then Hanasusanto et al. (2014) show that the distributionally robust problem (1) is NP-hard. Nevertheless, the problem can be reformulated as an SDP with $O(2^n)$ semidefinite inequality constraints, as stated in the following lemma.

Lemma 4.1. *Suppose that $\mathcal{F}$ is the set of all multivariate joint distribution of the n-dimensional random demand with mean $\mathbf{m}$ and covariance matrix $\boldsymbol{\Sigma} \succ 0$. Then (1) is equivalent to*

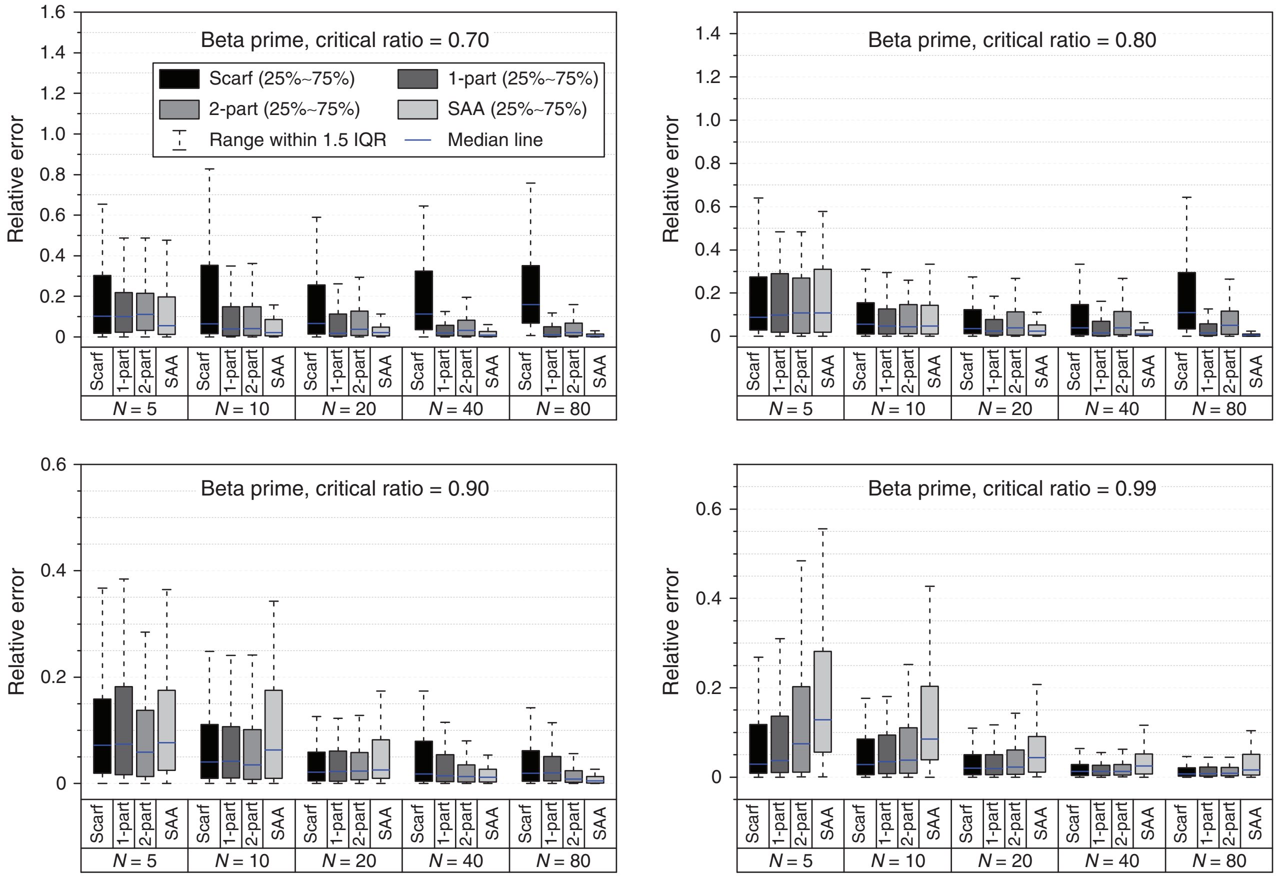
$$\begin{aligned} &\max_{\mathbf{q},\, y_0,\, \mathbf{y},\, \mathbf{Y}} \{\mathbf{p}^\top \mathbf{m} - y_0 - \mathbf{y}^\top \mathbf{m} - \mathbf{Y}\cdot\mathbf{Q} - \mathbf{c}^\top \mathbf{q}\} \\ &\text{s.t.} \begin{pmatrix} y_0 + \mathbf{p}_A^\top \mathbf{q} & \frac{1}{2}(\mathbf{y}-\mathbf{p}_A)^\top \\ \frac{1}{2}(\mathbf{y}-\mathbf{p}_A) & \mathbf{Y} \end{pmatrix} \succeq 0, \quad \forall A \in 2^{[n]}, \\ &\quad \mathbf{q} \in \mathcal{Q}, \quad y_0 \in \mathbb{R}, \quad \mathbf{y} \in \mathbb{R}^n, \quad \mathbf{Y} \in \mathcal{S}^n, \end{aligned}$$

where $\mathbf{Q} = \boldsymbol{\Sigma} + \mathbf{m}\mathbf{m}^\top$, and where we define $\mathbf{p}_A$ as a n-dimensional vector, with $p_{Ai} = p_i$ if $i \in A$ and $p_{Ai} = 0$ if $i \notin A$.

Proof. See Section A.6 in the electronic companion. □

The SDP can be solved using commercial off-the-shelf solvers for a small number of items; however, it is computationally intractable for a large number of items since the number of semidefinite inequality constraints is exponential in n. Alternatively, we can solve a lower bound to the problem by only considering the mean and variance information of the marginal distributions, which is equivalent to solving n separate Scarf (1958) problems. However, this does not take into account any joint distribution information; hence, the resulting solution might be very conservative. For instance, if the multiple demands are strongly correlated, then this information is ignored by solving independent Scarf (1958) problems that protect against the worst-case distribution of a conservatively large set of joint distributions, which includes all forms of correlations. Natarajan and Teo (2017) recently introduced a tighter lower bound on the worst-case expected profit in (1) with mean and covariance information, which involves solving an SDP with a single semidefinite inequality constraint.

Figure 8. (Color online) Box Plots Displaying the Errors of the Distributionally Robust Solutions with Partitioning and of SAA When the Models Are Trained with a Sample of Size N Drawn from a Heavy-Tailed Distribution (Beta Prime)



Notes. Similar to Figure 7, SAA has larger small-sample-size errors when the critical ratio is close to 1. Even if a small sample size results in estimation errors for 1-part and 2-part moments, in many cases these partitioned models still have smaller overall errors than Scarf. IQR, interquartile range.

