Data-Driven Distributionally Robust Newsvendor Models with Censored Demand

Abstract

In this paper we generalize classic maximin model (Scarf, 1958) of the distributionally robust newsvendor problem, creating *data-driven* models for *sales* (i.e., censored demand). In particular, we focus on a single-period single-item distributionally robust newsvendor context, where only observations of sales (censored demand) data are available. We first calculate confidence intervals for the first two sales moments and then derive necessary and sufficient conditions to ensure that a valid demand distribution exists for any combination of sales moments contained in the union of the confidence intervals, which lead to guidance on appropriate confidence levels. Once the confidence intervals are well-posed, their union forms the data-driven ambiguity set in our new distributionally robust newsvendor models. Focusing on the case where confidence intervals for only the first two moments are utilized, we derive a closed-form solution for the optimal order quantity. Moreover, we extend our models by incorporating information about the probability of demand censoring as well as allowing the sales data to be censored by multiple inventory levels. For the general case where demands are censored by multiple capacities, we show that the model can be solved via semidefinite programming. Extensive experiments, based on real data, explore the impact of various parameters and complement our theoretical contributions. Our paper provides the first robust newsvendor models that can accommodate censored demand using only data, whose solutions may be implemented easily via either closed-form solutions or tractable semidefinite programs.

Keywords: Robust Optimization, newsvendor, data-driven, censored demand.

1 Introduction

Distributionally-robust newsvendor problems have been widely studied in the operations management literature. For a comprehensive overview of robust models in this domain, we refer the reader to Lu & Shen (2021). Among the most influential approaches is the maximin model, originally introduced by (Scarf, 1958), which maximizes the worst-case (minimum) profit over all valid distributions. In the maximin model, summary statistics of the demand distribution, such as mean and variance, are utilized and can be estimated from historical demand data. However, in many applications the historical data are observations of *sales*, or *censored* demand. Furthermore, the censoring depends critically on the inventory level that had been previously chosen; if actual demand exceeds the available inventory, the excess demand is typically not tracked and hence unobservable. As a result, it is difficult for firms to accurately estimate the underlying demand distribution, or even moments of demand, using only sales data. While the problem of censored demand has been studied in the literature, the approaches are typically based on stochastic optimization, which require some distributional assumptions. In contrast, we develop data-driven distributionally-robust newsvendor models that are based on *sales* information, rather than demand information; in particular, we assume that only sales *data* are available and we use them to calculate confidence intervals for the sales moments, which serve as building blocks for ambiguity sets in our robust models. To the best of our knowledge, we are the first to combine censored demand data with data-driven distributionally-robust optimization.

1.1 Contributions

Our paper provides several contributions for the distributionally robust newsvendor problem:

1. **Introduction of Data-Driven Censored-Demand Models**: Unlike earlier models that assume demand moments are known precisely, we instead assume that only *sales data* are available. We propose the following framework to incorporate sales data:

    a) We first calculate confidence intervals for the first ℓ sales moments, focusing on the case of the first two moments, based on the sales data. Then, we derive necessary and sufficient conditions on the confidence intervals to ensure that a valid demand distribution exists for any combination of sales moments in the union of the confidence intervals, which lead to guidelines for setting confidence levels.

    b) Once it is ensured that the set of confidence intervals is well posed, their union serves as an ambiguity set in our distributionally robust newsvendor models.

2. **Generalizations**: We generalize the maximin model (Scarf, 1958) to accommodate a *data-driven* environment with *censored demand*. In particular, we show that:

    a) Given well-posed confidence intervals for the first two moments, we derive a closed-form expression for the optimal order quantity when the sales data is censored by a single censoring level.

    b) We extend our model by incorporating confidence intervals for the probability of demand censoring. We retain our ability to derive a closed-form solution under this extension. Overall, our single-censoring-level model is a data-driven DRO model for truncated demand/sales; and our novelty is not a new Scarf structure, but the construction of statistically valid ambiguity sets from censored sales data.

    c) We further extend our model by allowing the sales data to be censored by multiple historical censoring levels, which are efficiently solvable by semidefinite programming (SDP).

3. **Technical Contributions**: While our analyses are guided by existing results from the literature (e.g., the duality theory from Shapiro (2001) and the data-driven distributionally robust optimization of Delage & Ye (2010)), the introduction of censored demand creates technical challenges. Notably, in the case where there are M censoring levels, we find that the model requires solving for the maximum of M distinct SDP problems.

