Asymmetry and Ambiguity in Newsvendor Models

Karthik Natarajan* Melvyn Sim† Joline Uichanco‡

Submitted: July 14, 2008

Abstract

The traditional decision-making framework for newsvendor models is to assume a distribution of the underlying demand. However, the resulting optimal policy is typically sensitive to the choice of the distribution. A more conservative approach is to assume that the distribution belongs to a set parameterized by a few known moments. An ambiguity-averse newsvendor would choose to maximize the worst-case profit. Most models of this type assume that only the mean and the variance are known, but do not attempt to include asymmetry properties of the distribution. Other recent models address asymmetry by including skewness and kurtosis. However, closed-form expressions on the optimal bounds are difficult to find for such models. In this paper, we propose a framework under which the expectation of a piecewise linear objective function is optimized over a set of distributions with known asymmetry properties. This asymmetry is represented by the first two moments of multiple random variables that result from partitioning the original distribution. In the simplest case, this reduces to semivariance. The optimal bounds can be solved through a second-order cone programming (SOCP) problem. This framework can be applied to the risk-averse and risk-neutral newsvendor problems and option pricing. We provide a closed-form expression for the worst-case newsvendor profit with only mean, variance and semivariance information.

1 Introduction

The single-period newsvendor problem is the foundation of many inventory control models. In the classical version, a newsvendor decides before the sales period how many units of a product to order. The actual demand occurs during the sales period and is satisfied as much as possible with the units on hand. The newsvendor incurs a cost c for each ordered unit, and sells each unit for a price p. If the distribution of the demand is known, then the optimal ordering quantity is the $1 - \frac{c}{p}$ quantile of

*matkbn@nus.edu.sg. National University of Singapore, Department of Mathematics, NUS Risk Management Institute, Singapore-MIT Alliance.

†dscsimm@nus.edu.sg. NUS Business School, NUS Risk Management Institute, Singapore-MIT Alliance. The research of the author was partially supported by NUS academic research grant R-314-000-068-122.

‡uichanco@mit.edu. Massachusetts Institute of Technology, Operations Research Center. The research of the author was partially supported by the Singapore-MIT Alliance.

the distribution. However, as it is often in practical scenarios, a full characterization of the demand distribution may not be available. How does the newsvendor decide on an ordering policy in this case? Usually, if there is lack of information about a distribution, there is a tendency to assume a Gaussian distribution. Although this approach provides an easy way to compute the optimal quantile, the assumed distribution is symmetric. If the true demand distribution exhibits asymmetry, then this information will be lost by fitting a Gaussian distribution. In fact, the effect of asymmetry can sometimes prove to be devastating. Suppose that the unit cost c is very small relative to the price p. If the true distribution is in fact positively skewed, then the true optimal ordering quantity is greatly over-estimated. Bartezzaghi et al. [2] investigate the impact of asymmetry on manufacturing, planning and control systems. By simulating different distribution shapes, they discover that the shape has a huge impact on inventory levels. They conclude that practitioners must recognize whether the demand is regular or irregular (multi-modal and asymmetric) in order to better estimate the inventory levels. Their conclusion, however, stems only from observing simulated distributions and not from any analytical results.

Perhaps it is valid to think that a stochastic model is successful if it well describes actual uncertain scenarios. If a system is known to contain asymmetry, then it is important to have stochastic models that accurately describe this. One phenomenon in production is a "lumpy" or sporadic demand which is characterized by infrequent and large demand. This demand pattern has been observed in parts and supply inventory systems such as large compressors and textile machines (Silver [35]; Croston [12]; Ward [37]). Bartezzaghi et al. [3] identify possible sources which can contribute to demand lumpiness. These include number of potential customers, heterogeneity of customers, and frequency and variety of requests. Asymmetric distributions are also observed in many other applications. This is probably most established and well-documented in the area of finance. Cont [11] presents an overview of "stylized facts" that are observed from the statistical analysis of asset prices in various financial markets. For instance, there appears to be a gain and loss asymmetry in which one observes large downward movements in stock prices but not equally large upward movements. Another stylized fact is that the distribution of asset returns exhibits a shape that is far from normal. In particular, its distribution appears to show a substantial degree of excess kurtosis (Fama [15]). Indeed, one major criticism against the popular Black-Scholes model is that it fails to explain this leptokurtotic behavior, since the basic assumption of the model is the lognormality of the asset price distribution. Other models have tried to do better by relaxing the assumptions of the Black-Scholes model. For instance, the GARCH framework attempts to model heteroskedasticity of asset returns (see Engle [13]; Bollerslev [7]). However, the conditional density function is still typically assumed to be symmetric.

One of the most common and natural methods to introduce asymmetry into a stochastic model is by assuming that the true distribution is some well-known asymmetrical distribution. For example, in the newsvendor model, it is common to assume a lognormal or a Poisson distribution. Other popular methods forecast lumpy demand by applying traditional forecasting techniques based on past demand information (Croston [12]; Johnston and Boylan [24]). However, these techniques usually result in large forecast errors since the data is sporadic (Fildes and Beard [16]). In finance literature, Sortino and Forsey [36] suggest fitting the data to a three-parameter lognormal distribution. Another model proposed by Knight, Satchell and Tran [25] involves partitioning the distribution into the upper and lower partial distributions with respect to some benchmark. Each of the two distributions is then modeled by a gamma distribution. However, they acknowledge that the model is not successful in modeling small changes, and they are unable to build in conditional heteroskedasticity into their model. Bond [8] addresses these issues by using a GARCH framework and modeling the conditional variance using the double gamma distribution. However, after testing the model on various exchange rates, he admits that evidence on the performance of the double gamma model is not overwhelmingly convincing.

Assuming a distribution for a random variable (such as demand, stock price or returns), though giving us a complete picture of the randomness, does not usually result in robustness. Since the decision is made fully under the assumed distribution, it might not perform well under other distributions. Instead of assuming a particular distribution, it is then more reasonable to assume that the distribution belongs to the set of all distributions satisfying known parameters (e.g. mean and variance). These parameters may come from estimates using past realizations or some prediction by industry experts. A conservative approach, often called the maximin approach, optimizes the worst-case objective (e.g. expected profit or payoff) over the parametric family. This arguably models the behavior of an "ambiguity-averse" decision-maker who knows little information about the demand. Early research that uses the maximin approach describes the distribution set by some known mean and variance (see Scarf [33]; Lo [26]; Gallego and Moon [20]). However, one major criticism against the maximin approach is that the resulting policies can be too conservative. A less conservative approach, called minimax regret, minimizes the maximum opportunity cost from not making the optimal decision instead (Savage [32]; Perakis and Roels [31]). Due to their second-order nature, closed-form expressions for the optimal bounds have been found for most of these robust models with known mean and variance. However, these models do not include information about the asymmetry of the distribution. Recent literature attempts to address this by assuming knowledge of higher moments, such as skewness or kurtosis (Jansen et al. [23]; De Schepper and Heijnen [34]; He et al. [21]; Zuluaga et al. [38]). Since the higher moments result in a problem with third and fourth order constraints, it is usually not easy to find a closed-form expression for the optimal bounds. Even if they are found, the expression can be complicated and can give limited managerial insight. Moreover, there remains an important issue: how does one actually go about calculating these higher moments? To quote Sortino and Forsey [36] who state a popular belief: "Others might say it is difficult enough to estimate the mean and variance; attempting to estimate higher moments is pretentious or superfluous." More importantly, skewness and kurtosis are not so intuitive for many industry practitioners, so advocating a model based on these might prove to be difficult.

Our approach is to represent asymmetry using a well-known measure: semivariance. Semivariance is a measure of downside risk put forward by Markowitz [28] in his seminal paper on portfolio selection. He argues that it is a much better measure of risk compared to variance, since it only considers deviations that are below some specified target, whereas variance equally penalizes positive and negative deviations from the mean. The semivariance of a random variable $\tilde{x}$ for some target α is a special case of the lower partial moment

$$\mathrm{E}\left((\alpha - \tilde{x})^{+n}\right),$$

with $n = 2$, where $x^+ = \max(0, x)$. Fishburn [17] describes the parameter n as a reflection of the feeling about the relative consequences of falling short of the target. If the decision-maker's main concern is the failure to achieve the target, without particular regard to the amount, then a small value of n is appropriate. On the other hand, a larger value of n should be used if small deviations are relatively harmless compared to larger ones. We will use semivariance as a parameter that indicates the degree of asymmetry. The relative magnitude of semivariance compared to variance indicates how the deviations from the mean are split between the upper and lower parts of the distribution. On a technical note, one advantage of using semivariance in moment bounds is that asymmetry is introduced without needing to stray from the simplicity of second-order models. A consequence of this is that closed-form expressions for optimal bounds are more easily derived. Moreover, there is virtue in espousing an asymmetry model using semivariance since it is a largely accepted measure in the financial industry. The idea of using semivariance as a measure of asymmetry is not new. In fact, it was used by Berck and Hihn [6] to tighten the Chebychev inequality which uses mean and variance to bound the probability of an outcome k standard deviations below the mean. They show that by using semivariance instead of variance, a sharper bound can found. Their result is especially useful when the underlying distribution is asymmetric.

An important question remains: is it any easier to make accurate estimates of semivariance? Several approaches have been proposed to calculate semivariance. Methods by Sortino and Forsey [36] and Bond [8], for instance, calculate semivariance directly from some estimated density function. However, the simplest and most obvious approach is still through a sample-based calculation. Yet one common concern is that the volatility of the sample-based semivariance is so high as to make it impractical in applied work. In fact, Sortino and Forsey [36] are critical of using the sample-based semivariance, since they argue that it can easily over- or under-estimate the true semivariance due to its dependency on some target value. Instead, they suggest that fitting a continuous probability density function is superior to discrete sample calculations. However, in a recent paper by Bond and Satchell [9] which studies the statistical properties of the sample semivariance, it has been shown that sample semivariance is in fact less volatile than sample variance when the distribution is asymmetrical. Their results suggest that the major concern of practitioners against using sample semivariance is not valid.

In this paper, we make a distinction between "ambiguity-aversion" and "risk-aversion". Ambiguity-aversion has been demonstrated in the famous Ellsberg paradox[1]. An ambiguity-averse newsvendor would prefer a sure profit over something unsure but with a potentially higher payoff. Although this behavior has been observed empirically, it cannot be accommodated within an expected utility model (Epstein [14]). We attempt to model ambiguity-aversion by a maximin model. We can argue that if there is limited information, the strategy of an ambiguity-averse newsvendor would be to optimize the worst possible expected profit.

[1] People prefer to bet on an urn with 50 red and 50 blue balls, than in one with 100 balls but where the number of blue or red balls is unknown.

Our contributions can be classified as follows:

1. *Asymmetry Model for an Ambiguity-Averse, Risk-Neutral Newsvendor.* We propose the use of the normalized semivariance s (defined in (2.1)) as a measure of the degree of asymmetry in a distribution. In Section 2, under a risk-neutral setting, we use a robust maximin approach to derive optimal newsvendor policies with only mean, variance and semivariance information. We find a closed-form expression for the lower bound on the expected newsvendor profit under this model. This bound is tight, in the sense that there exists a distribution under which the bound is achieved. The worst-case distribution has at most three support points. We find that the optimal policy is less conservative than Scarf's [33] policy. Scarf's model recommends ordering above or below the mean depending on whether c/p is below or above $\frac{1}{2}$. Our asymmetric model recommends a slightly altered policy of ordering above or below the mean depending on whether this ratio is below or above $\frac{1}{2}(1-s)$.

2. *Asymmetry Model for an Ambiguity-Averse, Risk-Averse Newsvendor.* We apply the asymmetry model to the newsvendor problem under a risk-averse setting in Section 3. In particular, we consider the problem of minimizing the worst-case optimized certainty equivalent (OCE) risk of the newsvendor profit. We show that this problem has an equivalent second-order cone programming (SOCP) formulation. A special case is the problem of maximizing the worst-case expected utility of a newsvendor with a piecewise linear utility. We provide examples for a risk-averse newsvendor with a CVaR objective under the mean-variance and the asymmetry models. Through computational experiments, we find that the mean-variance policy is sometimes counterintuitive by recommending more conservatism for a less risk-averse newsvendor. Our asymmetry model on the other hand recommends adaptive policies that are appropriate for the level of risk tolerance.

3. *Option Pricing.* In Section 2, we apply our asymmetry bounds to option pricing. We find a closed-form expression for the tight upper bound of the expected option payoff under mean, variance and semivariance information on the stock price distribution. We also show that the tight lower bound can be found by solving an SOCP. We find that the range of expected payoffs can become very narrow if there is a high degree of asymmetry.

4. *General Partitioned Model.* In Appendix A, we generalize our approach to include multiple partitions of the distribution and under a piecewise linear utility function. Under knowledge of the first two moments of the partitioned random variables, we formulate the tight lower and upper bounds on the expected utility as moment problems. We introduce a method for solving these tight bounds through an SOCP problem, which can be efficiently solved using interior point algorithms. This general model can admit many variations to the newsvendor problem, such as a bounded support or partitions defined on the distribution quantiles.

The structure of the paper is as follows. We discuss the risk-neutral newsvendor model in Section 2 and introduce our asymmetry model. In Section 3, we consider a risk-averse newsvendor and develop methods to find optimal policies in this setting. Finally, in Appendix A, we generalize the model to multiple partitions and a piecewise linear objective.

2 Ambiguity-Averse, Risk-Neutral Newsvendor

Consider a newsvendor facing a random demand $\tilde{d}$ for the product observed during the sales period. He satisfies the demand as much as possible with the units he has preordered. Any unmet demand is assumed to be lost. Let c be the unit ordering cost and p the exogenously determined unit selling price. A standard assumption is $p > c$, since otherwise, the newsvendor will choose to order nothing. If random demand has a probability density function f, then for a given order quantity q, the newsvendor's expected profit is then

$$p\mathrm{E}_f\left(\min\{\tilde{d}, q\}\right) - cq.$$

Here the expectation $\mathrm{E}_f(\cdot)$ is taken with respect to the known distribution f. A risk-neutral newsvendor would be concerned with finding an ordering policy that maximizes the expected newsvendor profit. Suppose, rather than having a complete knowledge of the demand distribution, all the newsvendor knows are some of its parameters (e.g., known moments). Instead of maximizing the expected profit under some assumed distribution, an ambiguity-averse newsvendor will take a conservative approach by maximizing the worst-case profit

$$\inf_{f\in\mathbb{F}} p\mathrm{E}_f\left(\min\{\tilde{d}, q\}\right) - cq,$$

where $\mathbb{F}$ is the parametric family of distributions satisfying the known information.

2.1 Mean-Variance (MV) Model

Scarf [33] addressed the robust newsvendor model when the parametric family of distributions consist of those with mean μ, variance σ^2 and nonnegative support. The worst-case newsvendor profit under this setting is

$$\begin{aligned} MV(q) \quad = \quad & \inf_{f} \; p\mathrm{E}_f\left(\min\{\tilde{d}, q\}\right) - cq \\ & \text{s.t.} \;\; \mathrm{E}_f(\tilde{d}) = \mu, \quad \mathrm{E}_f(\tilde{d}^2) = \mu^2 + \sigma^2 \\ & \qquad \mathrm{E}_f(1) = 1, \quad f(\tilde{d}) \geq 0, \quad \forall \tilde{d} \geq 0. \end{aligned}$$

For any distribution belonging in this set, he found through a lengthy mathematical argument the optimal lower bound for the expected newsvendor profit. In particular,

$$p\mathrm{E}_f\left(\min\{\tilde{d}, q\}\right) - cq \geq \begin{cases} pq\frac{\mu^2}{\mu^2+\sigma^2} - cq, & \text{for } q \in \left[0, \frac{\mu^2+\sigma^2}{2\mu}\right], \\ p\left(\frac{\mu+q}{2} - \frac{1}{2}\sqrt{(q-\mu)^2+\sigma^2}\right) - cq, & \text{for } q \in \left[\frac{\mu^2+\sigma^2}{2\mu}, \infty\right). \end{cases}$$

This bound is tight, in the sense that there exists a feasible distribution with mean μ and variance σ^2 whose expected profit is exactly equal to the lower bound. The worst-case distribution is one that has a positive mass at exactly two points. If $q \leq (\mu^2 + \sigma^2)/(2\mu)$, the worst-case distribution has mass $\sigma^2/(\mu^2+\sigma^2)$ at 0 and mass $\mu^2/(\mu^2+\sigma^2)$ at $(\mu^2+\sigma^2)/\mu$. Otherwise, the worst-case two-point distribution is

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

Gallego and Moon [20] reach the same conclusion, but with a more concise proof that invokes the use of Cauchy-Schwartz inequality.