Lemma 4.2 (Natarajan and Teo 2017). *Suppose that $\mathcal{F}$ is the set of all multivariate joint distribution of the n-dimensional random demand with mean* $\mathbf{m}$ *and covariance matrix* $\boldsymbol{\Sigma} \succ 0$. *Then a lower bound on the optimal value of* (1) *is*

$$\begin{aligned}
&\max\ \{\mathbf{p}^\top\mathbf{m} - y_0 - \mathbf{e}^\top\mathbf{S}\mathbf{e} - \mathbf{m}^\top\mathbf{y} - \mathbf{Q}\cdot\mathbf{W} - \mathbf{c}^\top\mathbf{q}\}\\
&\text{s.t.}\ \begin{pmatrix} \mathbf{W} & \frac{1}{2}\mathbf{y} & -\frac{1}{2}\mathrm{diag}(\mathbf{p})\\ \frac{1}{2}\mathbf{y}^\top & y_0 & C_{23}\\ -\frac{1}{2}\mathrm{diag}(\mathbf{p}) & C_{32} & \frac{1}{2}(\mathbf{R}+\mathbf{R}^\top)-\mathbf{S}\end{pmatrix} \succeq 0,\\
&\quad C_{32} = \mathbf{S}\mathbf{e} + \tfrac{1}{2}(\mathbf{p}\circ(\mathbf{q}-\mathbf{m}) - \mathbf{R}^\top\mathbf{e}),\\
&\quad C_{23} = \mathbf{e}^\top\mathbf{S} + \tfrac{1}{2}(\mathbf{p}\circ(\mathbf{q}-\mathbf{m}) - \mathbf{R}^\top\mathbf{e})^\top,\\
&\quad \mathbf{R}\geq \mathbf{0},\quad \mathbf{S}\geq\mathbf{0},\quad \mathbf{S}=\mathbf{S}^\top,\quad \mathbf{W},\mathbf{R},\mathbf{S}\in\mathcal{S}^n,\\
&\quad \mathbf{q}\in\mathcal{Q},\quad y_0\in\mathbb{R},\quad \mathbf{y}\in\mathbb{R}^n,
\end{aligned}$$

where $\mathbf{Q} = \boldsymbol{\Sigma} + \mathbf{m}\mathbf{m}^\top$ *and* $\circ$ *is the Hadamard product operator.*

The details of the proof of Lemma 4.2 can be found in Natarajan and Teo (2017). We modify their approach to find a lower bound on the distributionally robust problem (1) with second-order partitioned statistics information. Suppose that we partition the random demand vector $\tilde{\mathbf{d}}$ above and below the mean vector $E(\tilde{\mathbf{d}}) = \mathbf{m}$. In particular, define the following partitioned random variables:

$$\tilde{\mathbf{z}}_+ = (\tilde{\mathbf{d}} - \mathbf{m})^+,$$
$$\tilde{\mathbf{z}}_- = (\mathbf{m} - \tilde{\mathbf{d}})^+.$$

Clearly, $E(\tilde{\mathbf{z}}_+) = E(\tilde{\mathbf{z}}_-)$, since we partition at the mean vector, and hence

$$\begin{aligned} E(\tilde{\mathbf{z}}_+) - E(\tilde{\mathbf{z}}_-) &= E(\tilde{\mathbf{d}} - \mathbf{m})\\ &= 0.\end{aligned}$$

Note that the inner objective of (1) can be rewritten in terms of $\tilde{\mathbf{z}}^\top = (\tilde{\mathbf{z}}_+^\top, \tilde{\mathbf{z}}_-^\top)$ as

$$\mathbf{p}^\top\mathbf{m} - E_f\left(\sum_{i=1}^n p_i(\tilde{z}_{+i} - \tilde{z}_{-i} - q_i + m_i)^+\right) - \mathbf{c}^\top\mathbf{q}.$$

Suppose that there is partial information known about the joint distribution of partitioned moments $(\tilde{\mathbf{z}}_+^\top, \tilde{\mathbf{z}}_-^\top)$. In particular, suppose that

$$E\left(\begin{pmatrix}\tilde{\mathbf{z}}_+\\ \tilde{\mathbf{z}}_-\end{pmatrix}\right) = \begin{pmatrix}\bar{\mathbf{m}}\\ \bar{\mathbf{m}}\end{pmatrix}, \tag{9}$$

$$E\left(\begin{pmatrix}\tilde{\mathbf{z}}_+\\ \tilde{\mathbf{z}}_-\end{pmatrix}\begin{pmatrix}\tilde{\mathbf{z}}_+\\ \tilde{\mathbf{z}}_-\end{pmatrix}^\top\right)=\begin{pmatrix}\bar{\mathbf{Q}}_+ & \bar{\mathbf{Q}}_0^\top\\ \bar{\mathbf{Q}}_0 & \bar{\mathbf{Q}}_-\end{pmatrix}. \tag{10}$$

Then the problem of choosing the worst-case distribution with the partitioned information is equivalent to solving

$$\begin{aligned}&\sup_{f\in\mathcal{F}} E_f\left(\sum_{i=1}^n p_i(\tilde z_{+i}-\tilde z_{-i}-q_i+m_i)^+\right)\\ &\quad=\sup_{f\in\mathcal{F}} E_f\left(\max_{\hat{\mathbf{x}}\in\{0,1\}^n}\sum_{i=1}^n p_i(\tilde z_{+i}-\tilde z_{-i}-q_i+m_i)\hat x_i\right) \qquad (11)\\ &\quad=\sup_{f\in\mathcal{F}} E_f\left(\sum_{i=1}^n p_i(\tilde z_{+i}-\tilde z_{-i}-q_i+m_i)\hat x_i^*(\tilde{\mathbf{z}})\right), \qquad (12)\end{aligned}$$

where $\hat{\mathbf{x}}^*(\mathbf{z})$ is the maximizer in (11) for any realization $\mathbf{z}^\top\in\mathbb{R}_+^{2n}$ of the random vector $(\tilde{\mathbf{z}}_+^\top,\tilde{\mathbf{z}}_-^\top)$. We introduce the decision variables:

$$\begin{aligned}\mathbf{x}&=E(\hat{\mathbf{x}}^*(\tilde{\mathbf{z}})),\\ \mathbf{Y}_+&=E(\hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\tilde{\mathbf{z}}_+^\top),\\ \mathbf{Y}_-&=E(\hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\tilde{\mathbf{z}}_-^\top),\\ \mathbf{X}&=E(\hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\hat{\mathbf{x}}(\tilde{\mathbf{z}})^\top).\end{aligned}$$

In terms of these variables, the objective function can be expressed as

$$\begin{aligned}&E_f\left(\sum_{i=1}^n p_i(\tilde z_{+i}-\tilde z_{-i}-q_i+m_i)^+\right)\\ &\quad=\langle \operatorname{diag}(\mathbf{p}),\mathbf{Y}_+-\mathbf{Y}_-\rangle-(\mathbf{p}\circ(\mathbf{q}-\mathbf{m}))^\top\mathbf{x}. \qquad (13)\end{aligned}$$