4. **Benchmarking**: We compare our models, using real data from Kaggle, with three benchmarks, namely 1) a simple de-censoring strategy, and 2) the Kaplan-Meier (KM) myopic model, a non-parametric estimate of the classic newsvendor order quantity for censored demand (Huh et al., 2011).We demonstrate that our models can outperform both benchmarks.

We next provide a literature review to survey the most relevant research to our paper and to properly position our contributions in terms of existing work.

1.1.1 The Newsvendor with Censored Demand.

We begin by discussing the literature on the newsvendor with censored demand. The first stream consists of parametric demand estimation techniques using censored demand. Under this approach, the underlying demand distribution is assumed to belong to a parametric family of distributions with unknown parameters; censored demand information is then used to estimate the underlying demand distribution's parameters. For example, Ding et al. (2002) and Lu et al. (2008) apply a Bayesian approach to the newsvendor model and its multi-period extension with censored demand. In both papers, the unknown parameters are first updated through prior and posterior observations, and then the newsvendor model is solved based on the updated demand distribution. The second stream of literature focuses on nonparametric approaches to analyze the same problem. For instance, Burnetas & Smith (2000) propose an adaptive algorithm to find approximations for the newsvendor's optimal order quantity and price. Godfrey & Powell (2001) develop algorithms for solving the newsvendor problem with censored demand. Huh & Rusmevichientong (2009) introduce a stochastic gradient descent approach to improve convergence rates of demand estimation. Huh et al. (2011) apply a Kaplan-Meier approach for a distribution-free newsvendor model with censored demand. In contrast, our paper adopts distributionally robust optimization in a data-driven setting to study the newsvendor problem under censored demand, which leads to tractable solution methods. In our most general setting, we formulate the resulting maxmin DRO problem and show that its dual can be reformulated as a finite-dimensional semidefinite program, yielding a tractable solution approach. For the simpler case with second-order sales information (mean and variance) under a single censoring level, we further derive closed-form characterizations.

1.1.2 Maximin Distributionally Robust Newsvendor.

We next discuss a stream of literature concerning newsvendor models that are studied via distributionally-robust optimization, where moment information of demand is available. As mentioned in the introduction, the approach is a maximin methodology that derives the optimal order quantity under the worst-case distribution of demand. Scarf (1958) and Gallego & Moon (1993) are two notable papers in this stream and both derive closed-form expressions for optimal quantities when only the mean and variance of demand are known. Yue et al. (2006) analyze the expected value of distributional information for the classic maximin order quantity from Scarf (1958). Ben-Tal & Hochman (1976) study a similar model but where the mean absolute deviation of demand is available. Ben-Tal et al. (2013) extend the single-item newsvendor model to a case with multiple items where the demand information for each item are available. Recently, Natarajan et al. (2018) introduce demand semivariance information into the model to measure distributional asymmetry and derive a closed-form solution for the order quantity. For a more general model, Zymler et al. (2013) develop a distributionally robust model with joint chance constraints given first- and second-order moment information, as well as the support of uncertain parameters, and provide an SDP to approximate the worst-case conditional value-at-risk. Our study focuses on a single-item distributionally robust newsvendor problem; however, in contrast, we assume that only sales data, rather than demand moments, are available. In particular, given an inventory level, sales are censored demand. We create confidence intervals for the first and second moments of sales, and we derive a closed-form expression for the optimal order quantity. Furthermore, for the case with multiple censoring levels, we are able to solve for the optimal ordering quantity using semidefinite programming.

2 Background and Baseline Model

The classic (stochastic) newsvendor model is concerned with a firm that needs to decide on an order quantity q at the beginning of a sales season where demand is uncertain; the demand is realized by the end of the sales season once the inventory is on hand. The unit revenue is r and the unit cost is c; we assume $0 < c < r$. We also assume, without loss of generality, that unsold units at the end of the sales season have zero salvage value.