It is straightforward to find that an optimal ordering policy q^* that maximizes the worst-case profit is given by:

$$q^* = \begin{cases} 0, & \text{if } \frac{c}{p} \geq \frac{\mu^2}{\mu^2+\sigma^2}, \\ \mu + \frac{\sigma}{2}\frac{(p-2c)}{\sqrt{c(p-c)}}, & \text{if } \frac{c}{p} \leq \frac{\mu^2}{\mu^2+\sigma^2}. \end{cases}$$

Under this ordering policy, the worst-case expected profit is

$$MV(q^*) = \begin{cases} 0, & \text{if } \frac{c}{p} \geq \frac{\mu^2}{\mu^2+\sigma^2}, \\ (p-c)\mu - \sigma\sqrt{c(p-c)}, & \text{if } \frac{c}{p} \leq \frac{\mu^2}{\mu^2+\sigma^2}. \end{cases}$$

It might be interesting to look at the flip side of the coin, or the best-case scenario. A naïve upper bound for the expected newsvendor profit can be found by invoking Jensen's inequality. Thus,

$$p\mathrm{E}_f\left(\min\{\tilde{d}, q\}\right) - cq \leq \min\{p\mu - cq, (p-c)q\}.$$

In fact, this upper bound is tight under mean and variance information (see De Schepper and Heijnen [34]).

2.2 Mean-Variance-Semivariance (MVS) Model

We introduce asymmetry into the robust newsvendor model through a characterization of the lower partial moments of the demand distribution. We will be focusing on the partial moments taken with respect to the mean demand (i.e. $\alpha = \mu$) to derive a closed-form expression for the worst-case newsvendor profit. The first-order lower partial moment does not capture asymmetry, because regardless of the distribution f, it is always true that

$$\mathrm{E}_f\left((\mu - \tilde{d})^+\right) = \mathrm{E}_f\left((\tilde{d} - \mu)^+\right).$$

This relationship results from the partial moment being defined with respect to the mean. Hence, we simply focus on the second order ($n = 2$) lower partial moment, or semivariance. It is entirely possible to include asymmetry information in the first partial moment if it is defined with respect to some value other than the mean. However, the resulting closed-form expression for the model with the first two lower partial moments is untidy at best and and does not give us much additional insights. Instead, we direct interested readers to Appendix A which shows how this model can be solved as an SOCP.

We introduce the notion of *normalized semivariance*, which we define as

$$s \equiv \frac{\mathrm{E}_f\left((\tilde{d} - \mu)^{+2}\right) - \mathrm{E}_f\left((\mu - \tilde{d})^{+2}\right)}{\sigma^2}. \tag{2.1}$$

This measure is only defined for random variables with a strictly positive and finite variance. We can immediately see that the normalized semivariance must take values in the range of -1 to 1. Clearly, s describes how the volatility of the demand is divided between the upper and lower parts of the distribution. Figure 2.1 shows examples of some common probability distributions and their semivariances.

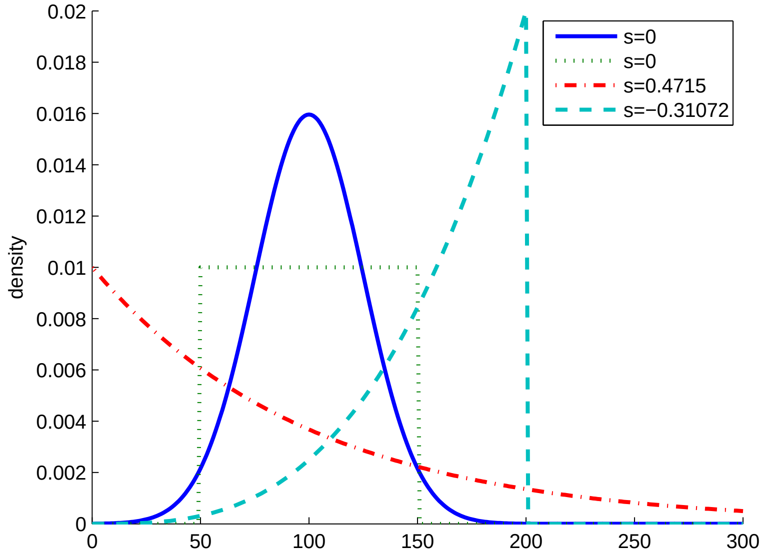

Figure 2.1: Some probability density functions and their normalized semivariances.

Normal and uniform distributions always have a normalized semivariance of zero. An exponential distribution always has $s = 4e^{-1} - 1 \approx 0.4715$. The s value of a beta distribution can be positive or negative depending on its parameters. In fact, we can think of a value of $s = 0$ as a weaker form of distributional symmetry. All symmetric distributions (e.g. uniform, normal) must have a normalized semivariance of zero. However, the converse is not true. We can also think of a distribution with $s > 0$ as roughly positively skewed. Similarly, $s < 0$ implies that the distribution is roughly negatively skewed. We have mentioned that s is in between -1 and 1. In fact, in the following proposition, we find a tighter bound for s for nonnegative distributions.

Proposition 2.1. *Consider a nonnegative random variable with mean $\mu > 0$, standard deviation $\sigma > 0$ and normalized semivariance s. Then the necessary and sufficient condition for the moments to be feasible is:*

$$\frac{\sigma^2 - \mu^2}{\sigma^2 + \mu^2} \leq s < 1. \tag{2.2}$$

Moreover, there exists a unique distribution for which the lower bound is tight.

Proof. Let us first prove that any feasible distribution satisfies condition (2.2). For notational convenience, we define the nonnegative random variables $\tilde{d}_1 = (\tilde{d} - \mu)^+$ and $\tilde{d}_2 = (\mu - \tilde{d})^+$. From the definition of s, we have $\mathrm{E}(\tilde{d}_1^2) = \frac{1}{2}(1 + s)\sigma^2$ and $\mathrm{E}(\tilde{d}_2^2) = \frac{1}{2}(1 - s)\sigma^2$. Since $\tilde{d}$ is nonnegative, $\tilde{d}_2$ must not exceed μ. This implies that $\mathrm{E}(\tilde{d}_2(\mu - \tilde{d}_2)) \geq 0$, or equivalently, $\mathrm{E}(\tilde{d}_2) \geq \mathrm{E}(\tilde{d}_2^2)/\mu = (1 - s)\sigma^2/(2\mu)$.

Also since $\mathrm{E}(\tilde{d}_2^2|\tilde{d}_2 > 0) \geq (\mathrm{E}(\tilde{d}_2|\tilde{d}_2 > 0))^2$, then we have

$$\Pr(\tilde{d}_2 > 0) \geq \frac{\Big(\mathrm{E}(\tilde{d}_2|\tilde{d}_2 > 0)\Pr(\tilde{d}_2 > 0) + 0 \cdot \Pr(\tilde{d}_2 = 0)\Big)^2}{\mathrm{E}(\tilde{d}_2^2|\tilde{d}_2 > 0)\Pr(\tilde{d}_2 > 0) + 0^2 \cdot \Pr(\tilde{d}_2 = 0)} = \frac{(\mathrm{E}(\tilde{d}_2))^2}{\mathrm{E}(\tilde{d}_2^2)} \geq \frac{(1-s)\sigma^2}{2\mu^2}. \tag{2.3}$$

By a similar argument, and since $\mathrm{E}(\tilde{d}_1) = \mathrm{E}(\tilde{d}_2)$, we also find that

$$\Pr(\tilde{d}_1 > 0) \geq \frac{\Big(\mathrm{E}(\tilde{d}_1|\tilde{d}_1 > 0)\Pr(\tilde{d}_1 > 0) + 0 \cdot \Pr(\tilde{d}_1 = 0)\Big)^2}{\mathrm{E}(\tilde{d}_1^2|\tilde{d}_1 > 0)\Pr(\tilde{d}_1 > 0) + 0^2 \cdot \Pr(\tilde{d}_1 = 0)} = \frac{(\mathrm{E}(\tilde{d}_1))^2}{\mathrm{E}(\tilde{d}_1^2)} \geq \frac{(1-s)^2}{(1+s)}\frac{\sigma^2}{2\mu^2}. \tag{2.4}$$

Finally, note that by our definition, $\Pr(\tilde{d}_1 > 0) = \Pr(\tilde{d} > \mu)$ and $\Pr(\tilde{d}_2 > 0) = \Pr(\tilde{d} < \mu)$. Thus, by inequalities (2.3)–(2.4), it follows that

$$1 \geq \Pr(\tilde{d} > \mu) + \Pr(\tilde{d} < \mu) \geq \frac{\sigma^2}{\mu^2}\frac{(1-s)}{(1+s)}, \tag{2.5}$$

which gives us the lower bound on s. Observe that $s = \frac{\sigma^2-\mu^2}{\sigma^2+\mu^2}$ is equivalent to having inequalities (2.3)–(2.5) tight. Then this would imply that

$$\begin{array}{ll} \Pr(\tilde{d} > \mu) = \frac{\mu^2}{\mu^2+\sigma^2}, & \mathrm{E}((\tilde{d}-\mu)^+) = \frac{\mu\sigma^2}{\mu^2+\sigma^2}, \\ \Pr(\tilde{d} < \mu) = \frac{\sigma^2}{\mu^2+\sigma^2}, & \mathrm{E}((\mu-\tilde{d})^+) = \frac{\mu\sigma^2}{\mu^2+\sigma^2}. \end{array}$$

Therefore, we have the following conditional moments:

$$\begin{array}{ll} \mathrm{E}((\tilde{d}-\mu)^+|\tilde{d} > \mu) = \frac{\sigma^2}{\mu}, & \mathrm{E}((\tilde{d}-\mu)^{+2}|\tilde{d} > \mu) = \frac{\sigma^4}{\mu^2}, \\ \mathrm{E}((\mu-\tilde{d})^+|\tilde{d} < \mu) = \mu, & \mathrm{E}((\mu-\tilde{d})^{+2}|\tilde{d} < \mu) = \mu^2. \end{array}$$

The only distribution satisfying these conditions is a two-point support distribution with positive mass at 0 and $(\mu^2+\sigma^2)/\mu$. In other words, if $s = \frac{\sigma^2-\mu^2}{\sigma^2+\mu^2}$, the feasible distribution set only contains a single distribution. Now let us verify that any s that satisfies (2.2) implies a feasible distribution. For a given (μ, σ, s) triplet, we can find a two-point distribution:

$$\tilde{d} = \begin{cases} \mu - \sigma\sqrt{\frac{1-s}{1+s}}, & \text{w.p. } \frac{1+s}{2}, \\ \mu + \sigma\sqrt{\frac{1+s}{1-s}}, & \text{w.p. } \frac{1-s}{2}. \end{cases}$$

These support points are nonnegative in the range (2.2). In fact, s can be arbitrarily close but never equal to one since, by definition,

$$s = \frac{\mathrm{E}((\tilde{d}-\mu)^{+2}) - \mathrm{E}((\mu-\tilde{d})^{+2})}{\sigma^2} = 1 - \frac{2}{\sigma^2}\mathrm{E}((\mu-\tilde{d})^{+2}).$$

Since we assume that $\sigma > 0$, then there exists $\tilde{d} < \mu$ with nonzero probability. □

As an interesting aside, note that the only feasible distribution when $s = \frac{\sigma^2-\mu^2}{\sigma^2+\mu^2}$ is the Scarf two-point distribution for the case when $q \leq \frac{\mu^2+\sigma^2}{2\mu}$.

Consider a newsvendor model where the exact demand distribution is unknown, but the mean μ, variance σ^2 and normalized semivariance s are known. For a given quantity q, the worst-case expected newsvendor profit is

$$\begin{aligned} MVS(q) \quad = \quad & \inf_{f} \quad p\mathrm{E}_f\left(\min\{\tilde{d}, q\}\right) - cq \\ & \text{s.t.} \quad \mathrm{E}_f(\tilde{d}) = \mu, \quad \mathrm{E}_f\left((\tilde{d}-\mu)^2\right) = \sigma^2, \\ & \qquad \mathrm{E}_f\left((\tilde{d}-\mu)^{+2}\right) - \mathrm{E}_f\left((\mu-\tilde{d})^{+2}\right) = s\sigma^2, \\ & \qquad \mathrm{E}_f(1) = 1, \quad f(\tilde{d}) \geq 0, \quad \forall \tilde{d} \geq 0. \end{aligned}$$

We can consider f to be an infinite dimensional vector indexed by $\tilde{d} \in \Re^+$, such that $f(\tilde{d}) : \Re^+ \mapsto \Re^+$. We assume that the conditions for Proposition 2.1 are satisfied so that the moment problem $MVS(\cdot)$ is well-defined. By Isii's [22] strong duality theorm, $MVS(q)$ is equivalent to the dual problem

$$\begin{aligned} MVS_D(q) \quad = \quad & \sup_{t,r,y_1,y_2} \quad t + r\mu + y_1\sigma^2 + y_2 s\sigma^2 \\ & \text{s.t.} \quad t + rx + y_1(x-\mu)^2 - y_2(x-\mu)^2 \leq px - cq, \quad \forall 0 \leq x \leq \mu, \\ & \qquad t + rx + y_1(x-\mu)^2 - y_2(x-\mu)^2 \leq (p-c)q, \quad \forall 0 \leq x \leq \mu, \\ & \qquad t + rx + y_1(x-\mu)^2 + y_2(x-\mu)^2 \leq px - cq, \quad \forall x \geq \mu, \\ & \qquad t + rx + y_1(x-\mu)^2 + y_2(x-\mu)^2 \leq (p-c)q, \quad \forall x \geq \mu. \end{aligned}$$

We can convert this dual problem into an equivalent SOCP formulation which can be solved efficiently through interior point methods. The details can be found in Appendix A for the more general case.

In fact, we can find a closed-form expression for the worst-case expected newsvendor profit $MVS(\cdot)$ with only knowledge of mean, variance, semivariance and nonnegativity. Theorem 2.1 presents this optimal bound. The proof of the theorem, which we relegate to Appendix B, is quite involved since it consists of constructing various forms of the dual feasible solutions. For each dual solution, we find a corresponding primal feasible distribution that achieves the same objective value. Note that in the theorem, the domain of q is partitioned into five different regions. Each of the regions implies a particular form of the dual feasible solution which is optimal. A distribution that gives the worst-case expected profit is in fact one with at most three support points.

Theorem 2.1. *Consider a newsvendor problem specified by a unit cost c and unit price p. Suppose the demand distribution has mean μ, standard deviation σ, and normalized semivariance s. For a given* *order quantity q, a lower bound for the expected newsvendor profit is*

$$\begin{cases} (p-c)q - \frac{p(1-s)\sigma^2}{2\mu^2}q, & \textit{for (i): } q \in \left[0, \frac{\mu}{2}\right], \\ (p-c)q - \frac{p(1-s)\sigma^2}{8(\mu-q)}, & \textit{for (ii): } q \in \left[\frac{\mu}{2}, \mu - \frac{\sigma}{2}\sqrt{\frac{1-s}{1+s}}\right], \\ p\left(\frac{(1-s)}{2}q + \frac{(1+s)}{2}\mu - \frac{\sigma}{2}\sqrt{1-s^2}\right) - cq, & \textit{for (iii): } q \in \left[\mu - \frac{\sigma}{2}\sqrt{\frac{1-s}{1+s}}, \mu + \frac{\sigma}{2}\sqrt{\frac{1+s}{1-s}}\right], \\ p\mu - cq - \frac{p(1+s)\sigma^2}{8(q-\mu)}, & \textit{for (iv): } q \in \left[\mu + \frac{\sigma}{2}\sqrt{\frac{1+s}{1-s}}, \mu + \frac{\mu(1+s)}{2(1-s)}\right], \\ \frac{p}{2}\left(\mu + bq - \sqrt{(bq-\mu)^2 - (1-b)^2\mu^2 + \frac{(1+s)\sigma^2 b}{2}}\right) - cq, & \textit{for (v): } q \in \left[\mu + \frac{\mu(1+s)}{2(1-s)}, \infty\right), \end{cases}$$

where

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

Moreover, among the set of nonnegative distributions parameterized by (μ, σ, s), there exists a distribution with at most three support points that attains this bound.

Proof. See Appendix B.

Note that $MVS(q)$ is a continuous and concave function of q. We can then easily find the order quantity q^* that maximizes the worst-case expected profit. Theorem 2.2 provides the optimal policy that depends on the magnitude of the ratio of unit cost to unit price c/p.