Furthermore, the following random matrix is positive semidefinite for any realization since it is a rank-one matrix:

$$\begin{pmatrix}\tilde{\mathbf{z}}_+\tilde{\mathbf{z}}_+^\top & \tilde{\mathbf{z}}_+\tilde{\mathbf{z}}_-^\top & \tilde{\mathbf{z}}_+ & \tilde{\mathbf{z}}_+\hat{\mathbf{x}}^*(\tilde{\mathbf{z}})^\top\\ \tilde{\mathbf{z}}_-\tilde{\mathbf{z}}_+^\top & \tilde{\mathbf{z}}_-\tilde{\mathbf{z}}_-^\top & \tilde{\mathbf{z}}_- & \tilde{\mathbf{z}}_-\hat{\mathbf{x}}^*(\tilde{\mathbf{z}})^\top\\ \tilde{\mathbf{z}}_+^\top & \tilde{\mathbf{z}}_-^\top & 1 & \hat{\mathbf{x}}^*(\tilde{\mathbf{z}})^\top\\ \hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\tilde{\mathbf{z}}_+^\top & \hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\tilde{\mathbf{z}}_-^\top & \hat{\mathbf{x}}^*(\tilde{\mathbf{z}}) & \hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\hat{\mathbf{x}}^*(\tilde{\mathbf{z}})^\top\end{pmatrix}=\begin{pmatrix}\tilde{\mathbf{z}}_+\\ \tilde{\mathbf{z}}_-\\ 1\\ \hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\end{pmatrix}\begin{pmatrix}\tilde{\mathbf{z}}_+\\ \tilde{\mathbf{z}}_-\\ 1\\ \hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\end{pmatrix}^\top\succeq 0.$$

Taking expectations hence implies that the following matrix is positive semidefinite:

$$\begin{pmatrix}\bar{\mathbf{Q}}_+ & \bar{\mathbf{Q}}_0^\top & \bar{\mathbf{m}} & \mathbf{Y}_+^\top\\ \bar{\mathbf{Q}}_0 & \bar{\mathbf{Q}}_- & \bar{\mathbf{m}} & \mathbf{Y}_-^\top\\ \bar{\mathbf{m}}^\top & \bar{\mathbf{m}}^\top & 1 & \mathbf{x}^\top\\ \mathbf{Y}_+ & \mathbf{Y}_- & \mathbf{x} & \mathbf{X}\end{pmatrix}=\mathrm{E}\left(\begin{pmatrix}\tilde{\mathbf{z}}_+\\ \tilde{\mathbf{z}}_-\\ 1\\ \hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\end{pmatrix}\begin{pmatrix}\tilde{\mathbf{z}}_+\\ \tilde{\mathbf{z}}_-\\ 1\\ \hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\end{pmatrix}^\top\right)\succeq 0. \tag{14}$$

Furthermore, since $\hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\in\{0,1\}^n$, we have

$$\begin{aligned}\begin{pmatrix}1 & \mathbf{x}^\top\\ \mathbf{x} & \mathbf{X}\end{pmatrix}&=\mathrm{E}\left(\begin{pmatrix}1\\ \hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\end{pmatrix}\begin{pmatrix}1\\ \hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\end{pmatrix}^\top\right)\\ &\in \operatorname{conv}\left\{\begin{pmatrix}1\\ \hat{\mathbf{x}}\end{pmatrix}\begin{pmatrix}1\\ \hat{\mathbf{x}}\end{pmatrix}^\top \,\middle|\, \hat{\mathbf{x}}\in\{0,1\}^n\right\}. \qquad (15)\end{aligned}$$

Finally, since $\tilde{\mathbf{z}}_+\ge 0$, $\tilde{\mathbf{z}}_-\ge 0$, and $\hat{\mathbf{x}}^*(\tilde{\mathbf{z}})\ge 0$, we have

$$\mathbf{Y}_+,\mathbf{Y}_-\ge 0. \tag{16}$$

We can thus define an upper bound to the moment problem (11) by using the objective function (13) and enforcing the set of necessary conditions (14)–(16) as constraints that the decision variables must satisfy:

$$\begin{aligned}\max_{\mathbf{Y}_+,\mathbf{Y}_-,\mathbf{x},\mathbf{X}}\quad & \{\langle \operatorname{diag}(\mathbf{p}),\mathbf{Y}_+-\mathbf{Y}_-\rangle-(\mathbf{p}\circ(\mathbf{q}-\mathbf{m}))^\top\mathbf{x}\}\\ \text{subject to}\quad & \begin{pmatrix}\bar{\mathbf{Q}}_+ & \bar{\mathbf{Q}}_0^\top & \bar{\mathbf{m}} & \mathbf{Y}_+^\top\\ \bar{\mathbf{Q}}_0 & \bar{\mathbf{Q}}_- & \bar{\mathbf{m}} & \mathbf{Y}_-^\top\\ \bar{\mathbf{m}}^\top & \bar{\mathbf{m}}^\top & 1 & \mathbf{x}^\top\\ \mathbf{Y}_+ & \mathbf{Y}_- & \mathbf{x} & \mathbf{X}\end{pmatrix}\succeq 0,\\ & \begin{pmatrix}1 & \mathbf{x}^\top\\ \mathbf{x} & \mathbf{X}\end{pmatrix}\in \mathrm{BQP}_n\\ &\qquad = \operatorname{conv}\left\{\begin{pmatrix}1\\ \hat{\mathbf{x}}\end{pmatrix}\begin{pmatrix}1\\ \hat{\mathbf{x}}\end{pmatrix}^\top \,\middle|\, \hat{\mathbf{x}}\in\{0,1\}^n\right\},\\ & \mathbf{Y}_+\ge\mathbf{0},\quad \mathbf{Y}_-\ge\mathbf{0},\\ & \mathbf{Y}_+,\mathbf{Y}_-\in\mathbb{R}^{n\times n},\quad \mathbf{x}\in\mathbb{R}^n,\quad \mathbf{X}\in\mathbb{R}^{n\times n}. \qquad (17)\end{aligned}$$

Note that BQP_n is the Boolean quadratic polytope, for which Natarajan and Teo (2017) consider a relaxation:

$$\left\{\begin{pmatrix}1\\ \mathbf{x}\end{pmatrix}\begin{pmatrix}1\\ \mathbf{x}\end{pmatrix}^\top\in\mathcal{S}^{n+1}\,\middle|\,\begin{array}{l}X_{ii}=x_i,\quad \forall\, i\\ X_{ij}\le x_i,\quad \forall\, i\le j\\ X_{ij}\ge x_i+x_j-1,\quad \forall\, i\le j\\ X_{ij}\ge 0,\quad \forall\, i\le j\\ X_{ij}\le x_j,\quad i\le j\end{array}\right\}.$$

Hence, replacing the Boolean quadratic constraint in (17) by these linear constraints results in an SDP whose optimal value is an upper bound for (17). By taking the dual of this semidefinite program, we have a tractable relaxation for the distributionally robust problem (1) with partitioned statistics information, given in the following theorem.