The optimization problems in this paper are defined over a probability space $(\Omega, \mathcal{D}, F)$, where $\Omega = \mathbb{R}_+$ is the support of demand, $\mathcal{D}$ is the Borel σ-algebra on Ω, and F is a Borel probability measure of demand on $\mathbb{R}_+$. Given a continuous probability distribution F of demand d, the firm chooses an order quantity q to maximize expected profit: $\max_{q \geq 0} \pi_F$, where $\pi_F \triangleq \mathbb{E}_F \left[r \min(d, q) - cq \right]$, with the well-known optimal solution $q^* = F^{-1}(1 - c/r)$.

However, knowing the precise distribution F is a very strong assumption, as detailed in the distributionally robust newsvendor literature. If there is ambiguity about the underlying distribution, one can maximize the expected profit under a worst-case demand distribution, using various approaches. In this section, we review the maximin distributionally robust newsvendor model analyzed by Scarf (1958). In Section 3, we generalize the model by incorporating censored demand and data-driven ambiguity sets of moment information into their formulations.

Assuming that only the mean μ and variance σ^2 of demand are known, the maximin distributionally robust newsvendor model is formulated as follows:

$$
\begin{aligned}
\max_{q \geq 0} \quad \min_{F \in \mathcal{D}} \quad & \mathbb{E}_F \left[r \min(x, q) - cq \right] \\
\text{s.t.} \quad & \int_\Omega dF(x) = 1, \\
& \int_\Omega x dF(x) = \mu, \\
& \int_\Omega x^2 dF(x) = \mu^2 + \sigma^2.
\end{aligned}
$$

Scarf (1958) showed that the optimal order quantity for this formulation is

$$q_B^* = \begin{cases} \mu + \frac{\sigma}{2}\left(\sqrt{\frac{r-c}{c}} - \sqrt{\frac{c}{r-c}}\right), & \text{if } \frac{\mu}{\sigma} \geq \sqrt{\frac{c}{r-c}} \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where the subscript B indicates the maximin *baseline* solution. Note that we obtain a trivial outcome $q_B^* = 0$ when the ratio μ/σ is small.

3 A Data-Driven Maximin Model with Censored Demand

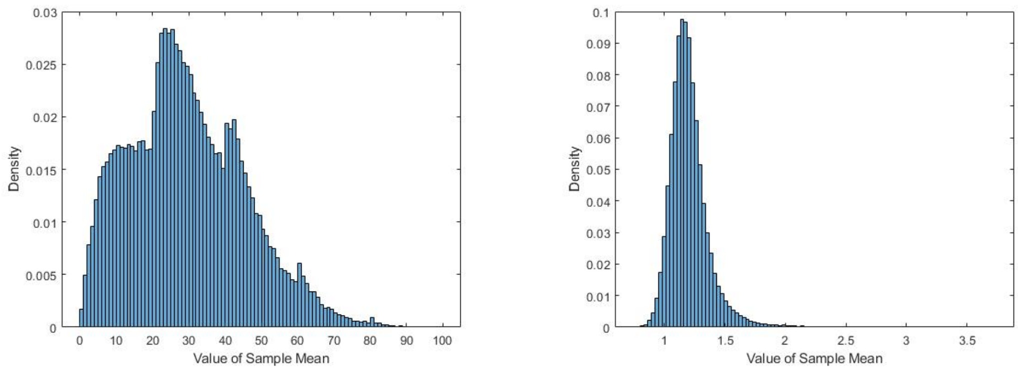
In this section, we propose and analyze a data-driven maximin distributionally robust newsvendor model that only has access to sales data, which we use to estimate the statistical properties of sales. We create statistical confidence intervals for moments and probabilities, which are then fed into a new distributionally robust newsvendor model; we use the percentile Bootstrap method to calculate these confidence intervals, though other approaches can be used to determine the intervals. In other words, we propose and analyze data-driven distributionally robust newsvendor models with censored demand. Our models generalize the baseline model of Section 2.

We begin by analyzing a base case in Section 3.1, where only the first two sales moments are considered, and we derive a closed-form optimal order quantity. In Section 3.2, we extend the base case model by considering the probability that demand is censored, and derive a closed-form optimal order quantity. Finally, in Section 4, we extend our model to allow the sales data to be censored by multiple inventory levels, which can be solved via semidefinite programming.