Theorem 2.2. *Consider a newsvendor problem specified by a unit cost c and unit price p. Suppose the demand distribution has mean μ, standard deviation σ, and normalized semivariance s. An ordering policy q^* that maximizes the worst-case expected profit is*

$$q^* = \begin{cases} 0, & \textit{if } 1 - \frac{(1-s)\sigma^2}{2\mu^2} \le \frac{c}{p}, \\ \mu - \frac{\sigma}{2}\sqrt{\frac{(1-s)p}{2(p-c)}}, & \textit{if } \frac{1}{2}(1-s) \le \frac{c}{p} < 1 - \frac{(1-s)\sigma^2}{2\mu^2}, \\ \mu + \frac{\sigma}{2}\sqrt{\frac{(1+s)p}{2c}}, & \textit{if } \frac{1}{2}\frac{(1-s)^2}{(1+s)}\frac{\sigma^2}{\mu^2} \le \frac{c}{p} < \frac{1}{2}(1-s), \\ \frac{\mu}{b} + \frac{(pb-2c)}{2b}\sqrt{\frac{(1+s)\sigma^2 b - 2(1-b)^2\mu^2}{2c(pb-c)}}, & \textit{if } \frac{c}{p} < \frac{1}{2}\frac{(1-s)^2}{(1+s)}\frac{\sigma^2}{\mu^2}, \end{cases}$$

where

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

The worst-case expected newsvendor profit attained by the policy is

$$MVS(q^*) = \begin{cases} 0, & \textit{if } 1 - \frac{(1-s)\sigma^2}{2\mu^2} \le \frac{c}{p}, \\ (p-c)\mu - \frac{\sigma}{2}\sqrt{2p(p-c)(1-s)}, & \textit{if } \frac{1}{2}(1-s) \le \frac{c}{p} < 1 - \frac{(1-s)\sigma^2}{2\mu^2}, \\ (p-c)\mu - \frac{\sigma}{2}\sqrt{2pc(1+s)}, & \textit{if } \frac{1}{2}\frac{(1-s)^2}{(1+s)}\frac{\sigma^2}{\mu^2} \le \frac{c}{p} < \frac{1}{2}(1-s), \\ \left(p - \frac{c}{b}\right)\left(\mu - \sqrt{\frac{c((1+s)\sigma^2 b - 2(1-b)^2\mu^2)}{2(pb-c)}}\right), & \textit{if } \frac{c}{p} < \frac{1}{2}\frac{(1-s)^2}{(1+s)}\frac{\sigma^2}{\mu^2}. \end{cases}$$

Proof. See Appendix B.

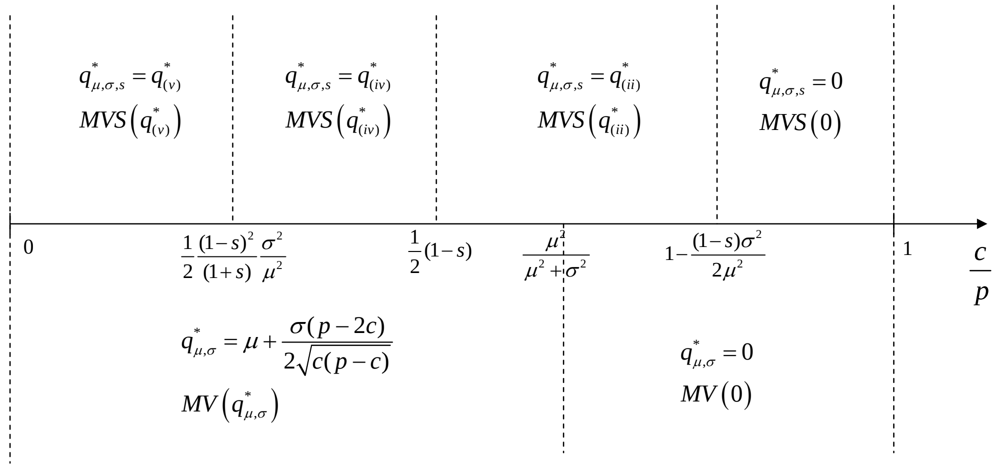

Figure 2.2: Optimal policies under different ranges of the unit cost to unit price ratio c/p.

Figure 2.2 illustrates how the ordering policies induced by the MV model and the MVS model are related. The values $q^*_{(ii)}, q^*_{(iv)}, q^*_{(v)}$ are defined in the proof of Theorem 2.2 (the subscript refers to the region where the optimum lies). Due to Proposition 2.1, we always have

$$\frac{1}{2}(1-s) \leq \frac{\mu^2}{\mu^2+\sigma^2} \leq 1 - \frac{(1-s)\sigma^2}{2\mu^2}.$$

Based on the diagram, we see that for any given (μ, σ, s) triplet, the difference between the two policies is clearly only dependent on the ratio c/p. Also observe that the MVS policy is less conservative than the MV policy in the sense that it recommends ordering nothing for a smaller range of c/p values. This degree of conservatism also decreases as s approaches 1. In fact, when s is approximately 1, the optimal policy is to almost always order approximately μ units.

It is also interesting to see how the best-case bound changes for different degrees of asymmetry. For a given quantity q, the best-case profit is given by

$$\begin{aligned} \sup_f \quad & p\mathrm{E}_f\left(\min\{\tilde{d}, q\}\right) - cq \\ \text{s.t.} \quad & \mathrm{E}_f(\tilde{d}) = \mu, \quad \mathrm{E}_f\left((\tilde{d}-\mu)^2\right) = \sigma^2, \\ & \mathrm{E}_f\left((\tilde{d}-\mu)^{+2}\right) - \mathrm{E}_f\left((\mu-\tilde{d})^{+2}\right) = s\sigma^2, \\ & \mathrm{E}_f(1) = 1, \quad f(\tilde{d}) \geq 0, \quad \forall \tilde{d} \geq 0. \end{aligned}$$

Unlike the worst-case profit, the best-case bound is not necessarily a concave function of q. In Appendix A, we show a technique of casting this problem into an equivalent SOCP formulation. If $q \leq \mu$, the best-case profit can be found by solving

$$\begin{aligned}
\inf_{t,r,y_1,y_2,\tau_1,\tau_2,\tau_3} \quad & t + \tfrac{1}{2}(1+s)\sigma^2 y_1 + \tfrac{1}{2}(1-s)\sigma^2 y_2 \\
\text{s.t.} \quad & t - (p-c)q + y_1 \geq \sqrt{(t-(p-c)q - y_1)^2 + (r-\tau_1)^2}, \\
& t - (p-c)q + y_2 + \tau_2 \geq \sqrt{(t-(p-c)q - y_2 - \tau_2)^2 + (r + (\mu - q)\tau_2)^2}, \\
& t - p\mu + cq + \mu(\mu - q)\tau_3 + y_2 + \tau_3 \\
& \qquad \geq \sqrt{(t - p\mu + cq + \mu(\mu-q)\tau_3 - y_2 - \tau_3)^2 + (r - p + (2\mu - q)\tau_3)^2}, \\
& \tau_1, \tau_2, \tau_3 \geq 0.
\end{aligned}$$

However, if $q \geq \mu$, then the best-case profit is found by solving

$$\begin{aligned}
\inf_{t,r,y_1,y_2,\tau_1,\tau_2,\tau_3} \quad & t + \tfrac{1}{2}(1+s)\sigma^2 y_1 + \tfrac{1}{2}(1-s)\sigma^2 y_2 \\
\text{s.t.} \quad & t - p\mu + cq + y_1 + \tau_1 \geq \sqrt{(t - p\mu + cq - y_1 - \tau_1)^2 + (r - p - (q-\mu)\tau_1)^2}, \\
& t - (p-c)q + (q-\mu)\tau_2 + y_1 \\
& \qquad \geq \sqrt{(t - (p-c)q + (q-\mu)\tau_2 - y_1)^2 + (r - \tau_2)^2}, \\
& t - p\mu + cq + y_2 + \tau_3 \geq \sqrt{(t - p\mu + cq - y_2 - \tau_3)^2 + (r - p + \mu\tau_3)^2}, \\
& \tau_1, \tau_2, \tau_3 \geq 0.
\end{aligned}$$

Note that this problem can only be solved as an SOCP if the quantity q is known.

In fact, the method in Appendix A can admit many variations to the newsvendor problem, such as if the demand has a bounded support, or if the newsvendor is risk-averse (Section 3). Different asymmetry models can also handled, such as multiple partitions of the distribution. This is especially useful if, aside from partitioning at the mean, we also include partitions one standard deviation away from the mean in both directions. This gives us a more complete picture of asymmetry than simply using semivariance. In all these variations of the model, the best- and worst-case objective can be solved through an SOCP.

2.3 Sensitivity Analysis

Figure 2.3 plots the different bounds on the newsvendor profit as a function of q. The four different plots correspond to four different values of the normalized semivariance s. The normalized semivariance of the upper left plot is the smallest possible value in which the model is still well-defined (by Proposition 2.1). Scarf and Jensen refer to the worst- and best-case expected profits under the given mean and variance information. MVS Worst and MVS Best are the worst- and best-case expected profits with the additional semivariance information. MVS Best is plotted from the solution of multiple SOCP problems. The other three bounds are found in closed-form. Since the feasible distribution set of the MVS model is a subset of the one in the MV model, then clearly its optimal bounds must be contained between the Scarf and Jensen bounds. We can see that, unlike the static Scarf and Jensen bounds, the MVS bounds are highly dependent on the asymmetry. Moreover, the difference between the MVS bounds becomes small as s approaches its upper and lower limits. One way to view this is that, at the limits of s, the feasible

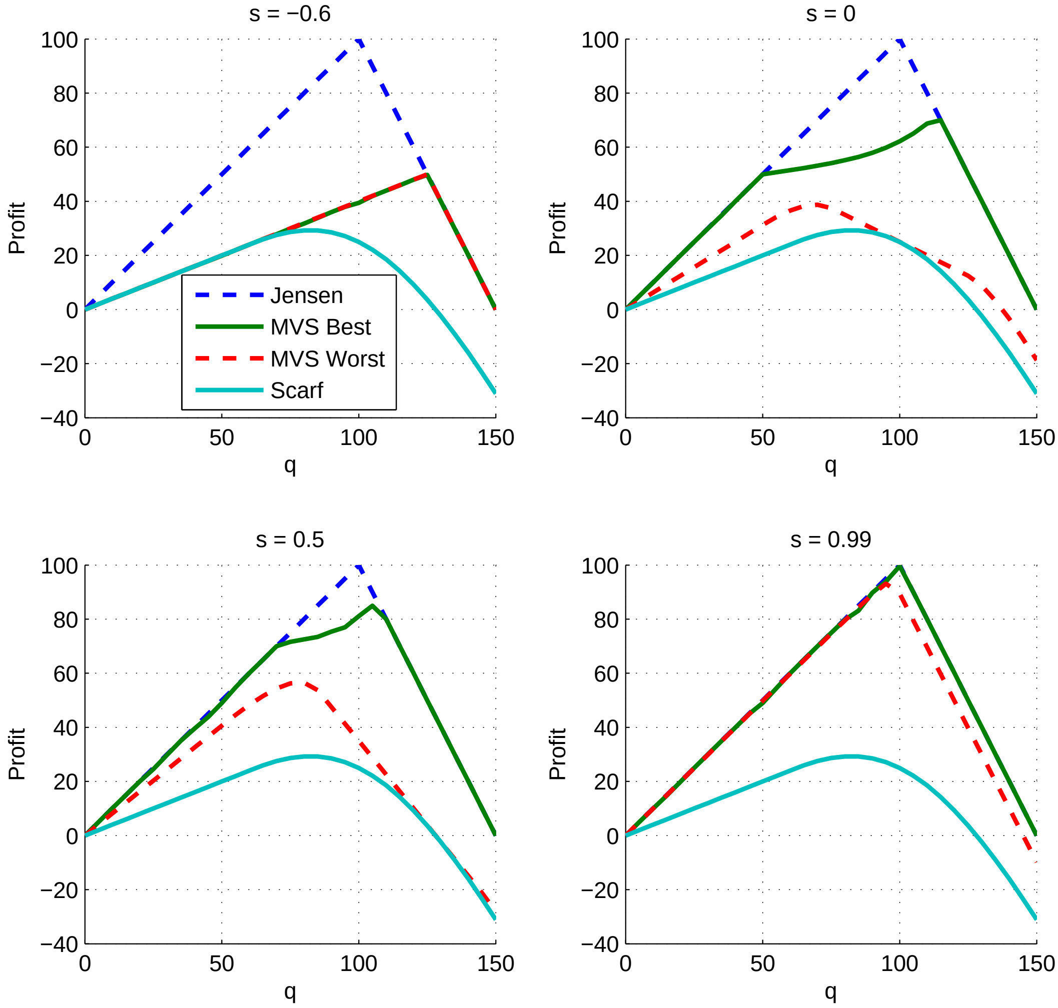

Figure 2.3: Bounds on the expected newsvendor profit ($p = 3, c = 2, \mu = 100, \sigma = 50$).

distribution set becomes more restrictive. 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.

We can perhaps view the difference between the best and worst optimal profits (i.e., the difference between the peaks) as some form of ambiguity risk. Suppose a newsvendor is ambiguity-averse, in the sense that he prefers a sure profit over something unsure but with a potentially higher payoff. Then, the scenario of the parameter s being as high or as low as possible is ideal for him. Put differently, if the distribution has a high degree of asymmetry, then by using the MVS ordering policy, the newsvendor can be fairly certain of getting a nearly optimal profit. This is actually rather intuitive, since a high degree of asymmetry tells us much about the distribution and narrows down possible demand scenarios. Contrast this with the Scarf ordering policy, which optimizes over a much larger set of scenarios, not all of which are probable. Clearly, the model that optimizes the profit while taking into account this

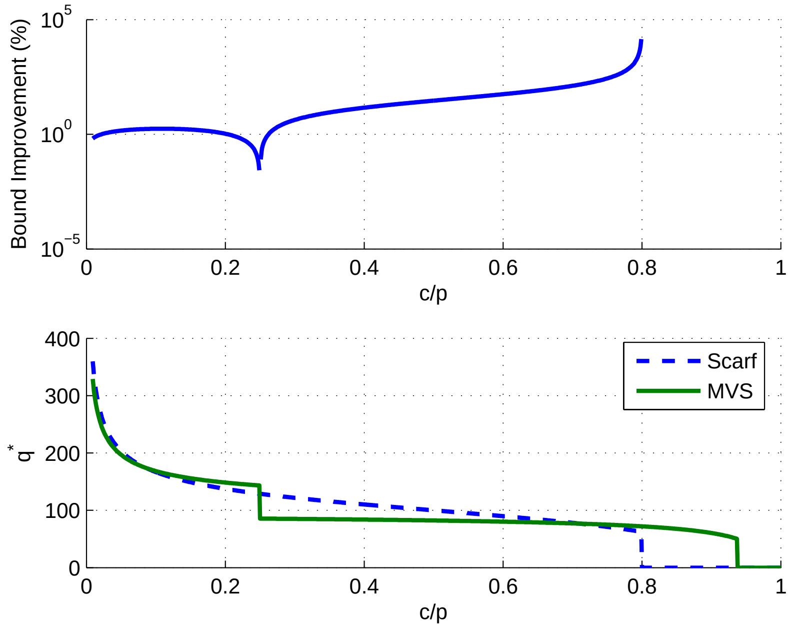

Figure 2.4: Sensitivity of the optimal order quantity and the worst-case bounds against the unit cost to unit price ratio c/p ($s = 0.5, c = 2, \mu = 100, \sigma = 50$).

high degree of asymmetry would fare much better.

Now let us examine how the optimal policies vary as the parameters of the newsvendor model are changed. Figures 2.4 and 2.5 plot the optimal ordering quantity q and percentage bound improvement as a function of the cost to price ratio and semivariance, respectively. The percentage bound improvement simply refers to the percentage increase of optimal MVS profit over the optimal Scarf profit. We can see from Figure 2.4 that the MVS ordering policy follows the same general trend as the Scarf policy over the c/p range. Scarf [33] observes that if $c/p < \frac{1}{2}$, the MV model suggests stocking more than the mean demand. If $c/p > \frac{1}{2}$, the policy is to stock less than the mean demand. The MVS model on the other hand has a slightly altered policy, which suggests stocking less than the mean demand if $c/p > \frac{1}{2}(1-s)$, and more otherwise. We observe that the bound improvement of the MVS model can be extremely large, especially for large values of c/p. We can also see in this figure that the MVS policy is less conservative since q^* is zero only for a small region. The jumps we observe in the plot of q^* indicate that there is in fact a range of quantities that maximize the worst-case profit. In Figure 2.5, we see that the Scarf policy remains static over all degrees of asymmetry. This figure further emphasizes the fact that the bound improvement is greatest at the limits of s. For instance, as s approaches 1, the MVS optimal profit is almost three times the Scarf optimal profit.