Theorem 4.3. *Suppose that $\mathcal{F}$ is the set of all multivariate joint distributions of the n-dimensional random demand with mean $\mathbf{m}$ and with the mean and covariance of the partitioned demand given in (9) and (10). Then a lower bound on the optimal value of (1) is*

$$\begin{aligned}\max\quad & \Bigg\{\mathbf{p}^\top\mathbf{m}-y_0-\mathbf{e}^\top\mathbf{S}\mathbf{e}-\bar{\mathbf{m}}^\top\mathbf{y}_+-\bar{\mathbf{m}}^\top\mathbf{y}_-\\ &\quad-\begin{pmatrix}\bar{\mathbf{Q}}_+ & \bar{\mathbf{Q}}_0^\top\\ \bar{\mathbf{Q}}_0 & \bar{\mathbf{Q}}_-\end{pmatrix}\cdot\begin{pmatrix}\mathbf{W}_+ & \mathbf{W}_0^\top\\ \mathbf{W}_0 & \mathbf{W}_-\end{pmatrix}-\mathbf{c}^\top\mathbf{q}\Bigg\}\\ \text{s.t.}\quad & \begin{pmatrix}\mathbf{W}_+ & \mathbf{W}_0^\top & \frac{1}{2}\mathbf{y}_+ & \mathbf{Z}_+^\top\\ \mathbf{W}_0 & \mathbf{W}_- & \frac{1}{2}\mathbf{y}_- & \mathbf{Z}_-^\top\\ \frac{1}{2}\mathbf{y}_+^\top & \frac{1}{2}\mathbf{y}_-^\top & y_0 & \mathbf{v}^\top\\ \mathbf{Z}_+ & \mathbf{Z}_- & \mathbf{v} & \mathbf{U}\end{pmatrix}\succeq 0,\\ & \mathbf{Z}_+\le-\tfrac{1}{2}\operatorname{diag}(\mathbf{p}),\end{aligned}$$

$$\begin{aligned}
&\mathbf{Z}_- \leq \tfrac{1}{2}\mathrm{diag}(\mathbf{p}),\\
&\mathbf{v} = -\tfrac{1}{2}\mathbf{R}^\top\mathbf{e} + \mathbf{S}^\top\mathbf{e} + \tfrac{1}{2}\mathbf{p}\circ(\mathbf{q}-\mathbf{m}),\\
&\mathbf{U} \leq \tfrac{1}{2}(\mathbf{R}+\mathbf{R}^\top) - \mathbf{S},\\
&\mathbf{R} \geq \mathbf{0}, \quad \mathbf{S} \geq \mathbf{0}, \quad \mathbf{S} = \mathbf{S}^\top,\\
&\mathbf{q} \in \mathcal{Q}, \quad y_0 \in \mathbb{R}, \quad \mathbf{y}_+, \mathbf{y}_-, \mathbf{v} \in \mathbb{R}^n,\\
&\mathbf{W}_0, \mathbf{W}_+, \mathbf{W}_-, \mathbf{R}, \mathbf{S}, \mathbf{Z}_+, \mathbf{Z}_-, \mathbf{U} \in \mathcal{S}^n,
\end{aligned} \tag{18}$$

where we define $\mathbf{e}$ *as the n-dimensional vector of all ones.*

We next conduct simulation experiments to compare the EVAI with partitioned moment information against other partial information models. Recall that the EVAI value represents the maximum willingness to pay for complete knowledge of the joint demand distribution. We scale the EVAI value by the optimal expected profit for a given distribution to compute the relative EVAI. In these experiments, we compare the relative EVAI of four models: mean and variance of the marginal demand distributions (Scarf); mean, variance, and semivariance of the marginal demand distributions (MVS Marginal); mean and covariance of the joint demand distribution (MV Joint); and mean and covariance of the joint partitioned demand distribution (MVS Joint). The MVS Marginal model is obtained by using the single item bounds derived in Section 2 with semivariance information. The MVS Joint model can be interpreted as a 1-partition model in multiple dimensions and is approximated using Theorem 4.3. The difference between the relative EVAI of the Scarf model and, say, the relative EVAI of the MV Joint model can be interpreted as the value of additional covariance information.

We describe the methodology to generate each problem instance in the experiment next. First, we randomly choose an integer n between 10 and 30 as the number of items. We assume that the unit price for all n items is \$100. For each item $i \in [n]$, we set the unit cost $c_i = (1-\beta_i) \times \$100$, where β_i is the critical ratio that is randomly drawn from $(0.9, 1.0)$. We model the multivariate joint distribution in terms of n univariate marginal distribution functions and a Gaussian copula that describes the dependence structure. We randomly generate parameters of the marginal demand distributions from the range of parameters specified in Table 3. We also randomly generate a $n \times n$ correlation matrix for the Gaussian copula, where the correlation coefficients are uniformly distributed within boundaries that are sequentially computed (Numpacharoen and Atsawarungruangkit 2012). We determine the mean vector and covariance matrix of the random demand based on the randomly generated parameters. We estimate the mean and covariance of the partitioned demand based on 500,000 samples drawn from the multivariate joint distribution. On the basis of the mean and covariance of the demand, we use Lemma 4.2 to find the order quantity that maximizes the lower bound using the results of Natarajan and Teo (2017) for the MV Joint problem. Similarly, on the basis of the estimated mean and covariance of the partitioned demand, we find the order quantity that maximizes the bound in SDP (18), which gives a lower bound to the MVS Joint problem. We also solve n individual Scarf (1958) problems using the marginal means and variances and n individual problems using the marginal means, variances, and semivariances. For each of these robust solutions, we compute the EVAI by finding the optimal expected profit minus the expected profit of the robust solution. Figure 9 shows plots of the relative EVAI values for the four different models (MV Marginal (Scarf), MVS Marginal, MV Joint, or MVS Joint) in randomly generated problem instances. Each panel in the figure represents a different distribution family for the marginal demand distributions (exponential, normal, Student's t, and beta prime). The figure shows that the information on the semivariance and information on the partitioned multivariate demand significantly reduces the EVAI as compared to MV Joint, particularly for the Student's t and beta prime distributions. For the normal and exponential distributions, the MVS Joint model provides improvement on the MVS Marginal model.

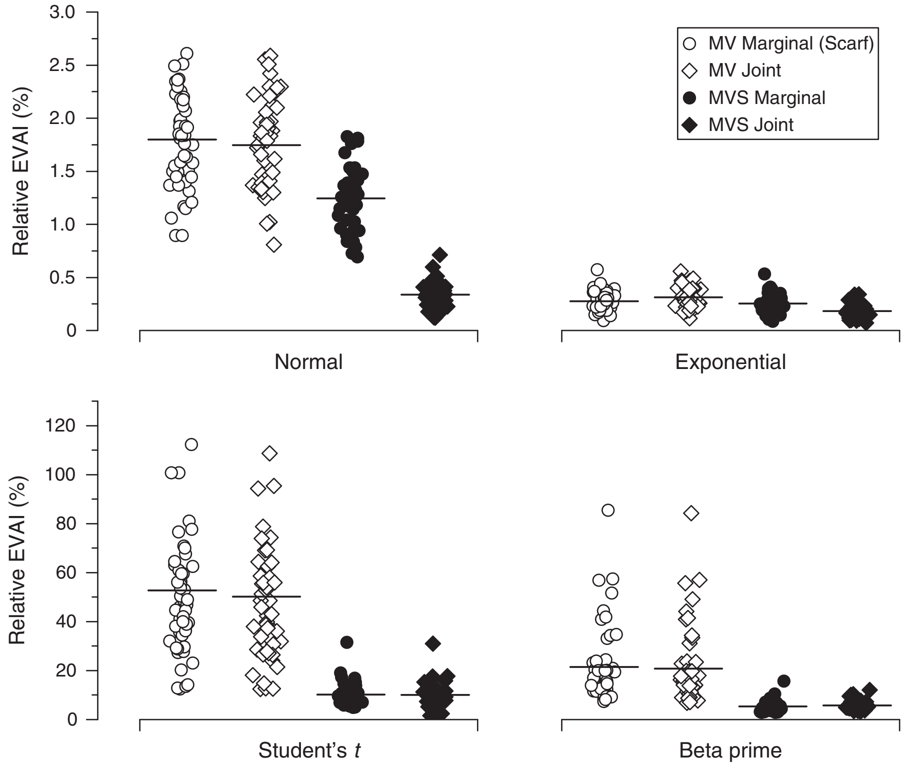
Table 3. Marginal Demand Distributions Used in the Multi-Item Experiments