Throughout this paper, let d denote the latent demand and $s = \min(d, k)$ denote the observed sales, where k represents the historical censoring threshold that generated the available data. In Sections 3.1–3.2, we analyze the single-threshold case where the entire historical dataset was generated under a common k. This k may represent a physical display cap, but it often represents a fixed historical order-up-to level Q_{up-to} that remained constant over the sampling period. Under this interpretation, the manager only observes sales censored by this censoring threshold. The current decision variable is the order quantity q. Because our historical observations are right-censored at k, the model optimizes the order quantity within this known data horizon, requiring $q \leq k$. In Section 4, we extend this to multiple historical censoring thresholds $\{k_i\}$ (when historical censoring thresholds vary). Finally, Section 4.2 provides an illustrative rolling implementation and discusses its inherent informational limitations.

3.1 A Model with Mean & Variance

Suppose we have N observations of sales $\{s_j\}_{j=1}^{N}$, where $s_j = \min(d_j, k)$. Here, d_j is the uncensored demand realization and k is the common historical censoring threshold. This setting is particularly relevant when a firm maintains a consistent order-up-to inventory level over several periods, thereby generating a stable censoring threshold, before seeking a distributionally robust adjustment to its order quantity. Given these N sales observations, we first construct confidence intervals for the first two moments of sales and then derive a closed-form solution for the optimal order quantity q (subject to $q \leq k$), expressed as a function of these data-driven intervals.



(a) $N = 5$, $x \sim exp(5)$, $k = 10$, $i = 2$.   (b) $N = 100$, $x \sim N(1, 1)$, $k = 1.2$, $i = 4$.

Figure 1: Possible distributions of $\hat{\mu}_i$.

Given the N observations of sales $\{s_j\}_{j=1,\ldots,N}$, an $\alpha\%$ confidence interval for the i-th moment μ_i can be obtained by using the student's t-distribution: $\mu_i \in \left[\hat{\mu}_i - t_{\alpha,N-1}\frac{\hat{\sigma}_i}{\sqrt{N}}, \hat{\mu}_i + t_{\alpha,N-1}\frac{\hat{\sigma}_i}{\sqrt{N}}\right]$, where $\hat{\mu}_i = \frac{1}{N}\sum_{j=1}^{N} s_j^i$ is the sample mean, $\hat{\sigma}_i^2 = \frac{1}{N-1}\sum_{j=1}^{N}\left(s_j^i - \hat{\mu}_i\right)^2$ is the associated sample variance, and $t_{\alpha,N-1}$ is the associated critical value of the student's t-distribution with $N-1$ degrees of freedom[1]. However, applying this approach is not appropriate when the sample size N is small. For example, given $(N, k) = (5, 10)$ and supposing that d_j follows an exponential distribution with mean 5, the sample distribution of the sales mean $\hat{\mu}_1 = \frac{1}{N}\sum_{j=1}^{N} s_j$ could be highly skewed, in contrast to the symmetrical student's t-distribution – see Figure 1a. Moreover, when the order index i is large, the distribution of s_j^i can become unreasonably skewed or kurtotic, which also conflicts with the assumption of a symmetric student's t-distribution. For example, given $(N, k) = (100, 1.2)$ and supposing that d_j follows a normal distribution with mean and standard deviation equal to one, the distribution of the sample moment $\hat{\mu}_i = \frac{1}{N}\sum_{j=1}^{N} s_j^i$ is long-tailed when $i \geq 3$ – see Figure 1b. Consequently, we take a new approach to estimate moment confidence intervals.

We propose using the percentile Bootstrap method to approximate confidence intervals of the sales moments μ_i; Algorithm 1 in Appendix A of the E-Companion Appendix provides the details. Although the percentile Bootstrap method is asymptotically valid only when the sample size N is large, it is a mature technique that has been widely utilized in the field of applied statistics even when N is small. Moreover, many experiments show that the percentile Bootstrap method provides fairly accurate estimation of confidence intervals for most distributions (e.g., see DiCiccio & Efron (1996) and Efron (1987)).

To obtain a good estimation of a moment's confidence interval, the M in the algorithm should be large. For example, if we set $M = 10^6$, the resampled $\{z_j^i\}_{j=1,\ldots,M}$ are symmetrically distributed under most conditions and an associated, say $\alpha_s = 95\%$, confidence interval is reasonably accurate. However, if the percentile Bootstrap distribution is not nearly symmetric (e.g., when the actual demand distribution is highly skewed), the percentile Bootstrap method may fail to provide satisfactory results.