2.4 Application: Payoff of a European Call Option

We can apply the worst-case bounds with mean, variance and semivariance to the payoff of an option. Consider a European call option with a strike price K which matures at some future time. Let $\tilde{S}$ be

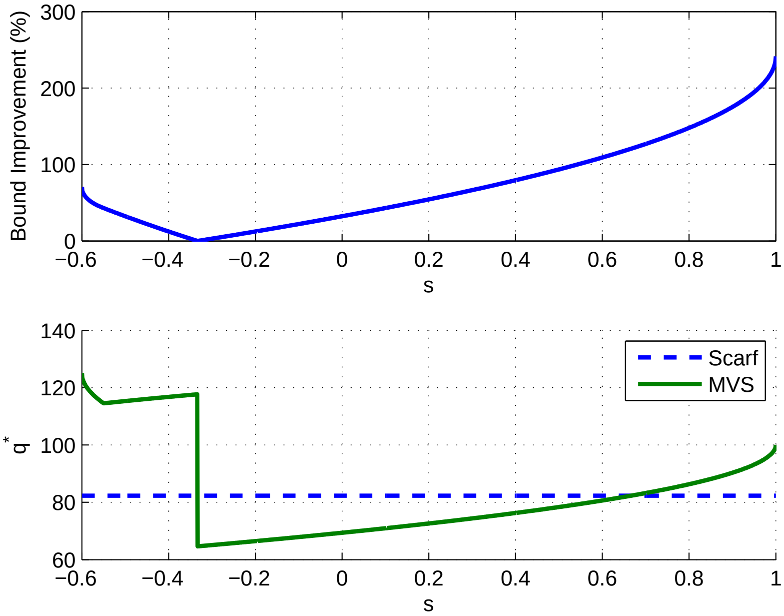

Figure 2.5: Sensitivity of the optimal order quantity and the worst-case bounds against the normalized semivariance s $(p = 3, c = 2, \mu = 100, \sigma = 50)$.

the uncertain stock price at the maturity date. The expected payoff of the option at maturity is

$$\mathrm{E}_f((\tilde{S} - K)^+),$$

where the expectation is taken with respect to the known probability distribution f. We mentioned earlier that the Black-Scholes framework assumes that the underlying stock price $\tilde{S}$ follows a lognormal distribution process. Instead of making such a strong assumption, we assume that the distribution f belongs to a parametric set of distributions of known moment information. Lo [26] finds the tight upper bound on the expected payoff of the call option under mean and variance information. In particular,

$$\mathrm{E}_f((\tilde{S} - K)^+) \leq \begin{cases} \mu - \frac{K\mu^2}{\mu^2+\sigma^2}, & \text{if } K \leq \frac{\mu^2+\sigma^2}{2\mu}, \\ \frac{\mu-K}{2} + \frac{1}{2}\sqrt{(\mu - K)^2 + \sigma^2}, & \text{if } K > \frac{\mu^2+\sigma^2}{2\mu}, \end{cases}$$

for all stock price distributions with mean μ and variance σ^2. This bound is in fact can be derived from the Scarf bound. The tight lower bound on the expected payoff is the Jensen's bound $(\mu - K)^+$. In fact, if the expectation is taken with respect to the risk-neutral stock price distribution, then the option price is simply the expected payoff discounted by the risk-free rate. Thus, if the moments are taken from the risk-neutral stock price distribution, then we can also find a range of option prices by using discounting the payoff bounds by the risk-free rate.

Since one of the stylized facts about the distribution of stock prices is its asymmetrical nature, then it is more reasonable to compute the optimal bounds only against distributions exhibiting asymmetry. A less conservative model must include some form of information about the degree of asymmetry. We assume that the set of distributions can be described by known mean μ, variance σ^2 and normalized semivariance s. Corollary 2.1 provides a tight closed-form expression for the upper bound of a call option's expected payoff. Corollary 2.2 on the other hand provides an SOCP formulation for the problem of finding the optimal lower bound for the expected payoff. These are direct corollaries from our results for the newsvendor model by setting $q = K$, $p = 1$ and $c = 0$. Similar bounds can also be found for the expected payoff of a European put option. If these moments are taken with respect to the risk-neutral distribution, then the corollaries also provide an upper and lower bound on the current option price.

Corollary 2.1. *Consider a European call option with a strike price K maturing at some future time. Suppose the distribution of the stock price at maturity has mean μ, standard deviation σ and normalized semivariance s. Then an upper bound for the expected payoff of the option is*

$$\begin{cases} \mu - K + \frac{(1-s)\sigma^2}{2\mu^2}K, & \text{if } K \in \left[0, \frac{\mu}{2}\right], \\ \mu - K + \frac{(1-s)\sigma^2}{8(\mu-K)}, & \text{if } K \in \left[\frac{\mu}{2}, \mu - \frac{\sigma}{2}\sqrt{\frac{1-s}{1+s}}\right], \\ \frac{(1-s)}{2}(\mu - K) + \frac{\sigma}{2}\sqrt{1-s^2}, & \text{if } K \in \left[\mu - \frac{\sigma}{2}\sqrt{\frac{1-s}{1+s}}, \mu + \frac{\sigma}{2}\sqrt{\frac{1+s}{1-s}}\right], \\ \frac{(1+s)\sigma^2}{8(K-\mu)}, & \text{if } K \in \left[\mu + \frac{\sigma}{2}\sqrt{\frac{1+s}{1-s}}, \mu + \frac{\mu(1+s)}{2(1-s)}\right], \\ \frac{\mu}{2} - \frac{bK}{2} + \frac{1}{2}\sqrt{(bK-\mu)^2 - (1-b)^2\mu^2 + \frac{(1+s)\sigma^2 b}{2}}, & \text{if } K \in \left[\mu + \frac{\mu(1+s)}{2(1-s)}, \infty\right), \end{cases}$$

where

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

Moreover, among the set of nonnegative distributions parameterized by (μ, σ, s), there exists a distribution with at most three support points that achieves this bound.

Corollary 2.2. *Consider a European call option with a strike price K maturing at some future time. Suppose the distribution of the stock price at maturity has mean μ, standard deviation σ and normalized semivariance s. If $K \leq \mu$, then a lower bound on the expected payoff is found by solving*

$$\begin{aligned} \sup_{t,r,y_1,y_2,\tau_1,\tau_2,\tau_3} \quad & \mu - t - \tfrac{1}{2}(1+s)\sigma^2 y_1 - \tfrac{1}{2}(1-s)\sigma^2 y_2 \\ \text{s.t.} \quad & t - K + y_1 \geq \sqrt{(t - K - y_1)^2 + (r - \tau_1)^2}, \\ & t - K + y_2 + \tau_2 \geq \sqrt{(t - K - y_2 - \tau_2)^2 + (r + (\mu - K)\tau_2)^2}, \\ & t - \mu + \mu(\mu - K)\tau_3 + y_2 + \tau_3 \\ & \qquad \geq \sqrt{(t - \mu + \mu(\mu - K)\tau_3 - y_2 - \tau_3)^2 + (r - 1 + (2\mu - K)\tau_3)^2}, \\ & \tau_1, \tau_2, \tau_3 \geq 0. \end{aligned}$$

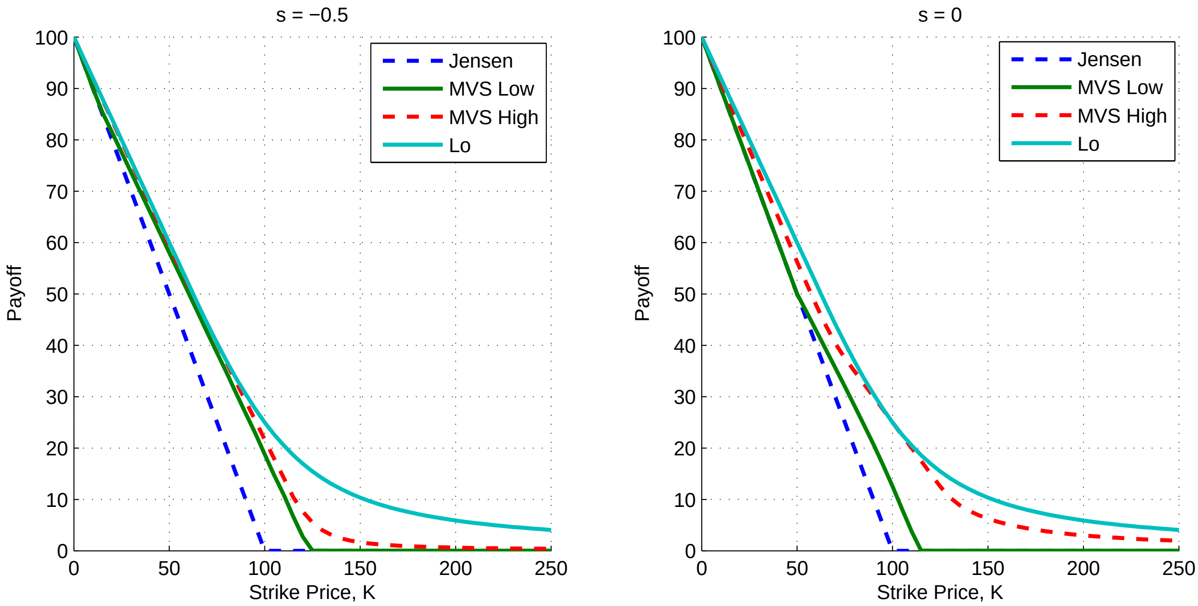

Figure 2.6: Bounds on expected payoff of a call option ($\mu = 100, \sigma = 50$).

Otherwise, if $K \geq \mu$, then a lower bound is found by solving

$$\begin{aligned}
\sup_{t,r,y_1,y_2,\tau_1,\tau_2,\tau_3} \quad & \mu - t - \tfrac{1}{2}(1+s)\sigma^2 y_1 - \tfrac{1}{2}(1-s)\sigma^2 y_2 \\
s.t. \quad & t - \mu + y_1 + \tau_1 \geq \sqrt{(t-\mu-y_1-\tau_1)^2 + (r-1-(K-\mu)\tau_1)^2}, \\
& t - K + (K-\mu)\tau_2 + y_1 \geq \sqrt{(t-K+(K-\mu)\tau_2 - y_1)^2 + (r-\tau_2)^2}, \\
& t - \mu + y_2 + \tau_3 \geq \sqrt{(t-\mu-y_2-\tau_3)^2 + (r-1+\mu\tau_3)^2}, \\
& \tau_1, \tau_2, \tau_3 \geq 0.
\end{aligned}$$

Figure 2.6 plots the upper and lower bounds on the call option's payoff. The plot on the left, which corresponds to a model with asymmetry, shows that for a fixed strike price, the range of payoffs (i.e. the difference between the bounds) is small under the MVS model. On the other hand, the second plot shows the two bounds being further apart. In both cases however, the range of possible payoffs under the MV model is much larger. What we conclude from this example is that if the stock price distribution is known to be asymmetric, then there is merit in using the MVS model to price an option, since it can narrow down the range of possible future expected payoffs without being too restrictive in its assumptions.

3 Ambiguity-Averse, Risk-Averse Newsvendor

In Section 2, the newsvendor is assumed to be risk-neutral whose primary concern is in maximizing the expected profit. Intuitively however, we expect knowledge of asymmetry to greatly affect the policies of a risk-averse newsvendor who greatly penalizes losses. A newsvendor who is risk-averse evaluates the quality of the ordering policy based on the risk he faces on the random loss. Is there a connection between a newsvendor's utility and how he defines risk? Ben-Tal and Teboulle [4],[5] explicitly define a connection between utility and risk by introducing the notion of the *optimized certainty equivalent* (OCE). For a random loss $\tilde{x}$ and a normalized concave utility function u, the OCE is defined as:

$$S_u(\tilde{x}) = \sup_{v \in \Re} \left\{ -v + \mathrm{E}_f \left(u(v - \tilde{x}) \right) \right\}.$$

The OCE can be interpreted as the sure present value of a future uncertain income $v - \tilde{x}$. Suppose a newsvendor expects an uncertain future profit of $v - \tilde{x}$ by investing part of it at present. If he chooses to invest v, the resulting present value is then $-v + \mathrm{E}\left(u(v - \tilde{x})\right)$. The optimized certainty equivalent is then a result of an optimal allocation of the payoffs between present and future consumption. The OCE risk is defined as:

$$\rho_u(\tilde{x}) = -S_u(\tilde{x}).$$

Consider the general class of functions $u(x) : \Re \mapsto [-\infty, \infty)$ that are proper, closed, concave and nondecreasing utility functions with effective domain $\mathrm{dom}(u) = \{t \in \Re : u(t) > -\infty\} \neq \emptyset$. Assume that the utility function satisfies the properties

$$u(0) = 0 \quad \text{and} \quad 1 \in \partial u(0),$$

where $\partial u(\cdot)$ denotes the subdifferential map of u. It is shown in [4] that for this class of utility functions, $\rho_u(\tilde{x})$ is a convex risk measure, in the sense that it satisfies a set of axiomatic properties (see Fölmer and Schied [18]; Fritelli and Gianin [19]). In particular, it meets the conditions of monotonicity, translation invariance and convexity. Moreover, for piecewise linear utility functions with two pieces of the form

$$u(x) = \begin{cases} \gamma_2 x, & \text{if } x \leq 0, \\ \gamma_1 x, & \text{if } x > 0, \end{cases}$$

for some $\gamma_2 > 1 > \gamma_1 \geq 0$, $\rho_u(\tilde{x})$ defines a coherent risk measure, in the sense that it satisfies a more stringent set of axiomatic properties (see Artzner et al. [1]). That is, in addition to satisfying the previous properties, it is also positive homogeneous. The following definition specializes OCE risk measures for the class of piecewise linear utility functions.

Definition 3.1. *Let $u(x) = \min_{k=1,\ldots,K} \{a_k x + b_k\}$ be a piecewise linear concave utility function satisfying the following three properties:*

1. $a_1 > a_2 > \cdots > a_l \geq 1 \geq a_{l+1} > \cdots > a_K \geq 0$,
2. $b_k \geq 0$ *for* $k = 1, \ldots, K$ *and* $b_l = 0$,
3. $a_l > 1$ *implies that* $b_{l+1} = 0$.

Property 1 implies that the utility is a piecewise linear increasing function. Property 2 implies that the utility is zero at $x = 0$ and this is attained at the lth piece. Properties 1, 2 and 3 ensure that the value 1 is a subgradient for the utility function at $x = 0$. For a piecewise utility function that satisfies Definition 3.1, the corresponding OCE risk

$$\rho_u(\tilde{x}) = \inf_{v\in\Re}\left\{v - \mathrm{E}_f\left(\min_{k=1,\ldots,K}\{a_k(v-\tilde{x}) + b_k\}\right)\right\}$$

is a convex risk measure.

An implicit assumption of the OCE risk is that the distribution f of the random loss $\tilde{x}$ must be known for the expectation to be defined. Again, this assumption seems unrealistic, since in most practical applications, the distribution of a random variable is unknown. Instead, we make a more conservative assumption that the actual distribution f lies in a parametric set of distributions $\mathbb{F}$. The worst-case OCE risk is then defined as:

$$\inf_{v\in\Re}\left\{v - \inf_{f\in\mathbb{F}}\mathrm{E}_f\left(\min_{k=1,\ldots,K}\{a_k(v-\tilde{x}) + b_k\}\right)\right\}. \tag{3.1}$$

If $\tilde{x}$ represents the loss of a newsvendor, an ambiguity-averse newsvendor will choose to minimize the worst-case OCE risk. We can think of the worst-case OCE risk as an ambiguous risk measure since it is defined over a set of distributions. Ambiguous risk measures have been studied in the context of portfolio management (Calafiore [10]; Natarajan, Sim and Uichanco [29]). Calafiore [10] considered the ambiguous variance of a portfolio defined over the set of distributions lying within some distance from a nominal one. On the other hand, Natarajan et al. [29] studied the worst-case OCE risk of portfolio returns when the multivariate distribution of asset returns is assumed to have a known mean and covariance matrix. With a positive and negative partitioning of the random returns, they provide a bound on the worst-case OCE risk (which is not necessarily tight) in the form of a compact SOCP.