Distribution	Range from which parameters are drawn
Normal	$m \in (100, 500)$, $\sigma \in (100, 800)$
Exponential	$m \in (100, 500)$
Student's t	$\nu \in (2, 3)$
Beta prime	$\alpha \in (0, 5)$, $\beta \in (2, 3)$

5. Computational Experiments: Automotive Industry

In this section, we conduct computational experiments to demonstrate in a real-world manufacturing setting the effectiveness of using the distributionally robust approach with partitioned statistics information. For these experiments, a European automotive manufacturer provided us with daily demand data over almost one year (250 weekdays) for 36 spare parts' SKUs. Manufacturing spare parts is an important function for this automotive manufacturer, since an aging vehicle fleet in Europe and North America has led to a higher spare part volume, with dealers placing orders for spare parts daily. Spare parts are necessary to prevent operational delays when a vehicle breaks down; therefore, spare parts must be available when needed. But, on the other hand, the carrying cost of spare parts can be significant. We chose this industry for the setting of

Figure 9. Scatter Plot Displaying Relative EVAI Values for the MV Marginal (Scarf), MVS Marginal, MV Joint, and MVS Joint Model for Randomly Generated Problem Instances



Notes. For many of these instances, the MV Marginal and MV Joint models are comparable in terms of relative EVAI. For a majority of the instances, asymmetry information through partitioned statistics results in a significant reduction in relative EVAI.

our computational experiments because of the inherent difficulty in forecasting spare parts demand since it is irregular and intermittent. Our goals in this section are twofold: (i) to show that second-order partitioned moments of a multivariate joint demand distribution can be estimated using data in a real-world setting and (ii) to compare the out-of-sample cost of distributionally robust solutions based on estimated moments against the sample average approximation solution.

5.1. Data and Methodology

Our data set consists of daily demand data for 36 spare parts' SKUs for almost one year (250 weekdays). Table 4 displays the summary statistics (mean, coefficient of variation, skewness, and normalized semivariance) of the daily demand data of the 36 spare parts' SKUs. The average daily demand ranges from 43 units to almost 95,000 units. The coefficient of variation ranges from 0.25 to almost 1.5. Almost three-quarters of the SKUs exhibit positive skewness, with the remainder having negative skewness. In the table, we also display the normalized semivariance of the 36 SKUs. Note that the skewness and the normalized semivariance have the same sign in all of the demand data. Based on the summary statistics, the marginal demand distributions all exhibit variability and asymmetry. Hence, a practice commonly used in other industries of fitting the data to a normal distribution is inappropriate. In fact, when we fit the demand data to as many as 17 distribution families, almost all of the SKUs have the best fit (i.e., have the smallest Bayesian information criterion) for heavy-tailed distribution families (generalized Pareto, generalized extreme value, and t location-scale).

We next describe the setting of the computational experiments. The automotive manufacturer needs to determine the number of units to manufacture for each SKU to meet the external demand from its customers (e.g., dealers). Each day, the manufacturer begins production before the customers' orders arrive. Hence, the manufacturer solves a newsvendor problem to determine the daily target inventory levels; $\mathbf{q} = (q_1, \ldots, q_{36})$. Suppose that the manufacturer knew the true multivariate joint demand distribution f of the random daily demand $\tilde{\mathbf{d}} = (\tilde{d}_1, \ldots, \tilde{d}_{36})$. If b_i is the per-unit per-day backorder cost for SKU i and if h_i is the per-unit per-day holding cost for SKU i, then the manufacturer needs to solve the following multi-item newsvendor problem:

$$\min_{q \in \mathcal{Q}} E_f\left(\sum_{i=1}^{36} b_i(\tilde{d}_i - q_i)^+ + h_i(q_i - \tilde{d}_i)^+\right). \tag{19}$$

Table 4. Descriptive Statistics (Mean, Coefficient of Variation, Skewness, and Normalized Semivariance) of the Daily Demand Data for the 36 Spare Parts' SKUs

SKU	Mean	Coeff. of variation	Skewness	Norm. semivar.	SKU	Mean	Coeff. of variation	Skewness	Norm. semivar.
1	43	0.94	2.19	0.458	19	10,761	0.80	0.22	0.027
2	10,471	1.47	1.63	0.503	20	30,797	0.66	0.18	0.021
3	22,661	0.99	1.14	0.338	21	47,763	0.65	0.16	0.043
4	17,990	1.15	0.94	0.294	22	10,458	0.84	0.13	0.045
5	17,873	1.12	0.90	0.306	23	209	0.41	0.13	0.001
6	1,123	0.41	0.70	0.247	24	2,790	0.61	0.12	0.004
7	10,736	0.89	0.66	0.153	25	44,161	0.54	0.04	0.038
8	69,216	0.60	0.58	0.170	26	53,902	0.58	0.02	0.017
9	30,772	0.87	0.45	0.083	27	23,951	0.78	−0.03	−0.034
10	23,440	0.79	0.43	0.086	28	43,287	0.50	−0.04	−0.021
11	1,990	0.37	0.41	0.183	29	351	0.54	−0.15	−0.068
12	52	0.52	0.37	0.094	30	94,636	0.46	−0.40	−0.117
13	27,409	0.72	0.36	0.069	31	7,997	0.54	−0.42	−0.168
14	55,617	0.72	0.32	0.092	32	64	0.41	−0.66	−0.197
15	64,019	0.64	0.32	0.107	33	72,586	0.39	−0.81	−0.256
16	44,332	0.78	0.28	0.076	34	303	0.29	−0.84	−0.162
17	39,355	0.62	0.24	0.011	35	3,820	0.25	−1.68	−0.383
18	60,746	0.70	0.23	0.063	36	193	0.28	−1.98	−0.475

Note. Note that the skewness and normalized semivariance have the same sign in all the data.

The objective is to minimize the total cost, which reduces to the profit maximization version by setting $p_i = b_i + h_i$ and $c_i = h_i$. The automotive manufacturer did not provide us with cost parameters since it is considered proprietary information. Hence, for the purpose of our study, we replicate the experiment on 50 problem instances, where in each instance we randomly generate new cost parameters. If there are no joint ordering constraints (i.e., $\mathcal{Q} = \mathbb{R}^n$), then problem (19) is decomposable by item, and the optimal quantity q_i is equal to the $b_i/(b_i + h_i)$ quantile of the marginal distribution of the demand for SKU i.