[1] The i-th (raw) moment of sales is defined as $\mu_i = \mathbb{E}[s^i]$, so the moment index i is exactly the power (exponent) applied to the sales random variable. In addition, the two-sided critical value satisfying $\mathbb{P}(|T_{N-1}| \leq t_{\alpha,N-1}) = \alpha$.

In such a situation, one may switch to the Bootstrap-t technique (e.g., see Efron & Tibshirani (1994) and Reiser et al. (2017)).

Given a significance level α_s, we define the upper bound of the i-th moment's confidence interval as $\overline{m}_i(\alpha_s) = \bar{z^i}_{(\lceil (1+\alpha_s)M/2 \rceil)}$ and the lower bound as $\underline{m}_i(\alpha_s) = \bar{z^i}_{(\lfloor (1-\alpha_s)M/2 \rfloor)}$, for $i = 1, \ldots, \ell$, where ℓ is the highest moment order considered. For example, given $\alpha_s = 95\%$, we have $\overline{m}_i(95\%) = \bar{z^i}_{(\lceil 0.975M \rceil)}$ and $\underline{m}_i(95\%) = \bar{z^i}_{(\lfloor 0.025M \rfloor)}$. To further simplify notation, we shall use $\overline{m}_i$ and $\underline{m}_i$ to denote the upper and lower bounds of the i-th moment, with α_s implied, and consequently we have the confidence intervals $\mu_i \in [\underline{m}_i, \overline{m}_i]$, for $i = 1, \ldots, \ell$.

Given confidence intervals for the first two moments of sales (i.e., censored demand), we formulate a new distributionally robust newsvendor problem as follows:

$$\begin{aligned} \max_{q \in [0,k]} \quad \min_{F \in \mathcal{D}} \quad & \mathbb{E}_F \left[r \min(x, q) - cq \right] \\ \text{s.t.} \quad & \int_\Omega dF(x) = 1, \\ & \int_\Omega \min(k, x)^i dF(x) \in [\underline{m}_i, \overline{m}_i], \ \forall F \in \mathcal{D}, \ i = 1, 2. \end{aligned} \tag{2}$$

The first constraint ensures that F is a valid cumulative distribution function on Ω. The moment constraints define the ambiguity set by requiring that, for $i \in \{1, 2\}$, $\mu_i(F) \triangleq \int_\Omega \min(k, x)^i \, dF(x) \in [\underline{m}_i, \overline{m}_i]$. Since k is the censoring threshold that generated the historical data, the order decision must satisfy $q \leq k$ when the same threshold applies in the decision period.

This single-threshold formulation should therefore be read as a model for datasets generated under a common historical censoring threshold. If the observed quantity in a particular application is $\min(d, q^{\text{hist}})$ and the historical order-up-to level is constant across the sample, then we are exactly in this setting with $k = q^{\text{hist}}$. When the historical censoring thresholds vary across observations, the appropriate extension is the multi-threshold model in Section 4.

3.1.1 Selection of the Significance Level α_s.

Before solving Problem (2), we select a significance level α_s such that the resulting confidence intervals for the first two moments of *sales* are jointly consistent. Specifically, we require that every pair of moments within the two confidence intervals corresponds to at least one valid probability distribution of sales. This requirement is imposed to ensure the ambiguity set defined by interval constraints is a coherent Cartesian-product uncertainty set and ensure that the moment bounds do not include unattainable (internally inconsistent) moment pairs. The following lemma provides a tractable way to allows us to verify whether a given α_s satisfies this joint-consistency requirement.

Lemma 1. *The following are true:*

1. *The non-negative vector (μ_1, μ_2) is a feasible pair of sales moments if and only if the associated moment matrix is positive semidefinite,* $\mathcal{M} \triangleq \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \succeq 0$, *and* $k \geq \frac{\mu_2}{\mu_1}$ *(assume $\mu_1 > 0$).*
2. *The conditions $\mathcal{M} \succeq 0$ and $k \geq \frac{\mu_2}{\mu_1}$, $\forall (\mu_1, \mu_2) \in [\underline{m}_1, \overline{m}_1] \times [\underline{m}_2, \overline{m}_2]$ are equivalent to $\underline{m}_2 \geq \overline{m}_1^2$ and $k \geq \frac{\overline{m}_2}{\underline{m}_1}$.*