In fact, we can provide a tight bound on the worst-case OCE risk owing to the one dimensional nature of the random variable in our problem. Let $\tilde{x}$ be the difference between some benchmark M and the random newsvendor profit for a given quantity q. That is,

$$\tilde{x} = M - p\min\{\tilde{d}, q\} + cq.$$

A possible benchmark can be $(p-c)\mu$, which is the optimal expected newsvendor profit under the scenario that the newsvendor can place the order after the demand is observed. Therefore, $\tilde{x}$ represents the cost of not knowing the demand beforehand. We can then write (3.1) as

$$\inf_{v\in\Re}\left\{v - \inf_{f\in\mathbb{F}}\mathrm{E}_f\left(\min_{k=1,\ldots,K}\left\{\min\left(a_k(v - M + p\tilde{d} - cq) + b_k, a_k(v - M + (p-c)q) + b_k\right)\right\}\right)\right\}.$$

The expression inside the expectation operator can be thought of simply as a piecewise utility function on $\tilde{d}$ with $2K$ linear pieces. Suppose $\mathbb{F}$ consists of the distributions of $\tilde{d}$ with mean μ, variance σ^2, normalized semivariance s and nonnegative support. In other words, we partition $\tilde{d}$ at μ such that $\tilde{d} - \mu = (\tilde{d}-\mu)^+ - (\mu - \tilde{d})^+$. We can use the technique in Appendix A to convert the problem into an

SOCP. The optimal quantity q that minimizes the worst-case OCE risk of the newsvendor profit can be found by solving:

$$
\begin{aligned}
\inf_{q,v,t,r,y_1,y_2,\lambda_i^k,\tau_i^k} \quad & v - t - \tfrac{1}{2}(1+s)\sigma^2 y_1 - \tfrac{1}{2}(1-s)\sigma^2 y_2 \\
\text{s.t.} \quad & a_k(v - M + p\mu - cq) + b_k - t - y_1 \\
& \quad \geq \sqrt{(a_k(v - M + p\mu - cq) + b_k - t + y_1)^2 + \left(a_k p - r - \tau_1^k\right)^2}, \quad \forall k = 1, \ldots, K, \\
& a_k(v - M + (p-c)q) + b_k - t - y_1 \\
& \quad \geq \sqrt{(a_k(v - M + (p-c)q) + b_k - t + y_1)^2 + \left(-r - \lambda_1^k\right)^2}, \quad \forall k = 1, \ldots, K, \\
& a_k(v - M + p\mu - cq) + b_k - t - y_2 + \tau_2^k \\
& \quad \geq \sqrt{\left(a_k(v - M + p\mu - cq) + b_k - t + y_2 - \tau_2^k\right)^2 + \left(-a_k p + r - \mu\tau_2^k\right)^2}, \\
& \qquad \qquad \forall k = 1, \ldots, K, \\
& a_k(v - M + (p-c)q) + b_k - t - y_2 + \lambda_2^k \\
& \quad \geq \sqrt{\left(a_k(v - M + (p-c)q) + b_k - t + y_2 - \lambda_2^k\right)^2 + \left(r - \mu\lambda_2^k\right)^2}, \quad \forall k = 1, \ldots, K, \\
& \lambda_i^k, \tau_i^k \geq 0, \quad \forall i = 1, 2, \ \forall k = 1, \ldots, K.
\end{aligned}
$$

Note that by setting $v = 0$ and $M = 0$, the problem is equivalent to maximizing the worst-case expected utility of the newsvendor profit. The OCE risk has a general form and can admit a wide variety of risk measures. We have been assuming throughout this section that the underlying utility function has a piecewise linear form. In fact, we can approximate general nonlinear utilities using a piecewise linear function. In other words, our approach also provides an approximate method to optimize the worst-case OCE risk if the underlying utility is a nonlinear function. In the following example, we show an application of our method to a popular risk measure.

3.1 Example: Conditional Value-at-Risk

The conditional value-at-risk (CVaR) is a popular risk measure in finance. It is related to the value-at-risk (VaR) measure. By definition, with respect to a specific probability level α, the α-VaR is the lowest amount of loss v such that, with probability α, the loss will exceed that amount v. The α-CVaR on the other hand is the conditional expectation of losses above that amount v. The conditional value-at-risk of a random loss $\tilde{x}$ can be written as

$$
\mathrm{CVaR}_\alpha(\tilde{x}) = \inf_{v \in \Re} \left\{ v + (1-\alpha)^{-1} \mathrm{E}_f \left(\max\{0, \tilde{x} - v\} \right) \right\},
$$

for some risk tolerance parameter $\alpha \in (0, 1)$. Values of α commonly considered are $0.90, 0.95$ and 0.99. We can define the loss as the difference between some benchmark M and the newsvendor profit, that is $\tilde{x} = M - p\min\{\tilde{d}, q\} + cq$. For a given order quantity q, the worst-case CVaR of the newsvendor profit is:

$$
\inf_{v \in \Re} \left\{ v - (1-\alpha)^{-1} \inf_{f \in \mathbb{F}} \mathrm{E}_f \left(\min\left\{ 0, v - M + p\tilde{d} - cq, v - M + (p-c)q \right\} \right) \right\}.
$$

Suppose $\mathbb{F}$ consists of all distributions f with known mean μ, variance σ^2, normalized semivariance s and nonnegative support. Then the problem of minimizing the worst-case α-CVaR is equivalent to solving the following SOCP problem:

$$
\begin{aligned}
\inf_{q,v,t,r,y_1,y_2,\tau_i^k} \quad & v - \frac{1}{1-\alpha}\left(t + \tfrac{1}{2}(1+s)\sigma^2 y_1 + \tfrac{1}{2}(1-s)\sigma^2 y_2\right) \\
\text{s.t.} \quad & -t - y_1 \geq \sqrt{(-t+y_1)^2 + \left(-r-\tau_1^1\right)^2}, \\
& v - M + p\mu - cq - t - y_1 \geq \sqrt{(v - M + p\mu - cq - t + y_1)^2 + \left(p - r - \tau_1^2\right)^2}, \\
& v - M + (p-c)q - t - y_1 \geq \sqrt{(v - M + (p-c)q - t + y_1)^2 + \left(-r - \tau_1^3\right)^2}, \\
& -t - y_2 + \tau_2^1 \geq \sqrt{\left(-t + y_2 - \tau_2^1\right)^2 + \left(r - \mu\tau_2^1\right)^2}, \\
& v - M + p\mu - cq - t - y_2 + \tau_2^2 \geq \sqrt{\left(v - M + p\mu - cq - t + y_2 - \tau_2^2\right)^2 + \left(-p + r - \mu\tau_2^2\right)^2}, \\
& v - M + (p-c)q - t - y_2 + \tau_2^3 \geq \sqrt{\left(v - M + (p-c)q - t + y_2 - \tau_2^3\right)^2 + \left(r - \mu\tau_2^3\right)^2}, \\
& \tau_i^k \geq 0, \qquad i = 1,2,\ k = 1,2,3.
\end{aligned}
$$

In fact, we can also find an SOCP formulation for the problem of minimizing the worst-case CVaR under just mean and variance information using the same technique (see Appendix A for details).

Figures 4.1 and 4.2 show how the optimal policy of a risk-averse newsvendor changes as the parameters of the model vary. As a point of reference, the optimal policy of a risk-neutral newsvendor under the MVS model is also plotted. The optimal risk-averse policies are found by solving SOCP formulations. Figure 4.1 shows how the policies of four newsvendors (with varying risk preferences) changes as the cost to price ratio increases. The plot on the top simply shows the optimal bounds which appear to be steadily decreasing as a function of c/p. Let us focus our attention on the second plot, which shows the optimal ordering quantities. Observe that for small c/p ratios, the optimal policies of the four newsvendors are almost similar. However, as the cost to price ratio increases, the risk-averse newsvendors become more and more conservative. The newsvendor with the lowest risk tolerance ($\alpha = 0.95$) is the most conservative of the four. The less risk-averse the newsvendor becomes, the more his policy resembles the risk-neutral policy.

Now let us examine Figure 4.2, which shows how the policy changes as the normalized semivariance increases. Observe that, for a wide range of values of s, the highly risk-averse newsvendors ($\alpha = 0.95, 0.90$) adopt a conservative policy of not ordering anything. Note that when s is small, the distribution's variance is mostly concentrated on the interval below the mean. Intuitively, a distribution that is negatively skewed must have a small s value. Put differently, if the demand distribution is negatively skewed, then a risk-averse newsvendor should choose not to order. This is in huge contrast with the risk-neutral newsvendor who in fact orders more when s is small. Intuitively, this makes sense because, when there is a large variance for small levels of demand, then we expect the newsvendor profit to also have a high variance. Then a risk-averse newsvendor might choose to protect against this risk by ordering nothing. We also compare the MVS policies and the policies under mean and variance information in Figure 4.3. The leftmost plots correspond to the policies of a less risk-averse newsvendor. We observe that the mean-variance policies are rather counterintuitive, since they recommend conservatism for less risk-averse newsvendors, even if the distribution is positively skewed. The MVS model on the other hand recommends adaptive policies that are appropriate for the level of risk tolerance.

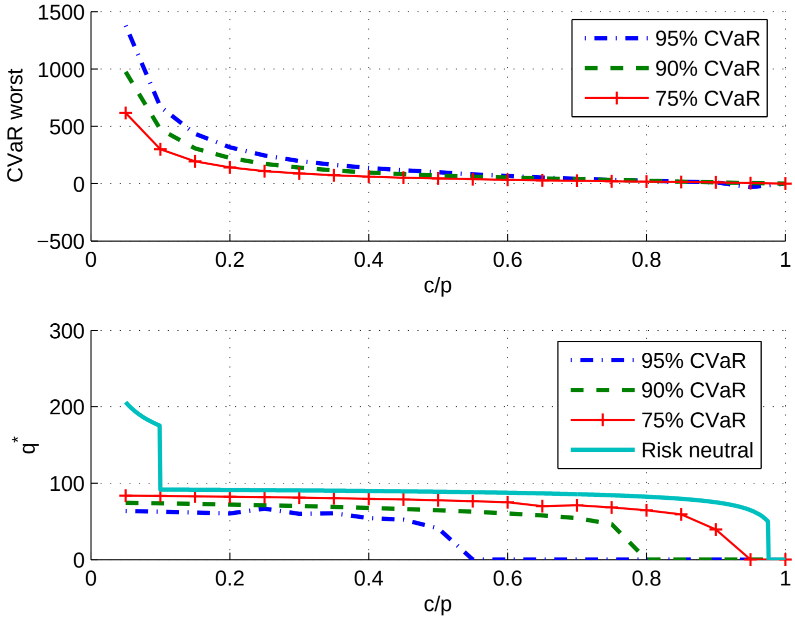

Figure 4.1: Sensitivity of the optimal order quantity and the worst-case CVaR against the unit cost to unit price ratio c/p $(s = 0.8, c = 1, \mu = 100, \sigma = 50, M = (p - c)\mu)$.

4 Conclusion

We have introduced a decision-making framework under which the expectation of a piecewise linear objective function is optimized over a set of distributions that have similar asymmetry properties. In particular, this asymmetry is represented by the first two moments of multiple random variables that result from partitioning the support of the original distribution. The method can be applied to several variants of the newsvendor problem and to option pricing. In a risk-neutral setting, a closed-form expression for the worst-case newsvendor profit can be found under mean, variance and semivariance information. In fact, due to the second-order nature of our proposed model, closed-form expressions for optimal bounds can generally be found. We have also shown through illustrative examples that if the distribution is known to have a high degree of asymmetry, then the upper and lower bounds on our proposed model can be very close. This suggests that our model is useful especially for applications where the underlying distribution is known to be asymmetric.

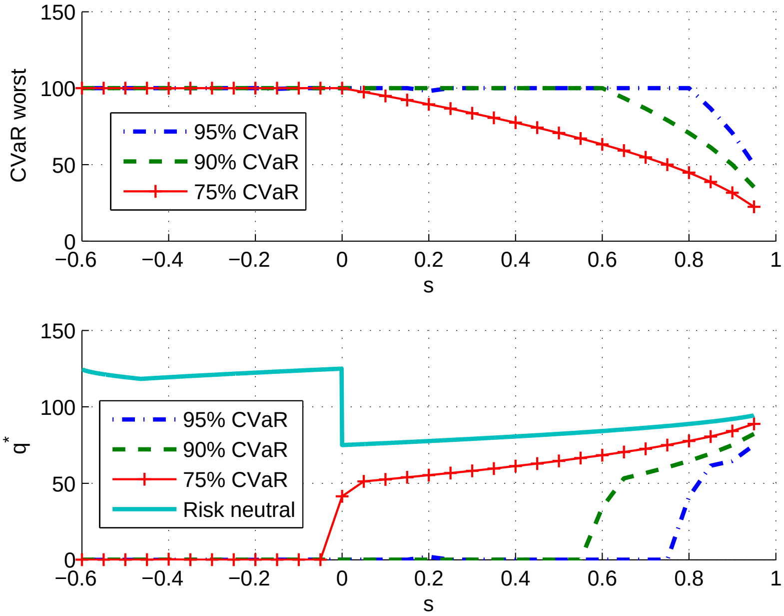

Figure 4.2: Sensitivity of the optimal order quantity and the worst-case CVaR against the normalized semivariance s ($p = 2, c = 1, \mu = 100, \sigma = 50, M = (p - c)\mu$).

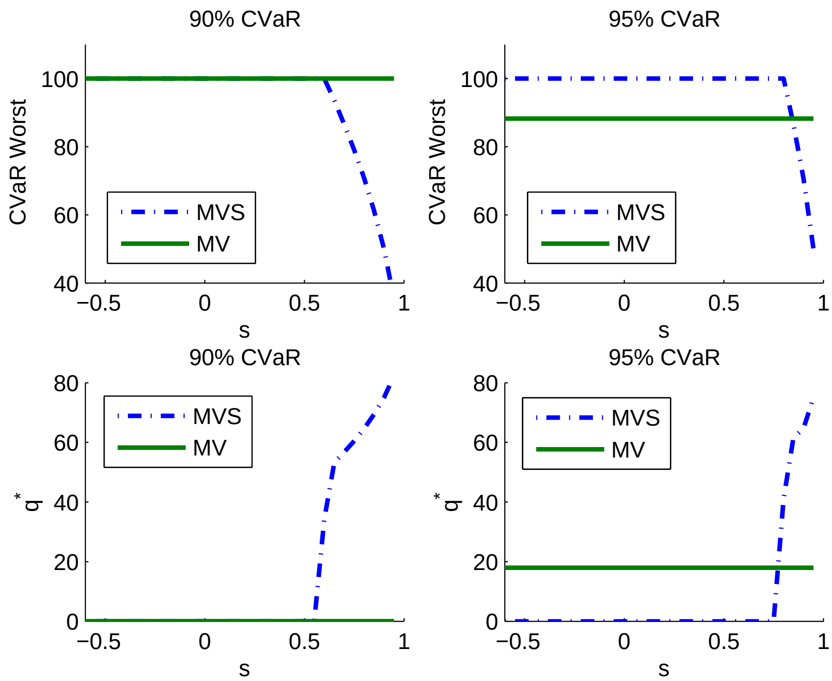

Figure 4.3: Comparison between mean-variance and mean-variance-semivariance risk-averse policies ($p = 2, c = 1, \mu = 100, \sigma = 50, M = (p - c)\mu$).

Appendix A General Partitioned Model

Consider a single random variable $\tilde{x}$ with a support set $A \subseteq \Re$. We can partition the support set in the following manner. Let $\{A_i\}_{i=1}^p$ being non-overlapping intervals whose union is A (see Figure A.1). Therefore,

$$\tilde{x} = \sum_{i=1}^{p} h_i(\tilde{x}),$$

where

$$h_i(x) = \begin{cases} x, & \text{if } x \in A_i, \\ 0, & \text{otherwise.} \end{cases}$$

We can view each $h_i(\tilde{x})$ as a new partitioned random variable. If we have limited information on the distribution of $\tilde{x}$, say, the mean and variance, then this gives no light as to any asymmetry information. However, introducing the p partitioned random variables gives us an idea of how distribution is divided among the support subintervals. Consider the expectation over a piecewise linear function of $\tilde{x}$

$$\mathrm{E}_f\left(\min_{k=1,\ldots,K}\{a_k\tilde{x}+b_k\}\right).$$

If $\tilde{x}$ represents a random payoff, we can interpret the piecewise linear function as some form of utility function. Using the partitioning, we can rewrite the expectation as

$$\mathrm{E}_f\left(\min_{k=1,\ldots,K}\left\{a_k\sum_{i=1}^{p} h_i(\tilde{x})+b_k\right\}\right).$$

If we have mean and variance information on each of the partitioned random variables, then a robust approach is to optimize the objective subject to the known information. Consider the problem of minimizing the objective, that is

$$\begin{aligned} GP_m \quad = \quad \inf_f \quad & \mathrm{E}_f\left(\min_{k=1,\ldots,K}\left\{a_k\sum_{i=1}^{p} h_i(\tilde{x})+b_k\right\}\right) \\ \text{s.t.} \quad & \mathrm{E}_f(h_i(\tilde{x})) = \mu_i, \quad \forall i = 1,\ldots,p \\ & \mathrm{E}_f(h_i(\tilde{x})^2) = \mu_i^2+\sigma_i^2, \quad \forall i = 1,\ldots,p, \\ & \mathrm{E}_f(1) = 1, \\ & f(\tilde{x}) \geq 0, \quad \forall \tilde{x} \in A. \end{aligned}$$

We can think of the MVS model as a special case of the partitioning with $p = 2$. But for a model with more than two partitions, a valid question remains: how do you define reasonable partitions? We believe that partitions can naturally result from the model. For instance, they may come from quantiles of the empirical distribution. This choice of partitioning not only has theoretical value, but also leads to “fair” estimates of the moments given finite realizations of $\tilde{x}$. In other words, each partitioned random variable will have moment estimates using roughly the same number of samples. This enforces estimation errors to be spread out among the partitions.