Suppose now that the spare parts manufacturer does not know the true joint distribution f of the random demand vector. In practice, the only information available to the manufacturer is a data sample $\{\hat{\mathbf{d}}^1, \hat{\mathbf{d}}^2, \ldots, \hat{\mathbf{d}}^N\}$ of the daily demand, where each sample point can be assumed to be drawn from the unknown joint distribution f. In our experiments, we assume that this data sample is the N most recent historical daily demand. One well-known approach, if given a sample from the unknown true distribution, is to solve the sample average approximation counterpart (Levi et al. 2015):

$$\min_{q\in\mathcal{Q}} \frac{1}{N} \sum_{k=1}^{N} \left(\sum_{i=1}^{36} b_i(\hat{d}_i^k - q_i)^+ + h_i(q_i - \hat{d}_i^k)^+ \right). \tag{20}$$

However, if the sample size for the demand data is small, then the empirical distribution might be a poor approximation of the true joint demand distribution. An alternative approach is to use the demand data sample to estimate moments of the joint distribution and to use these estimated moments in a distributionally robust newsvendor problem:

$$\min_{q\in\mathcal{Q}} \sup_{f\in\hat{\mathcal{F}}_N} E_f\left(\sum_{i=1}^{36} b_i(\tilde{d}_i - q_i)^+ + h_i(q_i - \tilde{d}_i)^+ \right), \tag{21}$$

where $\hat{\mathcal{F}}_N$ is the set of all joint distributions with moments equal to the estimated moments from the sample with size N. If the distribution family consists of information on the joint distribution, then (21) is not separable by item even if there are no joint constraints. On the other hand, if $\hat{\mathcal{F}}_N$ only has information on the marginal distributions, then the problem is separable by item. We observed from the data that there is significant information contained in the joint distribution that would be lost if we only estimate marginal distribution information. This can be seen in Figures 1 and 2, which show the scatter plot matrix of demand for three specific SKUs and the scatter plot matrix of the corresponding partitioned demand. From these figures, we observe that the demand for SKU pairs is strongly positively correlated. This strongly positive correlation is also seen in the partitioned demands of the SKUs. Notice that pairs of positive partitions are strongly positively correlated. Similarly, several pairs of negative partitions are also positively correlated.

Hence, because of the observed correlation in the demand of several SKUs, we chose to test in our experiments the performance of the distributionally robust solution to (21) with mean and covariance information estimated from the data sample. That is, we use the data sample $\{\hat{\mathbf{d}}^1, \ldots, \hat{\mathbf{d}}^N\}$ to estimate the mean and covariance of the joint demand distribution:

$$\mathbf{m}_N = \frac{1}{N}\sum_{k=1}^{N}\hat{\mathbf{d}}^k, \tag{22}$$

$$\mathbf{Q}_N = \frac{1}{N}\sum_{k=1}^{N}\hat{\mathbf{d}}^k\hat{\mathbf{d}}^{k\mathsf{T}}. \tag{23}$$

Because of the large number of SKUs ($n = 36$), solving the resulting distributionally robust problem using Lemma 4.1 is computationally intractable since it involves solving an SDP with 2^{36} semidefinite inequality constraints. Instead, we approximate the solution by solving the SDP lower bound reformulation by Natarajan and Teo (2017) (see Lemma 4.2).

Finally, we also test the performance of the distributionally robust solution to (21) with partitioned statistics information. That is, we estimate the following partitioned moments using the demand sample:

$$\bar{\mathbf{m}}_N = \frac{1}{2N}\sum_{k=1}^{N}(\hat{\mathbf{d}}^k - \mathbf{m}_N)^+ + \frac{1}{2N}\sum_{k=1}^{N}(\mathbf{m}_N - \hat{\mathbf{d}}^k)^+, \tag{24}$$

$$\bar{\mathbf{Q}}_{+N} = \frac{1}{N}\sum_{k=1}^{N}(\hat{\mathbf{d}}^k - \mathbf{m}_N)^+(\hat{\mathbf{d}}^k - \mathbf{m}_N)^{+\mathsf{T}}, \tag{25}$$

$$\bar{\mathbf{Q}}_{-N} = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{m}_N - \hat{\mathbf{d}}^k)^+(\mathbf{m}_N - \hat{\mathbf{d}}^k)^{+\mathsf{T}}, \tag{26}$$

$$\bar{\mathbf{Q}}_{0N} = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{m}_N - \hat{\mathbf{d}}^k)^+(\hat{\mathbf{d}}^k - \mathbf{m}_N)^{+\mathsf{T}}. \tag{27}$$

We then approximate the solution to the distributionally robust problem by solving the SDP given in Theorem 4.3.

Since backorder and holding cost parameter data were not available to us, we conducted the same experiments on 50 randomly generated problem instances. For each problem instance, we set the backorder cost $b_i = \$1$ and randomly generate critical ratios $\beta_i \in (0.80, 1)$, where $h_i = b_i(1-\beta_i)/\beta_i$ for $i = 1, \ldots, 36$. In what follows, we describe the methodology for conducting the experiments on each problem instance. We aimed to conduct the computational experiments to simulate implementation in a real-world setting. Therefore, we chose a rolling horizon implementation where daily inventory target levels are recalculated every 21 weekdays using one of the three methods (SAA, MV Joint, and MVS Joint). The models solved by each method are trained with the daily demand data from the most recent 62 weekdays (i.e., $N = 62$).

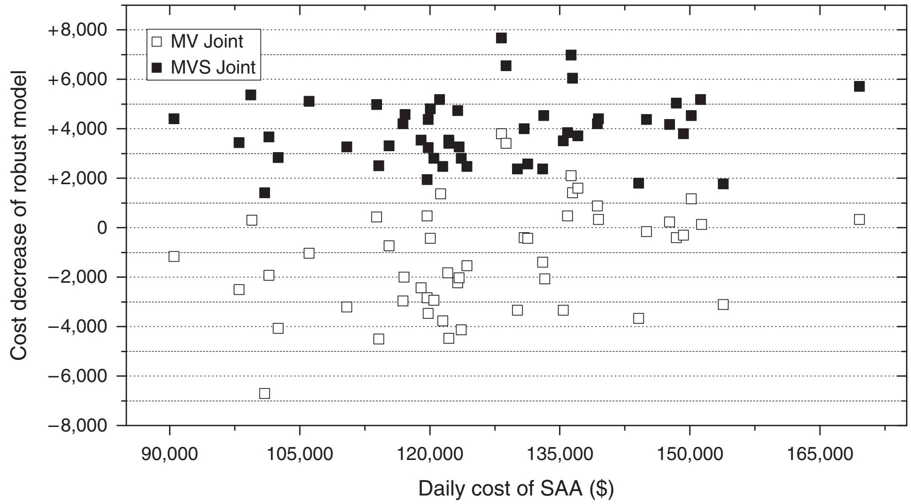
We start the experiments at the beginning of day 63 with zero starting inventory and determine daily target inventory levels for the 36 SKUs based on the three methods described earlier (SAA, MV Joint, and MVS Joint). The manufacturer then starts the production of spare parts so that the inventory levels match the targets. At the end of the day, demand is realized based on the actual day 63 demand, and inventories are depleted. Backorder costs or holding costs are incurred depending on whether the inventory position is negative or positive. Starting from day 64 until before the next time target inventory levels are recalculated, we set starting day inventory levels to match the targets, deplete inventory based on actual day demand, and incur backorder cost or holding cost. After 21 days, the daily target inventory levels are recalculated according to the three methods based on the most recent $N = 62$ demand data and implemented for the next 21 days. We repeat the recalibration-implementation cycle until the simulation reaches the end of day 250. The total cost (sum of holding and backorder costs) incurred by each of the three methods is finally recorded.