Lemma 1 provides necessary and sufficient conditions for the existence of a valid probability distribution for any pair of sales moments in the uncertainty set, thus ensuring the well-posedness of Problem (2). If an exogenous significance level α_s does not satisfy the necessary and sufficient conditions, $\underline{m}_2 \geq \overline{m}_1^2$ and $k \geq \frac{\overline{m}_2}{\underline{m}_1}$, we can lower the significance level α_s until the conditions hold[2]. In Section 5.1, we will demonstrate numerically that, for a given sample size and censoring threshold (N, k), there exists an upper bound $\bar{\alpha}_s \in [0, 1]$, such that the necessary and sufficient conditions of Lemma 1 are satisfied when $\alpha_s \leq \bar{\alpha}_s$.

3.1.2 A Closed-Form Optimal Order Quantity.

Now, we are ready to characterize the optimal order quantity in closed-form. The subscript C represents the maximin model with confidence intervals for *censored* demands.

Theorem 1. *If $\underline{m}_2 \geq \overline{m}_1^2$ and $k \geq \frac{\overline{m}_2}{\underline{m}_1}$, then the optimal order quantity q_C^* is*

(1) When $\frac{\underline{m}_1}{\sqrt{\nu}} > \sqrt{\frac{c}{r-c}}$,

$$q_C^* = \begin{cases} \underline{m}_1 + \frac{\sqrt{\nu}}{2}\left(\sqrt{\frac{r-c}{c}} - \sqrt{\frac{c}{r-c}}\right), & \text{if } k > \underline{m}_1 + \sqrt{\frac{(r-c)\nu}{c}}, \\ k, & \text{otherwise.} \end{cases}$$

(2) When $\frac{\underline{m}_1}{\sqrt{\nu}} \leq \sqrt{\frac{c}{r-c}}$,

$$q_C^* = \begin{cases} 0, & \frac{c}{r} > \frac{\underline{m}_1}{k}, \\ k, & \text{otherwise,} \end{cases}$$

where $\nu = \overline{m}_2 - \underline{m}_1^2$ is the largest possible variance of sales.

In Theorem 1, the optimal order quantity is a function of the lower bound of the first moment, $\underline{m}_1$, together with the upper bound of the second moment, $\overline{m}_2$. In other words, when moment information is estimated via sales data, we use the lowest mean $\underline{m}_1$ and the highest variance $\nu = \overline{m}_2 - \underline{m}_1^2$ to characterize the order quantity, which implicitly illustrates the conservatism of the maximin model. Moreover, our optimal maximin order quantity restores that of the maximin baseline in Section 2 when both $N \to \infty$ and $k \to \infty$: $\lim_{k,N\to\infty} q_C^* \to q_B^*$.

3.1.3 Worst-Case Sales Distribution.

Given the first two moments of demand, Scarf (1958) demonstrates that the underlying worst-case demand distribution that determines the order quantity q_B^* in the baseline model is a two-point distribution. Interestingly, if we only have access to confidence intervals for the first two *sales* moments, the worst-case *sales* distribution, that determines the order quantity q_C^*, is also a two-point distribution.

[2] As $\alpha_s \to 0$, we have $\underline{m}_i = \overline{m}_i = \bar{z}_{M/2}^i$, where $i = 1, 2$. Note that the index parameter $M/2$ is fixed, so $(\mu_1, \mu_2) = (\bar{z}_{M/2}^1, \bar{z}_{M/2}^2)$. By definition, for a given $M/2$, the $\bar{z}_{M/2}^1 = \sum_{j=1}^N z_j/N$ is the sample mean and $\bar{z}_{M/2}^2 = \sum_{j=1}^N z_j^2/N$ is the sample second moment. Since the random variable z_j follows a valid probability distribution, its associated sample mean and sample second moment must satisfy the necessary and sufficient conditions in Lemma 1. As a result, we conclude that when $\alpha_s \to 0$, the conditions in Lemma 1 must be satisfied.