By Isii's [22] strong duality theorem, if the moments of the problem lie strictly in the interior of the feasible moment cone, then the problem is equivalent to the dual formulation

$$\begin{array}{ll} \sup\limits_{t,r_i,y_i} & t + \sum\limits_{i=1}^{p} r_i\mu_i + \sum\limits_{i=1}^{p} y_i(\mu_i^2 + \sigma_i^2) \\ \text{s.t.} & t + r_i x + y_i x^2 \leq a_k x + b_k, \quad \forall x \in A_i,\ \forall i = 1,\ldots,p,\ \forall k = 1,\ldots,K. \end{array} \tag{A.1}$$

There are a total of Kp constraints in the dual problem. In the previous section, we found a closed-form expression to the dual problem in the simple case of $K = 2$ and $p = 2$. Closed-form expressions may prove to be complicated for models with multiple partitions or piecewise functions with multiple linear pieces. However, we show that we can in fact write the dual problem into a second-order cone programming problem by invoking the widely used S-lemma, which we state here for completeness.

Proposition A.1 (S-lemma). *Consider two quadratic functions of $\boldsymbol{z} \in \Re^N$, $q_i(\boldsymbol{z}) = \boldsymbol{z}'\boldsymbol{B}_i\boldsymbol{z} + 2\boldsymbol{b}_i'\boldsymbol{z} + c_i$, $i = 0, 1$, with $q_1(\bar{\boldsymbol{z}}) > 0$ for some $\bar{\boldsymbol{z}}$. Then*

$$q_0(\boldsymbol{z}) \geq 0 \quad \forall \boldsymbol{z} \in \{\boldsymbol{z} : q_1(\boldsymbol{z}) \geq 0\}$$

if and only if there exists $\tau \geq 0$ such that

$$\begin{pmatrix} c_0 & \boldsymbol{b}_0' \\ \boldsymbol{b}_0 & \boldsymbol{B}_0 \end{pmatrix} - \tau \begin{pmatrix} c_1 & \boldsymbol{b}_1' \\ \boldsymbol{b}_1 & \boldsymbol{B}_1 \end{pmatrix} \succeq 0.$$

Each interval $A_i \subset \Re$ can either be bounded, unbounded above, or unbounded below. We show how to convert the constraints in (A.1) to a second-order constraints if the interval is bounded. The other two cases can be handled in a similar manner. Suppose $A_i = [\underline{x}_i, \overline{x}_i]$, where $\underline{x}_i < \overline{x}_i$ and $\underline{x}_i, \overline{x}_i$ are finite. Then the set of all x that belong in A_i can be written as

$$\{x \in \Re : (x - \underline{x}_i)(x - \overline{x}_i) \leq 0\}.$$

By directly applying S-lemma, we can write the K constraints

$$t + r_i x + y_i x^2 \leq a_k x + b_k, \quad \forall x \in A_i,\ \forall k = 1,\ldots,K$$

as the positive semidefinite constraints

$$\begin{pmatrix} b_k - t & \dfrac{a_k - r_i}{2} \\ \dfrac{a_k - r_i}{2} & -y_i \end{pmatrix} - \tau_i^k \begin{pmatrix} -\underline{x}_i\overline{x}_i & \dfrac{\underline{x}_i + \overline{x}_i}{2} \\ \dfrac{\underline{x}_i + \overline{x}_i}{2} & -1 \end{pmatrix} \succeq 0, \quad \tau_i^k \geq 0, \quad \forall k = 1,\ldots,K.$$

Notice that the constraints only require 2×2 matrices to be positive semidefinite. In fact, we can rewrite each of the K positive semidefiniteness constraints into the following simpler form:

$$\begin{aligned} & b_k - t + \underline{x}_i\overline{x}_i\tau_i^k \geq 0, \\ & \left(b_k - t + \underline{x}_i\overline{x}_i\tau_i^k\right)\left(-y_i + \tau_i^k\right) \geq \left(\frac{a_k - r_i - (\underline{x}_i + \overline{x}_i)\tau_i^k}{2}\right)^2, \end{aligned}$$

for $k = 1, \dots, K$. The last inequality simply means that the determinant of the matrix must be nonnegative. Note that this constraint includes a hyperbolic term. We can further cast the constraints into the following second-order constraints (see Lobo et al. [27]):

$$b_k - t + \underline{x}_i \overline{x}_i \tau_i^k - y_i + \tau_i^k \geq \sqrt{\left(b_k - t + \underline{x}_i \overline{x}_i \tau_i^k + y_i - \tau_i^k\right)^2 + \left(a_k - r_i - (\underline{x}_i + \overline{x}_i)\tau_i^k\right)^2}, \quad k = 1, \dots, K.$$

By a similar idea, if A_i is unbounded above with $A_i = [\underline{x}_i, \infty)$, then the corresponding K constraints in the dual form (A.1) can be written as:

$$b_k - t + \underline{x}_i \tau_i^k - y_i \geq \sqrt{\left(b_k - t + \underline{x}_i \tau_i^k + y_i\right)^2 + \left(a_k - r_i - \tau_i^k\right)^2}, \qquad k = 1, \dots, K.$$

On the other hand, if $A_i = (\infty, \overline{x}_i]$, then the corresponding constraints are:

$$b_k - t - \underline{x}_i \tau_i^k - y_i \geq \sqrt{\left(b_k - t - \underline{x}_i \tau_i^k + y_i\right)^2 + \left(a_k - r_i + \tau_i^k\right)^2}, \qquad k = 1, \dots, K.$$

Finally, we can write the dual problem as an SOCP. Let I_0 be the set of all indices i for which A_i is bounded. Also denote I_∞ (and $I_{-\infty}$) for indices in which A_i is unbounded above (and below). Thus, minimization problem GP_m can be solved through the following SOCP:

$$\begin{aligned}
\sup_{t, r_i, y_i, \tau_i^k} \quad & t + \sum_{i=1}^p r_i \mu_i + \sum_{i=1}^p y_i (\mu_i^2 + \sigma_i^2) \\
\text{s.t.} \quad & b_k - t + \underline{x}_i \overline{x}_i \tau_i^k - y_i + \tau_i^k \geq \sqrt{\left(b_k - t + \underline{x}_i \overline{x}_i \tau_i^k + y_i - \tau_i^k\right)^2 + \left(a_k - r_i - (\underline{x}_i + \overline{x}_i)\tau_i^k\right)^2}, \\
& \qquad \forall i \in I_0,\ \forall k = 1, \dots, K, \\
& b_k - t + \underline{x}_i \tau_i^k - y_i \geq \sqrt{\left(b_k - t + \underline{x}_i \tau_i^k + y_i\right)^2 + \left(a_k - r_i - \tau_i^k\right)^2}, \\
& \qquad \forall i \in I_\infty,\ \forall k = 1, \dots, K, \\
& b_k - t - \underline{x}_i \tau_i^k - y_i \geq \sqrt{\left(b_k - t - \underline{x}_i \tau_i^k + y_i\right)^2 + \left(a_k - r_i + \tau_i^k\right)^2}, \\
& \qquad \forall i \in I_{-\infty},\ \forall k = 1, \dots, K, \\
& \tau_i^k \geq 0, \quad \forall i = 1, \dots, K,\ \forall k = 1, \dots, K.
\end{aligned}$$

The SOCP formulation can be solved efficiently using interior point methods (see Nesterov and Nemirovski [30]). Its optimal solution can be found using modern solvers such as CPLEX, MOSEK, SDPT3 or SeDuMi.

In fact, with some minor modifications, we can also find an equivalent SOCP formulation for the problem of maximizing the objective:

$$\begin{aligned}
GP_M \quad = \quad \sup_f \quad & \mathrm{E}_f \left(\min_{k=1,\dots,K} \left\{ a_k \sum_{i=1}^p h_i(\tilde{x}) + b_k \right\} \right) \\
\text{s.t.} \quad & \mathrm{E}_f(h_i(\tilde{x})) = \mu_i, \qquad \forall i = 1, \dots, p \\
& \mathrm{E}_f(h_i(\tilde{x})^2) = \mu_i^2 + \sigma_i^2, \quad \forall i = 1, \dots, p, \\
& \mathrm{E}_f(1) = 1, \\
& f(\tilde{x}) \geq 0, \quad \forall \tilde{x} \in A.
\end{aligned}$$

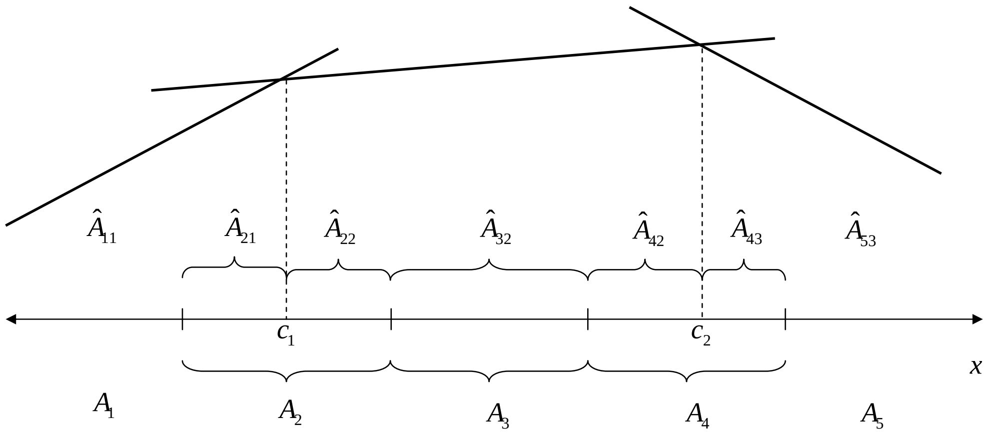

Figure A.1: Example of a partition for $A = \Re$ with $K = 3$ and $p = 5$.

We can assume without loss of generality that $a_1 > a_2 > \cdots > a_K$. Let c_k be the point of intersection between the kth and $(k+1)$th linear piece, i.e. $a_k c_k + b_k = a_{k+1} c_k + b_{k+1}$. Let us define new subintervals $\hat{A}_{ik} = A_i \cap [c_{k-1}, c_k]$, for $i = 1, \ldots, p$ and $k = 1, \ldots, K$. We define $c_0 = -\infty$ and $c_K = \infty$. Figure A.1 illustrates how this new partition is constructed.

If the moments of the problem lie strictly in the interior of the feasible moment cone, then the problem is equivalent to

$$\begin{array}{ll}\inf\limits_{t,r_i,y_i} & t + \sum\limits_{i=1}^{p} r_i \mu_i + \sum\limits_{i=1}^{p} y_i(\mu_i^2 + \sigma_i^2) \\ \text{s.t.} & t + r_i x + y_i x^2 \geq \min\limits_{k=1,\ldots,K} \{a_k x + b_k\}, \quad \forall x \in A_i,\ \forall i = 1, \ldots, p,\end{array}$$

or, equivalently,

$$\begin{array}{ll}\inf\limits_{t,r_i,y_i} & t + \sum\limits_{i=1}^{p} r_i \mu_i + \sum\limits_{i=1}^{p} y_i(\mu_i^2 + \sigma_i^2) \\ \text{s.t.} & t + r_i x + y_i x^2 \geq a_k x + b_k, \quad \forall x \in \hat{A}_{ik},\ \forall i = 1, \ldots, p,\ \forall k = 1, \ldots, K.\end{array} \tag{A.2}$$

Note that if the interval $\hat{A}_{ik}$ is empty, then the corresponding constraint is ignored. We can therefore apply the same method of casting the constraints (A.2) into SOCP constraints as before. Suppose $\underline{x}_{ik}$ and $\overline{x}_{ik}$ are the lower and upper bounds of interval $\hat{A}_{ik}$ (if they exist). Let $\hat{I}_0$ consist of all (i,k) for which $\hat{A}_{ik}$ is bounded, and $\hat{I}_\infty$ ($\hat{I}_{-\infty}$) consist of all (i,k) for which $\hat{A}_{ik}$ is unbounded above (below). The maximization problem GP_M can be solved through the following SOCP:

$$\begin{array}{ll}\inf\limits_{t,r_i,y_i,\tau_i^k} & t + \sum\limits_{i=1}^{p} r_i \mu_i + \sum\limits_{i=1}^{p} y_i(\mu_i^2 + \sigma_i^2) \\ \text{s.t.} & -b_k + t + \underline{x}_{ik}\overline{x}_{ik}\tau_i^k + y_i + \tau_i^k \geq \sqrt{\left(-b_k + t + \underline{x}_{ik}\overline{x}_{ik}\tau_i^k - y_i - \tau_i^k\right)^2 + \left(-a_k + r_i - (\underline{x}_{ik} + \overline{x}_{ik})\tau_i^k\right)^2}, \\ & \hfill \forall (i,k) \in \hat{I}_0, \\ & -b_k + t + \underline{x}_{ik}\tau_i^k + y_i \geq \sqrt{\left(-b_k + t + \underline{x}_{ik}\tau_i^k - y_i\right)^2 + \left(-a_k + r_i - \tau_i^k\right)^2}, \qquad \forall (i,k) \in \hat{I}_\infty, \\ & -b_k + t - \underline{x}_{ik}\tau_i^k + y_i \geq \sqrt{\left(-b_k + t - \underline{x}_{ik}\tau_i^k - y_i\right)^2 + \left(-a_k + r_i + \tau_i^k\right)^2}, \qquad \forall (i,k) \in \hat{I}_{-\infty}, \\ & \tau_i^k \geq 0, \quad \forall (i,k) \in \hat{I}_0 \cup \hat{I}_\infty \cup \hat{I}_{-\infty}.\end{array}$$

Appendix B

B.1 Proof of Theorem 2.1

To derive the closed-form solution for the worst-case expected newsvendor profit, we consider various forms of dual feasible solutions. For each form, we will construct a primal feasible distribution that satisfies the dual objective value. For ease of derivation, define $\tilde{d}_1 = (\tilde{d} - \mu)^+$ and $\tilde{d}_2 = (\mu - \tilde{d})^+$. Let $\sigma^2 = m_1 + m_2$ and $s\sigma^2 = m_1 - m_2$. The lower bound in Proposition 2.1 is now equivalent to the condition that $\frac{m_2}{m_1}(m_1 + m_2) \leq \mu^2$. We can write the dual problem as

$$\begin{aligned} MVS_D(q) \quad = \quad \sup_{t,r,y_1,y_2} \quad & t + y_1 m_1 + y_2 m_2 \\ \text{s.t.} \quad & t - rx + y_2 x^2 \leq p(x+\mu) - cq, \quad \forall 0 \leq x \leq \mu, \\ & t - rx + y_2 x^2 \leq (p-c)q, \quad \forall 0 \leq x \leq \mu, \\ & t + rx + y_1 x^2 \leq p(x+\mu) - cq, \quad \forall x \geq 0, \\ & t + rx + y_1 x^2 \leq (p-c)q, \quad \forall x \geq 0. \end{aligned}$$

Note that primal feasibility requires the following conditions on the moments: $\mathrm{E}(\tilde{d}_1) = \mathrm{E}(\tilde{d}_2)$, $\mathrm{E}(\tilde{d}_1^2) = m_1$, $\mathrm{E}(\tilde{d}_2^2) = m_2$; and the support conditions: $\tilde{d}_1 \geq 0$, $\tilde{d}_2 \in [0, \mu]$, and $\tilde{d}_1 \tilde{d}_2 = 0$. Observe that zero is always one of the support points for $\tilde{d}_1$ and $\tilde{d}_2$.