5.2. Results and Discussion

Figure 10 summarizes the final results from the simulation experiments described in this section. The figure shows a scatter plot of the SAA total cost per day against the reduction in daily cost by the robust model. The figure shows us, for example, that in the problem instance where SAA incurred an average total cost of about \$90,000 per day, the MV Joint model resulted in a total cost *increase* equal to about \$1,000 per day, and the MVS Joint model resulted in a total cost *decrease* of about \$4,500 per day.

Note that to compute the target inventory levels, the three methods use a small sample size ($N = 62$) drawn from a 36-dimensional joint demand distribution of past demands. The sample average approximation method assumes that the joint distribution of future demands matches the empirical distribution based on the 63 data samples. Hence, if the sample size is small, then sample average approximation can misspecify the distribution, and the SAA solution might incur a high cost under the true distribution of future demands. A distributionally robust approach protects against adverse effects of distribution misspecification by choosing inventory levels that perform well under a family of distributions. However, we observe that for the majority of the problem instances shown in Figure 10, the distributionally robust method with only mean and covariance information incurs a higher cost than SAA. As we discussed earlier in this paper, mean and covariance do not provide information about the asymmetry of the joint distribution. Hence, the MV Joint distribution set includes a large number of distributions with different asymmetry forms, resulting in a conservative solution that incurs a higher cost than SAA. On the other hand, we observe that in all of the problem instances shown in Figure 10, the distributionally robust solution based on second-order partitioned moments incurs a lower total cost than the SAA cost. By including asymmetry information, the MVS Joint model protects against a smaller set of candidate distributions while still achieving a robust solution that

Figure 10. Scatter Plot Displaying the Out-of-Sample Daily Cost Incurred from the SAA-Based Inventory Policy Against the Cost Reduction Achieved from a Distributionally Robust Inventory Policy in Randomly Generated Problem Instances, Where the Out-of-Sample Costs Are Computed Based on a Rolling Horizon Implementation on the Spare Parts Demand Data



Notes. The dotted line is the zero-reduction reference line. The plot shows that in the majority of instances, the distributionally robust model with mean and covariance information results in a cost increase. In all instances, information on mean and covariance of partitioned demand results in positive daily cost reduction.

performs well under the true distribution of future demands.

We further investigate the differences of the robust methods in the cost components (holding cost and backorder cost). The following table shows the daily cost components averaged over the 50 instances:

Method	Average holding cost per day ($)	Average backorder cost per day ($)
SAA	90,056	36,416
MV Joint	88,650	39,056
MVS Joint	87,572	34,956

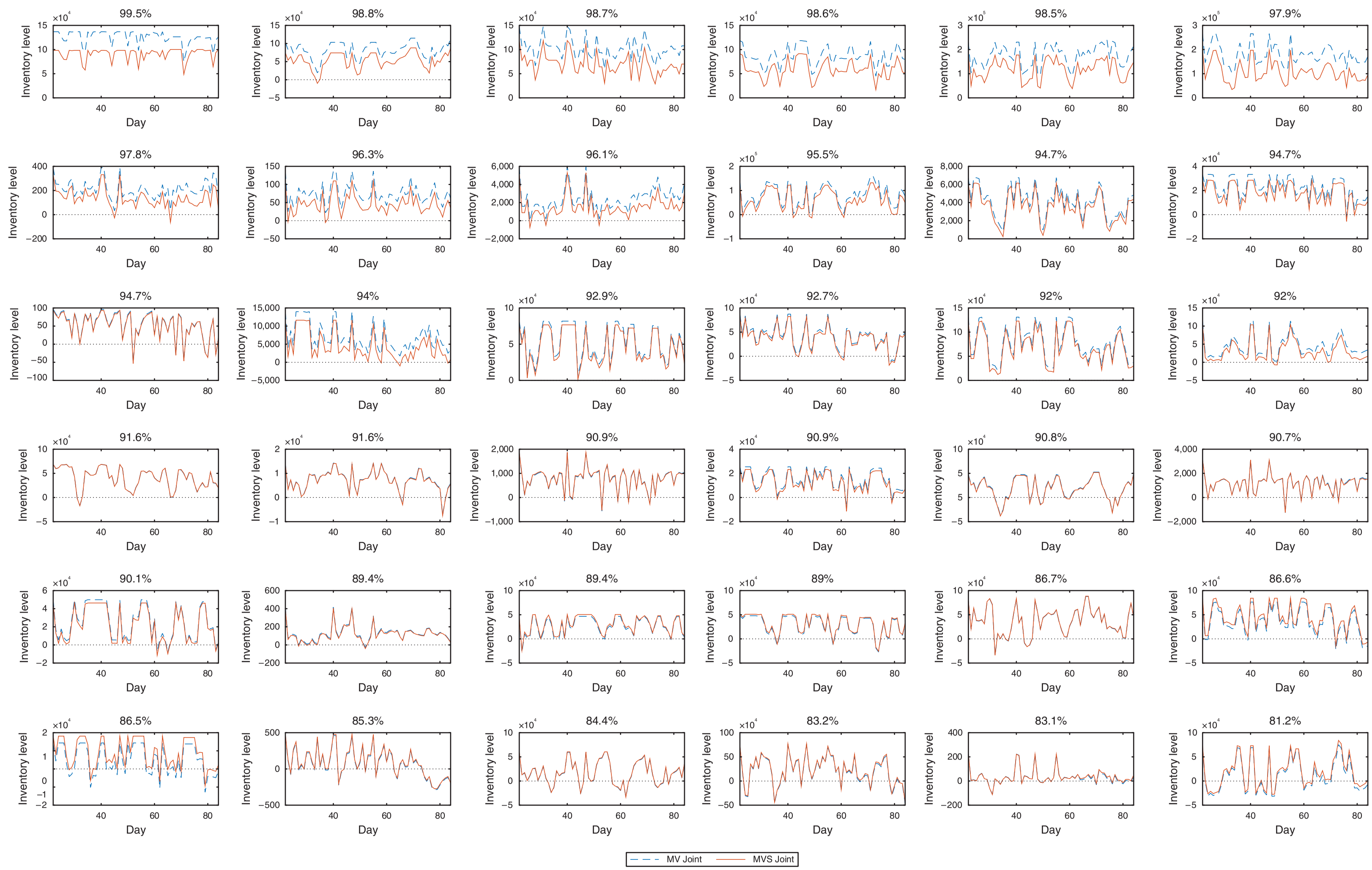
We observe from this table that the MVS Joint method incurs a lower average holding cost and a lower backorder cost compared with the MV Joint method. One reason behind why the MVS Joint method can reduce both types of costs can be seen by inspecting the end-of-day inventory levels. Figure 11 shows the end-of-day inventory levels for the 36 SKUs in one of the problem instances. From the figure, we can see that for SKUs with a high critical ratio (greater than 94%), the MVS Joint method has a lower target inventory level than the MV Joint method while incurring little to no backorders. For low critical ratios (less than 86.7%), the MVS Joint method slightly raises the inventory levels compared to the MV Joint method, resulting in less number of backorders.