Theorem 2. *The worst-case sales distribution that determines the order quantity q_C^* can be a two-point distribution with first moment equaling $\underline{m}_1$ and second moment equaling $\overline{m}_2$.*

Remark 1 (Connection to a truncated-demand DRO counterpart). Let $y = \min(d, k)$ denote the truncated demand (equivalently, sales). For any feasible $q \leq k$, $\min(d, q) = \min(y, q)$ holds. Consequently, Problem (2) can be interpreted as a distributionally robust newsvendor model, in the spirit of Scarf (1958), written directly in terms of the truncated variable y. The structural logic of the solution in Theorem 1 follows the classical two-moment maximin approach: it is driven by the smallest admissible mean and the largest admissible variance of the relevant random variable. The difference is that, in our setting, this variable is the sales distribution with support $[0, k]$, rather than the latent demand. The core contribution of our formulation lies in the data-driven construction of the ambiguity set from censored observations, the well-posedness conditions established in Lemma 1, and the fact that Theorem 2 identifies a worst-case sales distribution. We also see in Section 3.2 that adding censoring-probability information further tightens this truncated-demand ambiguity set and changes the resulting solution.

Thus, after taking $y = \min(d, k)$ as the primitive random variable, Problem (2) is the natural counterpart of Scarf (1958). The contribution here is therefore not a fundamentally new maximin structure, but a data-driven construction of a valid ambiguity set from censored sales observations, together with the associated well-posedness conditions and tractable robust decision rule.

3.2 The Probability of Demand Censoring

To allow our model to use additional available information, we now extend it by incorporating information about the complementary probability of demand censoring: $p \triangleq P(d \leq k) = \int_0^k dF(x)$. Given N observations of sales, we count the number of observations that are equal to the censoring threshold k: $n(k) = \sum_{i=1}^N \mathbb{1}\{s_i = k\}$. By Borel's Law of Large Numbers (Chandra, 2012), we have

$$\lim_{N \to \infty} \frac{n(k)}{N} = 1 - p,$$

where $1 - p$ is the probability of demand censoring (almost-sure convergence). For finite N, there is a non-trivial estimation error, and we calculate the confidence interval for p using the percentile Bootstrap method.

Given the confidence interval for p, $[\underline{m}_0, \overline{m}_0]$ (i.e., $\overline{m}_0 = \bar{p}_{(\lfloor (1+\alpha_s)M/2 \rfloor)}$ and $\underline{m}_0 = \bar{p}_{(\lceil (1-\alpha_s)M/2 \rceil)}$), we update the overall robust constraints to $(p, \mu_1, \mu_2) \in \bigtimes_{i=0}^2 [\underline{m}_i, \overline{m}_i]$. Next, we study the data-driven distributionally robust newsvendor model with the updated robust constraints $(p, \mu_1, \mu_2) \in \bigtimes_{i=0}^2 [\underline{m}_i, \overline{m}_i]$ as a demonstrative example of how this probabilistic information can be incorporated into our models. We again derive a closed-form order quantity, which we compare with the order quantity q_C^* from Section 3.1.2, that does not consider p. In other words, we evaluate the benefit of explicitly incorporating information regarding the probability of demand censoring. We modify the maximin model as follows:

$$
\begin{aligned}
\max_{q\in[0,k]} \min_{F\in\mathcal{D}} \quad & \mathbb{E}_F\left[r\min\left(x,q\right)-cq\right] \\
\text{s.t.} \quad & \int_\Omega dF(x)=1, \\
& \int_\Omega \mathbb{1}\{x<k\}dF(x)\in[\underline{m}_0,\overline{m}_0],\ \forall F\in\mathcal{D} \\
& \int_\Omega \min(k,x)^i dF(x)\in[\underline{m}_i,\overline{m}_i]\ \forall F\in\mathcal{D},\ i=1,2.
\end{aligned}
\tag{3}
$$

3.2.1 Selection of the Significance Level α_s.

To solve Problem (3), we must again select an appropriate significance level α_s, so that there exists a valid probability distribution F for any $(p,\mu_1,\mu_2)\in\bigtimes_{i=0}^{2}[\underline{m}_i,\overline{m}_i]$. Since Problem (3) is an extension of Problem (2), the conditions in Lemma 1 must hold, namely $\underline{m}_2\geq\overline{m}_1^2$ and $k\geq\frac{\overline{m}_2}{\underline{m}_1}$, to ensure that any vector $(\mu_1,\mu_2)\in[\underline{m}_1,\overline{m}_1]\times[\underline{m}_2,\overline{m}_2]$ admits a valid distribution F. We extend the logic of Lemma 1 for p in the following lemma.