Case 1: $q \leq \mu$

Under this case, it is not difficult to verify that $(p-c)q \leq p(x+\mu) - cq$ for all $x \geq 0$. Therefore,

$$\begin{aligned} MVS_D(q) = \sup_{t,r,y_1,y_2} \quad & t + y_1 m_1 + y_2 m_2 \\ \text{s.t.} \quad & t + rx + y_1 x^2 \leq (p-c)q, \quad \forall x \geq 0 \\ & t - rx + y_2 x^2 \leq (p-c)q \quad \forall 0 \leq x \leq \mu, \\ & t - rx + y_2 x^2 \leq -p(x-\mu) - cq, \quad \forall 0 \leq x \leq \mu. \end{aligned}$$

Let us define the piecewise functions $f_1(x) = (p-c)q$ and $f_2(x) = \min\{(p-c)q, -p(x-\mu) - cq\}$. Define also the quadratic functions $g_1(x) = t + rx + y_1 x^2$ and $g_2(x) = t - rx + y_2 x^2$. The quadratic functions are related in exactly two ways. First, they share the same y-intercept. Second, at $x = 0$, $g_1(x)$ increases (or decreases) at exactly the same rate of decrease (or increase) of $g_2(x)$. Figure B.1 illustrates a feasible solution to the dual problem.

Case 1(a): Arbitrary $\tilde{d}_1$ distribution, two-point distribution for $\tilde{d}_2$ (with one support point at μ). Suppose $g_1(x) = f_1(x)$. Furthermore, suppose $g_2(x)$ intersects with $f_2(x)$ at the points 0 and μ. Therefore, we have the relationships:

$$t = (p-c)q, \tag{B.1}$$

$$r = 0, \tag{B.2}$$

$$y_1 = 0, \tag{B.3}$$

$$t - r\mu + y_2 \mu^2 = -cq, \tag{B.4}$$

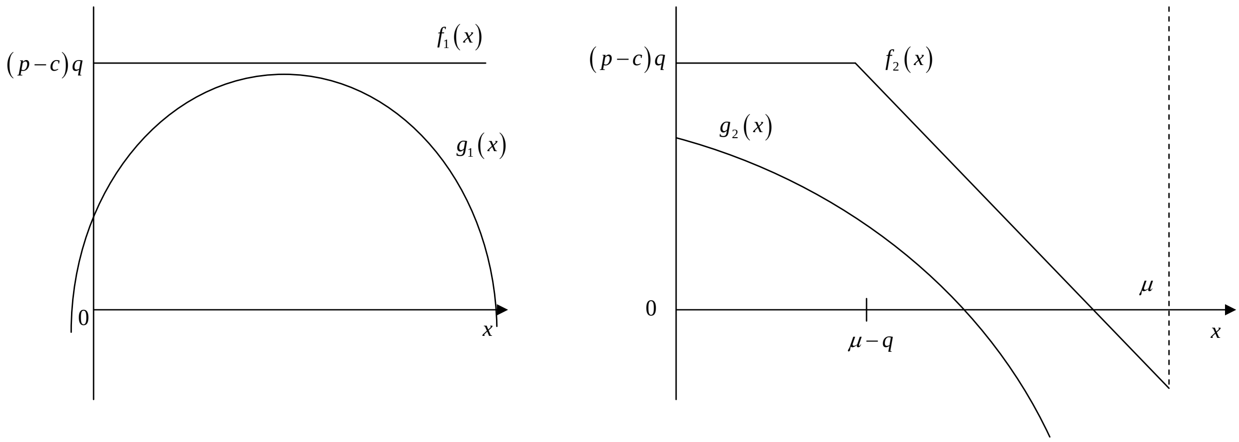

Figure B.1: Graphical illustration of functions satisfying feasibility conditions of the dual problem under the case when $q \leq \mu$.

which define a unique dual solution. The solution is dual feasible if $-r + 2y_2\mu \geq -p$, or equivalently if $q \leq \frac{\mu}{2}$. The dual objective value is

$$Z_{1a} = (p-c)q - \frac{pm_2}{\mu^2}q.$$

We now construct a primal feasible distribution from the dual solution. For a $\tilde{d}_2$ distribution with support points $\{0, \mu\}$ to have second moment m_2, the probability of being nonzero must be exactly $\frac{m_2}{\mu^2}$. By Proposition 2.1, this probability does not exceed 1. From the primal feasibility conditions, it follows that $\Pr(\tilde{d}_1 > 0) = \Pr(\tilde{d}_2 = 0) = 1 - \frac{m_2}{\mu^2}$, $\mathrm{E}(\tilde{d}_1) = \mathrm{E}(\tilde{d}_2) = \frac{m_2}{\mu}$ and $\mathrm{E}(\tilde{d}_1^2) = m_1$. Therefore, we have the conditional moments for $\tilde{d}_1$:

$$\begin{aligned} \mathrm{E}[\tilde{d}_1|\tilde{d}_1 > 0] &= \mathrm{E}[\tilde{d} - \mu|\tilde{d} > \mu] = \tfrac{\mu m_2}{\mu^2 - m_2}, \\ \mathrm{E}[\tilde{d}_1^2|\tilde{d}_1 > 0] &= \mathrm{E}[(\tilde{d} - \mu)^2|\tilde{d} > \mu] = \tfrac{\mu^2 m_1}{\mu^2 - m_2}. \end{aligned}$$

Observe that

$$\mathrm{E}[\tilde{d}_1^2|\tilde{d}_1 > 0] - \Big(\mathrm{E}[\tilde{d}_1|\tilde{d}_1 > 0]\Big)^2 \geq 0$$

is always true by Proposition 2.1. In fact, we can always find a two-point $\tilde{d}_1$ distribution that satisfies these conditional moments. Thus, we can construct a three-point $\tilde{d}$ distribution satisfying these conditions (with one support at zero and two above μ). We find that this distribution achieves a primal objective equal to Z_{1a}. Therefore, if $0 \leq q \leq \frac{\mu}{2}$, we have

$$MVS(q) = (p-c)q - \frac{pm_2}{\mu^2}q.$$

Case 1(b): Arbitrary $\tilde{d}_1$ distribution, two-point distribution for $\tilde{d}_2$. Suppose $g_1(x) = f_1(x)$. Due to the relationship of the quadratic functions, this implies that $g_2(x)$ intersects the line $(p-c)q$

exactly once at $x = 0$. Further suppose that $g_2(x)$ intersects $-p(x-\mu) - cq$ exactly once. Therefore, we have that

$$t = (p-c)q, \tag{B.5}$$

$$r = 0, \tag{B.6}$$

$$y_1 = 0, \tag{B.7}$$

$$(r-p)^2 - 4y_2(t - p\mu + cq) = 0, \tag{B.8}$$

which define a unique dual feasible solution achieving the dual objective

$$Z_{1b} = (p-c)q - \frac{pm_2}{4(\mu - q)}.$$

Let us construct a primal feasible distribution from the dual solution. Note that $g_2(x)$ intersects $-p(x-\mu) - cq$ at $2(\mu - q)$. For a $\tilde{d}_2$ distribution with support points $\{0, 2(\mu - q)\}$ to have second moment m_2, the probability of being nonzero must be exactly $\frac{m_2}{4(\mu-q)^2}$. For now, assume that $\frac{m_2}{4(\mu-q)^2} \le 1$. From the primal feasibility conditions, it follows that $\Pr(\tilde{d}_1 > 0) = \Pr(\tilde{d}_2 = 0) = 1 - \frac{m_2}{4(\mu-q)^2}$, $\mathrm{E}(\tilde{d}_1) = \mathrm{E}(\tilde{d}_2) = \frac{m_2}{2(\mu-q)}$ and $\mathrm{E}(\tilde{d}_1^2) = m_1$. Therefore, we have the conditional moments for $\tilde{d}_1$:

$$\begin{aligned} \mathrm{E}[\tilde{d}_1|\tilde{d}_1 > 0] &= \mathrm{E}[\tilde{d} - \mu|\tilde{d} > \mu] = \tfrac{2m_2(\mu-q)}{4(\mu-q)^2 - m_2}, \\ \mathrm{E}[\tilde{d}_1^2|\tilde{d}_1 > 0] &= \mathrm{E}[(\tilde{d} - \mu)^2|\tilde{d} > \mu] = \tfrac{4m_1(\mu-q)^2}{4(\mu-q)^2 - m_2}. \end{aligned}$$

These conditional moments are valid if and only if

$$\mathrm{E}[\tilde{d}_1^2|\tilde{d}_1 > 0] - \left(\mathrm{E}[\tilde{d}_1|\tilde{d}_1 > 0]\right)^2 \ge 0,$$

or equivalently, if $q \le \mu - \frac{m_2}{2}\sqrt{\frac{m_1+m_2}{m_1 m_2}}$. Note that this range automatically implies that $\frac{m_2}{4(\mu-q)^2} \le 1$. Under the range of q, we can always find a two-point $\tilde{d}_1$ distribution having these conditional moments. Moreover, the support points of $\tilde{d}_2$ are feasible only if $2(\mu - q) \le \mu$. Therefore, in this range, we can construct a primal feasible three-point $\tilde{d}$ distribution (with one support at $2q - \mu$ and two above μ). We find that the constructed distribution achieves a primal objective equal to Z_{1b}. Therefore, if

$$\frac{\mu}{2} \le q \le \mu - \frac{m_2}{2}\sqrt{\frac{m_1+m_2}{m_1 m_2}},$$

we have

$$MVS(q) = (p-c)q - \frac{pm_2}{4(\mu - q)}.$$

Case 1(c): Two-point distribution for $\tilde{d}_1$, two-point distribution for $\tilde{d}_2$. Suppose $\tilde{d}_1 + \tilde{d}_2 > 0$ and $\tilde{d}_1\tilde{d}_2 = 0$. Under this case, there is a unique primal feasible two-point support distribution for $\tilde{d}_1$ and $\tilde{d}_2$. Define

$$\begin{aligned} x_1 &= m_1\sqrt{\tfrac{m_1+m_2}{m_1 m_2}}, \\ x_2 &= m_2\sqrt{\tfrac{m_1+m_2}{m_1 m_2}}. \end{aligned}$$

Then x_1 and x_2 are the nonzero support points of $\tilde{d}_1$ and $\tilde{d}_2$, respectively. Suppose $g_1(x)$ intersects $f_1(x)$ exactly once at x_1. Further suppose that $g_2(x)$ intersects the line $-p(x-\mu)-cq$ exactly once at x_2. Then the following relationships must hold:

$$r^2 - 4y_1(t-(p-c)q) = 0, \tag{B.9}$$

$$\frac{-r}{2y_1} = x_1, \tag{B.10}$$

$$(r-p)^2 - 4y_2(t-p\mu+cq) = 0, \tag{B.11}$$

$$\frac{r-p}{2y_2} = x_2, \tag{B.12}$$

$$r \geq 0, \tag{B.13}$$

where the last inequality is necessary in order for the dual variables to be feasible. There is a unique solution to the equality constraints attained by the following variables

$$\begin{array}{rcl} t & = & p\left(\frac{qm_2+\mu m_1}{m_1+m_2} - \frac{1}{2}\sqrt{\frac{m_1m_2}{m_1+m_2}}\right) - cq, \\ r & = & \frac{2p}{m_1}\sqrt{\frac{m_1m_2}{m_1+m_2}}\left(\frac{m_1(q-\mu)}{m_1+m_2} + \frac{1}{2}\sqrt{\frac{m_1m_2}{m_1+m_2}}\right), \\ y_1 & = & \frac{-pm_2}{m_1(m_1+m_2)}\left(\frac{m_1(q-\mu)}{m_1+m_2} + \frac{1}{2}\sqrt{\frac{m_1m_2}{m_1+m_2}}\right), \\ y_2 & = & \frac{pm_1}{m_2(m+1+m_2)}\left(\frac{m_2(q-\mu)}{m_1+m_2} - \frac{1}{2}\sqrt{\frac{m_1m_2}{m_1+m_2}}\right), \end{array}$$

which achieves the dual objective value

$$Z_{1c} = p\left(\frac{qm_2+\mu m_1}{m_1+m_2} - \sqrt{\frac{m_1m_2}{m_1+m_2}}\right) - cq.$$

The solution is feasible if $r \geq 0$, or equivalently, $q \geq \mu - \frac{m_2}{2}\sqrt{\frac{m_1+m_2}{m_1m_2}}$. Now let us construct a primal feasible solution from the intersection points of the dual solution. Consider the two-point distribution

$$\tilde{d} = \begin{cases} \mu + m_1\sqrt{\frac{m_1+m_2}{m_1m_2}}, & \text{w.p. } \frac{m_2}{m_1+m_2}, \\ \mu - m_2\sqrt{\frac{m_1+m_2}{m_1m_2}}, & \text{w.p. } \frac{m_1}{m_1+m_2}, \end{cases}$$

We can easily verify that the constructed distribution satisfies the known moments. Also, because of Proposition 2.1, it follows that the support points are nonnegative. We find that the distribution achieves an expected payoff equal to Z_{1c}. Therefore, for

$$\mu - \frac{m_2}{2}\sqrt{\frac{m_1+m_2}{m_1m_2}} \leq q \leq \mu,$$

we have

$$MVS(q) = p\left(\frac{qm_2+\mu m_1}{m_1+m_2} - \sqrt{\frac{m_1m_2}{m_1+m_2}}\right) - cq.$$

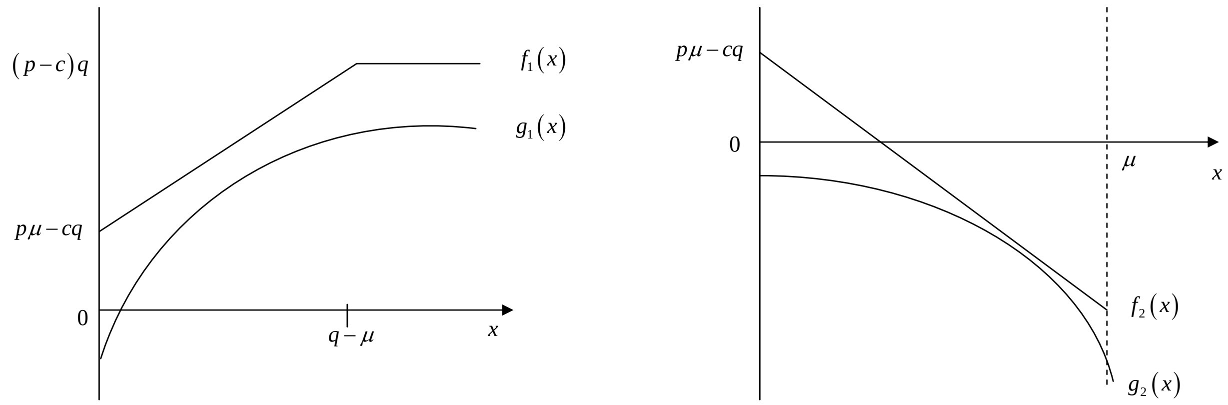

Figure B.2: Graphical illustration of functions satisfying feasibility conditions of the dual problem under the case when $q \geq \mu$.

Case 2: $q \geq \mu$
Under this case, it is not difficult to verify that $-p(x-\mu) - cq \leq (p-c)q$ for all $x \geq 0$. Therefore,

$$\begin{aligned} MVS_D(q) = \sup_{t,r,y_1,y_2} \quad & t + y_1 m_1 + y_2 m_2 \\ \text{s.t.} \quad & t + rx + y_1 x^2 \leq p(x+\mu) - cq, && \forall x \geq 0, \\ & t + rx + y_1 x^2 \leq (p-c)q && \forall x \geq 0, \\ & t - rx + y_2 x^2 \leq -p(x-\mu) - cq, && \forall 0 \leq x \leq \mu. \end{aligned}$$

Define the piecewise functions $f_1(x) = \min\{p(x+\mu) - cq, (p-c)q\}$ and $f_2(x) = -p(x-\mu) - cq$. Define also the quadratic functions $g_1(x) = t + rx + y_1 x^2$ and $g_2(x) = t - rx + y_2 x^2$. Like before, the quadratic functions share the same y-intercept. And at $x = 0$, $g_1(x)$ increases (or decreases) at exactly the same rate of decrease (or increase) of $g_2(x)$. Figure B.2 illustrates a feasible solution to the dual problem.