6. Conclusion

In this paper, we have proposed a novel information set for the distributionally robust multi-item newsvendor problem. This information set captures asymmetry through second-order partitioned statistics. We have demonstrated for the special case of $n = 1$ that the resulting distributionally robust problem has a closed-form solution. In this special case, asymmetry can be represented through the normalized semivariance. For the general case, the distributionally robust problem is NP-hard. However, we derive a semidefinite program lower bound to the robust problem. Hence, second-order partitioned statistics provides a tractable method of capturing asymmetry in a distributionally robust model.

A major insight from our paper is that when the true demand distribution is heavy tailed, then a distributionally robust model with only mean and covariance information can have a high expected profit loss under the true demand distribution. By including the proposed second-order partitioned statistics in the distributionally robust model, the expected profit loss can be significantly reduced. We demonstrated this through extensive numerical experiments for both the univariate case and for the multivariate case. Hence, second-order partitioned statistics is an informative statistic in the distributionally robust model. We believe that this insight is practically important since it has been empirically observed in several industries that demand has a heavy-tailed distribution. This has been demonstrated in various works such as by Chevalier and Goolsbee (2003) and Gaffeo et al. (2008) for book sales and by Bimpikis and Markakis (2016) for DVD demand and sneaker shoe sales. We also observe this for the automotive spare parts industry when we find that heavy-tailed distributions provide the best fit for almost all of the 36 spare parts' SKUs demand data. This suggests

Figure 11. (Color online) End-of-Day Inventory Levels Throughout 64 Days in the Rolling-Horizon Implementation of the Distributionally Robust Policies for One Problem Instance



Notes. Each panel shows the inventory level of one of the 36 spare parts' SKUs, where the critical ratio is indicated on top. The panels show that the inventory levels with the two partial information models are similar except for SKUs with a critical ratio close to 1. In those SKUs, the robust model with partitioned statistics information sets a significantly lower order-up-to level.

that our approach is promising in practical applications. Moreover, through computational experiments with the automotive spare parts data, we demonstrate that our approach outperforms the distributionally robust approach, which only uses mean and covariance information.

Acknowledgments
The authors thank the associate editor and three reviewers for their comments that helped improve the paper significantly.

Endnote
[1] For example, consider a random variable that takes a value of 0 (with probability (w.p.) 0.2) or a value of 5 (w.p. 0.8). The normalized semivariance is equal to 0, but the distribution is not symmetric.

References
Andersson J, Jörnsten K, Nonås SL, Sandal L, Ubøe J (2013) A maximum entropy approach to the newsvendor problem with partial information. *Eur. J. Oper. Res.* 228(1):190–200.
Ardestani-Jaafari A, Delage E (2016) Robust optimization of sums of piecewise linear functions with application to inventory problems. *Oper. Res.* 64(2):474–494.
Ben-Tal A, Hochman E (1976) Stochastic programs with incomplete information. *Oper. Res.* 24(2):336–347.
Ben-Tal A, den Hertog D, Waegenaere AD, Melenberg B, Rennen G (2013) Robust solutions of optimization problems affected by uncertain probabilities. *Management Sci.* 59(2):341–357.
Bertsimas D, Popescu I (2002) On the relation between option and stock prices: A convex optimization approach. *Oper. Res.* 50(2):358–374.
Bertsimas D, Doan XV, Natarajan K, Teo C (2010) Models for minimax stochastic linear optimization problems with risk aversion. *Math. Oper. Res.* 35(3):580–602.
Bimpikis K, Markakis M (2016) Inventory pooling under heavy-tailed demand. *Management Sci.* 62(6):1800–1813.
Chen L, He S, Zhang S (2011) Tight bounds for some risk measures, with applications to robust portfolio selection. *Oper. Res.* 59(4):847–865.
Chevalier J, Goolsbee A (2003) Measuring prices and price competition online: Amazon.com and BarnesandNoble.com. *Quant. Marketing Econom.* 1(2):203–222.
Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* 58(3):595–612.
De Schepper A, Heijnen B (2007) Distribution-free option pricing. *Insurance: Math. Econom.* 40(2):179–199.
Gaffeo E, Scorcu AE, Vici L (2008) Demand distribution dynamics in creative industries: The market for books in Italy. *Inform. Econom. Policy* 20(3):257–268.
Gallego G, Moon I (1993) The distribution free newsboy problem: Review and extensions. *J. Oper. Res. Soc.* 44(8):825–834.
Gorissen BL, den Hertog D (2013) Robust counterparts of inequalities containing sums of maxima of linear functions. *Eur. J. Oper. Res.* 227(1):30–43.
Han Q, Du D, Zuluaga LF (2014) A risk and ambiguity-averse extension of the max-min newsvendor order formula. *Oper. Res.* 62(3):535–542.
Hanasusanto GA, Kuhn D, Wallace SW, Zymler S (2014) Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Math. Programming* 152(1):1–32.
Isii K (1962) On the sharpness of Tchebyshev-type inequalities. *Ann. Institute Statist. Math.* 14(1):185–197.
Jansen K, Haezendonck J, Goovaerts MJ (1986) Upper bounds on stop-loss premiums in case of known moments up to the fourth order. *Insurance: Math. Econom.* 5(4):315–334.
Levi R, Perakis G, Uichanco J (2012) Regret optimization for stochastic inventory models with spread information. Working paper, Massachusetts Institute of Technology, Cambridge, MA.
Levi R, Perakis G, Uichanco J (2015) The data-driven newsvendor problem: New bounds and insights. *Oper. Res.* 63(6):1294–1306.
Markowitz HM (1959) *Portfolio Selection: Efficient Diversification of Investments* (John Wiley & Sons, New York).
Natarajan K, Teo CP (2017) On reduced semidefinite programs for second order moment bounds with applications. *Math. Programming* 161(1):487–518.
Natarajan K, Sim M, Uichanco J (2010) Tractable robust expected utility and risk models for portfolio optimization. *Math. Finance* 20(4):695–731.
Numpacharoen K, Atsawarungruangkit A (2012) Generating correlation matrices based on the boundaries of their coefficients. *PLoS One* 7(11):e48902.
Perakis G, Roels G (2008) Regret in the newsvendor model with partial information. *Oper. Res.* 56(1):188–203.
Saghafian S, Tomlin B (2016) The newsvendor under demand ambiguity: Combining data with moment and tail information. *Oper. Res.* 64(1):167–185.
Scarf H (1958) A min-max solution of an inventory problem. Arrow KJ, Karlin S, Scarf H, eds. *Studies in the Mathematical Theory of Inventory and Production* (Stanford University Press, Palo Alto, CA), 201–209.
Wiesemann W, Kuhn D, Sim M (2014) Distributionally robust convex optimization. *Oper. Res.* 62(6):1358–1376.
Yue J, Chen B, Wang M-C (2006) Expected value of distribution information for the newsvendor problem. *Oper. Res.* 54(6): 1128–1136.
Zhu Z, Zhang J, Ye Y (2013) Newsvendor optimization with limited distribution information. *Optim. Methods Software* 28(3): 640–667.
Zhu Z, Zhang J, Ye Y (2016) Likelihood robust optimization for data-driven problems. *Comput. Management Sci.* 13(2):241–261.
Zuluaga LF, Peña J, Du D (2009) Third-order extensions of Lo's semiparametric bound for European call options. *Eur. J. Oper. Res.* 198(2):557–570.