Case 2(a): Two-point distribution for $\tilde{d}_1$, two-point distribution for $\tilde{d}_2$. Similar to our discussion in Case 1(c), we know that there is a unique two-point distribution for each $\tilde{d}_1$ and $\tilde{d}_2$ that satisfies the given moments if $\tilde{d}_1 + \tilde{d}_2 > 0$ and $\tilde{d}_1\tilde{d}_2 = 0$. These nonzero support points for $\tilde{d}_1$ and $\tilde{d}_2$ are x_1 and x_2 (defined in Case 1(c)), respectively. Suppose $g_1(x)$ intersects $f_1(x)$ exactly once at x_1. Also, suppose $g_2(x)$ intersects $f_2(x)$ exactly once at x_2. In fact, this case is similar to Case 1(c) since relationships (B.9)–(B.12) must hold. Instead of constraint (B.13), we must require $r \leq p$ for feasibility. The solution of the dual problem is the same as the previous case, with the dual objective value

$$Z_{2a} = p\left(\frac{qm_2 + \mu m_1}{m_1 + m_2} - \sqrt{\frac{m_1 m_2}{m_1 + m_2}}\right) - cq.$$

Note that the solution is dual feasible if $r \leq p$, or equivalently, $q \leq \mu + \frac{m_1}{2}\sqrt{\frac{m_1+m_2}{m_1 m_2}}$. We have shown earlier, there is a two-point $\tilde{d}$ distribution that is feasible and achieves the dual objective Z_{2a}. Therefore, for

$$\mu \leq q \leq \mu + \frac{m_1}{2}\sqrt{\frac{m_1 + m_2}{m_1 m_2}},$$

we have

$$MVS(q) = p\left(\frac{qm_2 + \mu m_1}{m_1 + m_2} - \sqrt{\frac{m_1 m_2}{m_1 + m_2}}\right) - cq.$$

Case 2(b): Two-point distribution for $\tilde{d}_1$, arbitrary $\tilde{d}_2$ distribution. Suppose $g_2(x)$ is exactly the line $-p(x-\mu) - cq$. Due to the relationship between the two quadratic functions, this also implies that $g_1(x)$ intersects the line $p(x+\mu) - cq$ exactly once at $x = 0$. Further suppose that $g_2(x)$ intersects $(p-c)q$ exactly once. In other words, the following relationships must hold:

$$r = p, \tag{B.14}$$

$$t = p\mu - cq, \tag{B.15}$$

$$y_2 = 0, \tag{B.16}$$

$$r^2 - 4y_1(t - (p-c)q) = 0. \tag{B.17}$$

This gives a unique solution for the dual variables. The dual objective value is

$$Z_{2b} = p\mu - cq - \frac{pm_1}{4(q-\mu)}.$$

Now let us construct a primal feasible distribution from the dual solution. Note that $g_1(x)$ intersects $(p-c)q$ at exactly $2(q-\mu)$. For a $\tilde{d}_1$ distribution with support points $\{0, 2(q-\mu)\}$ to have a second moment m_1, the probability of being nonzero must be exactly $\frac{m_1}{4(q-\mu)^2}$. For now, suppose that $\frac{m_1}{4(q-\mu)^2} \leq 1$. Primal feasibility implies that $\Pr(\tilde{d}_2 > 0) = \Pr(\tilde{d}_1 = 0) = 1 - \frac{m_1}{4(q-\mu)^2}$, $\mathrm{E}(\tilde{d}_2) = \mathrm{E}(\tilde{d}_1) = \frac{m_1}{2(q-\mu)}$ and $\mathrm{E}(\tilde{d}_2^2) = m_2$. Therefore, the conditional moments of $\tilde{d}_2$ must be:

$$\begin{aligned}\mathrm{E}[\tilde{d}_2|\tilde{d}_2 > 0] = \mathrm{E}[\mu - \tilde{d}|\tilde{d} < \mu] &= \tfrac{2m_1(q-\mu)}{4(q-\mu)^2 - m_1}, \\ \mathrm{E}[\tilde{d}_2^2|\tilde{d}_2 > 0] = \mathrm{E}[(\mu - \tilde{d})^2|\tilde{d} < \mu] &= \tfrac{4m_2(q-\mu)^2}{4(q-\mu)^2 - m_1}.\end{aligned}$$

These conditional moments are valid if and only if:

$$\begin{aligned}\mathrm{E}[\tilde{d}_2(\mu - \tilde{d}_2)|\tilde{d}_2 > 0] &\geq 0, \\ \mathrm{E}[\tilde{d}_2^2|\tilde{d}_2 > 0] - \left(\mathrm{E}[\tilde{d}_2|\tilde{d}_2 > 0]\right)^2 &\geq 0.\end{aligned}$$

This is equivalent to the condition that

$$\mu + \frac{m_1}{2}\sqrt{\frac{m_1 + m_2}{m_1 m_2}} \leq q \leq \mu + \frac{\mu m_1}{2m_2}.$$

Moreover, this range of q automatically implies that $\frac{m_1}{4(q-\mu)^2} \leq 1$. In fact, under this range of q, we can always find a two-point $\tilde{d}_2$ distribution with these conditional moments. Thus, we can construct a primal feasible $\tilde{d}$ distribution with three support points (one at $2q-\mu$ and two below μ). This distribution achieves a primal objective value equal to Z_{2b}. Thus, for this range of q,

$$MVS(q) = p\mu - cq - \frac{pm_1}{4(q-\mu)}.$$

Case 2(c): Three-point distribution for $\tilde{d}_1$, two-point distribution for $\tilde{d}_2$ (with one support point at μ). Suppose $g_1(x)$ intersects $f_1(x)$ exactly twice at strictly positive values. Further suppose that $g_2(x)$ is convex and intersects $f_2(x)$ at μ. Therefore the following relationships must hold:

$$(r-p)^2 - 4y_1(t - p\mu + cq) = 0, \tag{B.18}$$

$$r^2 - 4y_1(t-(p-c)q) = 0, \tag{B.19}$$

$$t - r\mu + y_2\mu^2 = -cq, \tag{B.20}$$

$$r \geq p. \tag{B.21}$$

The last inequality constraint is needed in order for the quadratic functions to remain below piecewise functions. By eliminating the variable y_1 in the first two equations, we can get a second-order equation in r:

$$(q-\mu)r^2 + 2(t-(p-c)q)r - p(t-(p-c)q) = 0.$$

One of the two roots of the equation violates the condition that r does not exceed p. Therefore, we must choose

$$r = \frac{-(t-(p-c)q) + \sqrt{(t-(p-c)q)(t-p\mu+cq)}}{(q-\mu)}.$$

We can also write the remaining variables as functions of t:

$$\begin{aligned} y_1 &= \tfrac{(t-(p-c)q)+(t-p\mu+cq)-2\sqrt{(t-(p-c)q)(t-p\mu+cq)}}{4(q-\mu)^2}, \\ y_2 &= -\tfrac{(t-p\mu+cq)}{\mu^2} + \tfrac{-(t-p\mu+cq)+\sqrt{(t-(p-c)q)(t-p\mu+cq)}}{\mu(q-\mu)}. \end{aligned}$$

Thus, the dual objective as a function of t is:

$$\begin{aligned} t + \tfrac{m_1}{4(q-\mu)^2}(t-(p-c)q) + \left(\tfrac{m_1}{4(q-\mu)^2} - \tfrac{m_2}{\mu^2} - \tfrac{m_2}{\mu(q-\mu)}\right)(t-p\mu+cq) \\ + \left(\tfrac{m_2}{\mu(q-\mu)} - \tfrac{m_1}{2(q-\mu)^2}\right)\sqrt{(t-(p-c)q)(t-p\mu+cq)}. \end{aligned}$$

Let us define the variables

$$\begin{aligned} a &= \tfrac{m_1}{2(q-\mu)^2} - \tfrac{m_2}{\mu(q-\mu)}, \\ b &= 1 - \tfrac{m_2}{\mu^2}. \end{aligned}$$

The derivative of the objective function is then

$$a + b - \frac{a}{2}\frac{(t-(p-c)q) + (t-p\mu+cq)}{\sqrt{(t-(p-c)q)(t-p\mu+cq)}}.$$

Assume that a is negative, or equivalently, $q \geq \mu + \frac{m_1 \mu}{2m_2}$. It is not difficult to verify that due to the lower bound in Proposition 2.1, $2a + b \geq 0$ for all values of q. Therefore, $a + b \geq 0$. Equating the derivative to zero and solving for t, we find that

$$t = \frac{p}{2}\left(q + \mu - \frac{(q-\mu)(a+b)}{\sqrt{b(2a+b)}}\right) - cq.$$

We can easily check that this value does not exceed $p\mu - cq$. Thus, we have the following dual feasible variables:

$$\begin{aligned} r &= \tfrac{p}{2}\left(1 + \sqrt{\tfrac{b}{2a+b}}\right), \\ y_1 &= -\tfrac{p}{4(q-\mu)}\sqrt{\tfrac{b}{2a+b}}, \\ y_2 &= \tfrac{p(q-\mu)}{2\mu^2}\left(\tfrac{(a+b)}{\sqrt{b(2a+b)}} - 1\right) + \tfrac{p}{2\mu}\left(\sqrt{\tfrac{b}{2a+b}} - 1\right), \end{aligned}$$

and the dual objective simplifies to

$$Z_{2c} = \frac{p}{2}\left(\mu + bq - (q-\mu)\sqrt{b(2a+b)}\right) - cq.$$

Now let us construct a primal feasible distribution from the intersection points of the dual solution. Consider the distribution:

$$\tilde{d} = \begin{cases} \mu + (q-\mu)\left(1 - \sqrt{\frac{2a+b}{b}}\right), & \text{w.p. } \frac{1}{2}\left(b + \left(b - \frac{(1-b)\mu}{(q-\mu)}\right)\sqrt{\frac{b}{2a+b}}\right), \\ \mu + (q-\mu)\left(1 + \sqrt{\frac{2a+b}{b}}\right), & \text{w.p. } \frac{1}{2}\left(b - \left(b - \frac{(1-b)\mu}{(q-\mu)}\right)\sqrt{\frac{b}{2a+b}}\right), \\ 0, & \text{w.p. } 1 - b. \end{cases}$$

All the probabilities take values within the range of 0 to 1 by Proposition 2.1. Moreover, the moments of the primal problem are satisfied. We can verify that expected profit achieved by the constructed distribution is exactly equal to Z_{2c}. Thus, for

$$q \geq \mu + \frac{\mu m_1}{2m_2},$$

we have

$$\begin{aligned} MVS(q) &= \tfrac{p}{2}\left(\mu + bq - (q-\mu)\sqrt{b(2a+b)}\right) - cq \\ &= \tfrac{p}{2}\left(\mu + bq - \sqrt{(bq-\mu)^2 - (1-b)^2\mu^2 + m_1 b}\right) - cq. \end{aligned}$$

Combining all the cases and letting $m_1 = (1+s)\sigma^2/2$ and $m_2 = (1-s)\sigma^2/2$, we get the closed-form expression. □

B.2 Proof of Theorem 2.2

The worst-case profit $MVS(q)$ is concave and continuous in q. Note that unless the profit is maximized by a range of q, the optimal quantity can only occur at 0 or in Regions (ii), (iv) or (v) (as defined in Theorem 2.1). Let q^* be an order quantity that maximizes $MVS(q)$. Define

$$\begin{aligned} q^*_{(ii)} &= \mu - \tfrac{\sigma}{2}\sqrt{\tfrac{(1-s)p}{2(p-c)}}, \\ q^*_{(iv)} &= \mu + \tfrac{\sigma}{2}\sqrt{\tfrac{(1+s)p}{2c}}, \\ q^*_{(v)} &= \tfrac{\mu}{b} + \tfrac{(pb-2c)}{2b}\sqrt{\tfrac{(1+s)\sigma^2 b - 2(1-b)^2\mu^2}{2c(pb-c)}}, \end{aligned}$$

which are the unconstrained maximizers of the aggregate functions in Regions (ii), (iv) and (v), respectively. Suppose $\frac{c}{p} > 1 - \frac{(1-s)\sigma^2}{2\mu^2}$. Then $MVS(q)$ is strictly decreasing in Region (i). This implies that the function is strictly decreasing for $q \geq 0$. Thus, $q^* = 0$. Note that $q^*_{(ii)}$ lies in Region (ii) if $\frac{1}{2}(1-s) \leq \frac{c}{p} \leq 1 - \frac{(1-s)\sigma^2}{2\mu^2}$. Under this case, $q^* = q^*_{(ii)}$. Similarly, $q^*_{(iv)}$ lies in Region (iv) if $\frac{1}{2}\frac{(1-s)^2\sigma^2}{(1+s)\mu^2} \leq \frac{c}{p} \leq \frac{1}{2}(1-s)$. If this is true, then $q^* = q^*_{(iv)}$. Finally, if $\frac{c}{p} < \frac{1}{2}\frac{(1-s)^2\sigma^2}{(1+s)\mu^2}$, then $MVS(q)$ is increasing in Regions (i) to (iv). Therefore, the maximum is attained in Region (v). Thus, $q^* = q^*_{(v)}$. $\square$

References

[1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[2] E. Bartezzaghi, R. Verganti, and G. Zotteri. Measuring the impact of asymmetric demand distributions on inventories. *International Journal of Production Economics*, 60, 1999.

[3] E. Bartezzaghi, R. Verganti, and G. Zotteri. A simulation framework for forecasting uncertain lumpy demand. *International Journal of Production Economics*, 58:499–510, 1999.

[4] A. Ben-Tal and M. Teboulle. Expected utility, penalty functions and duality in stochastic nonlinear programming. *Management Science*, 32:1445–1466, 1986.

[5] A. Ben-Tal and M. Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476, 2007.

[6] P. Berck and J. M. Hihn. Using the semivariance to estimate safety-first rules. *American Journal of Agricultural Economics*, 64(2):298–300, 1982.

[7] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[8] S. A. Bond. Asymmetry and downside risk in foreign exchange markets. University of Cambridge. Working Paper, 2000.

[9] S. A. Bond and S. E. Satchell. Statistical properties of the sample semi-variance. *Applied Mathematical Finance*, 9:219–239, 2002.

[10] G. C. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18(3):853–877, 2007.

[11] R. Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

[12] J. D. Croston. Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 23:289–303, 1972.

[13] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1006, 1982.

[14] L. G. Epstein. A definition of uncertainty aversion. *Review of Economic Studies*, 66:579–608, 1999.

[15] E. F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38(1):34–105, 1965.

[16] R. Fildes and C. Beard. Forecasting systems for production and inventory control. *International Journal of Operations and Production Management*, 12(5):4–27, 1992.

[17] P. C. Fishburn. Mean-risk analysis with risk associated with below-target returns. *The American Economic Review*, 67(2):116–126, 1977.

[18] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.

[19] M. Fritelli and E. R. Gianin. Putting order in risk measures. *Journal of Banking and Finance*, 26:1473–1486, 2002.

[20] G. Gallego and I. Moon. The distribution free newsboy problem: Review and extensions. *Journal of the Operational Research Society*, 44(8):825–834, 1993.

[21] S. He, J. Zhang, and S. Zhang. Bounding probability of small deviations: A fourth moment approach. Working Paper, 2007.

[22] K. Isii. On the sharpness of Chebyshev-type inequalities. *Annals of the Institute of Statistical Mathematics*, 12:185–197, 1963.

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

[24] F. R. Johnston and J. E. Boylan. Forecasting for items with intermittent demand. *Journal of the Operational Research Society*, 47:113–121, 1996.

[25] J. L. Knight, S. E. Satchell, and K. C. Tran. Statistical modelling of asymmetric risk in asset returns. *Applied Mathematical Finance*, 2:155–172, 1995.

[26] A. W. Lo. Semi-parametric upper bounds for option prices and expected payoffs. *Journal of Financial Economics*, 19(2):373–387, 1987.

[27] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and Its Applications*, 284:193–228, 1998.

[28] H. M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. Wiley, New York, 1959.

[29] K. Natarajan, M. Sim, and J. Uichanco. Tractable robust expected utility and risk models for portfolio optimization. Working Paper, 2008.

[30] Y. Nesterov and A. Nemirovski. *Interior-point polynomial methods in convex programming*, volume 13 of *Studies in Applied Mathemathics*. SIAM, Philadelphia, PA, 1994.

[31] G. Perakis and G. Roels. Regret in the newsvendor model with partial information. *Operations Research*, 56(1):188–203, 2008.

[32] L. J. Savage. The theory of statistical decisions. *Journal of American Statistical Association*, 46:55–67, 1951.

[33] H. Scarf. A min-max solution to an inventory problem. In K. J. Arrow, S. Karlin, and H. Scarf, editors, *Studies in the Mathematical Theory of Inventory and Production*, chapter 12, pages 201–209. Stanford University Press, 1958.

[34] A. De Schepper and B. Heijnen. Distribution-free option pricing. *Insurance: Mathematics and Economics*, 40:179–199, 2007.

[35] E. A. Silver. Some ideas related to the inventory control of items having erratic demand patterns. *Canadian Operational Research Journal*, 8:87–100, 1970.

[36] F. A. Sortino and H. J. Forsey. On the use and misuse of downside risk. *Journal of Portfolio Management*, 22(2):35–42, 1996.

[37] J. B. Ward. Determining reorder points when demand is lumpy. *Management Science*, 24:623–632, 1978.

[38] L. F. Zuluaga, J. Peña, and D. Du. Extensions of Lo's semiparametric bound for European call options. Working Paper, 2006.