Lemma 2. *The following are true:*

1. *For a valid probabilistic distribution F having statistics p and μ_1, $1-\frac{\mu_1}{k}\leq p$ must be satisfied.*
2. *The condition $1-\frac{\mu_1}{k}\leq p,\ \forall(p,\mu_1)\in[\underline{m}_0,\overline{m}_0]\times[\underline{m}_1,\overline{m}_1]$, is equivalent to $1-\frac{\underline{m}_1}{k}\leq\underline{m}_0$.*

3.2.2 A Closed-Form Optimal Order Quantity.

Supposing that confidence intervals for p and μ_i, $i\in\{1,2\}$, satisfy all conditions listed in Lemmas 1 and 2, the firm's optimal order quantity, in closed-form, is shown in the next theorem, where the subscript S denotes the case where the probability of censoring is incorporated.

Theorem 3. *If $\underline{m}_2\geq\overline{m}_1^2$, $k\geq\frac{\overline{m}_2}{\underline{m}_1}$, and $1-\frac{\underline{m}_1}{k}\leq\underline{m}_0$, then the optimal order quantity q_S^* is*

(1) When $\frac{\underline{m}_1}{\sqrt{\nu}}\geq\sqrt{\frac{c}{r-c}-\frac{\gamma}{c-\gamma r}\frac{(ck-r\underline{m}_1)^2}{\nu(r-c)}}$,

$$
q_S^*=\begin{cases}\frac{\underline{m}_1-\gamma k}{1-\gamma}+\frac{\sqrt{\nu(1-\gamma)-\gamma(k-\underline{m}_1)^2}}{2(1-\gamma)}\left(\sqrt{\frac{r-c}{c-\gamma r}}-\sqrt{\frac{c-\gamma r}{r-c}}\right), & \text{if } k\geq\underline{m}_1+\sqrt{\frac{r-c}{c}\nu} \text{ and } \gamma<\frac{c}{r},\\ k, & \text{otherwise,}\end{cases}
$$

(2) When $\frac{\underline{m}_1}{\sqrt{\nu}}\leq\sqrt{\frac{c}{r-c}-\frac{\gamma}{c-\gamma r}\frac{(ck-r\underline{m}_1)^2}{\nu(r-c)}}$,

$$
q_S^*=\begin{cases}0, & \text{if } \frac{c}{r}>\frac{\underline{m}_1}{k},\\ k, & \text{otherwise,}\end{cases}
$$

where $\nu=\overline{m}_2-\underline{m}_1^2$ and $\gamma=1-\overline{m}_0$.

The order quantity q_S^* is determined by the lower bound on the sales mean $\underline{m}_1$, the upper bound on the sales variance $\nu=\overline{m}_2-\underline{m}_1^2$, and the lower bound on the probability of censoring $\gamma=1-\overline{m}_0$ ($\overline{m}_0$ is the upper bound of $P(d\leq k)$). In particular, q_S^* restores q_C^*, from Section 3.1.2, if $\gamma=0$. To better measure the impact of incorporating the probability of demand censoring, we compare the two order quantities q_S^* and q_C^* in the next lemma. We focus on the "non-trivial" quantities, ignoring the cases where 0 or k are ordered; in other words, we focus on the first order quantities in the first cases of Theorems 1 and 3.

Lemma 3. *The non-trivial order quantity of Problem* (3) *is always at most the non-trivial order quantity of Problem* (2)*:* $q_S^* \leq q_C^*$.

Lemma 3 demonstrates that incorporating the censoring probability $p = \mathbb{P}(s = k)$ leads to a more conservative order quantity. From a distributional perspective, this shift occurs because the ambiguity set in Problem (3) is a subset of the one in Problem (2). By explicitly accounting for the mass at the censoring point k, we effectively restrict the worst-case sales distributions to a mixed-type setting with an atom at k. This additional information reduces the maximum admissible variance within the ambiguity set, leading to a tighter robust solution. Finally, we obtain the same limit conclusion as that in Section 3.1.2: $\lim_{k,N\to\infty} q_S^* \to q_B^*$.

3.2.3 Worst-Case Sales Distribution.

We now examine the worst-case sales distribution that determines the optimal order quantity q_S^*.

Theorem 4. *The sales distribution that determines the optimal order quantity* q_S^* *can be a three-point distribution with first moment* $\underline{m}_1$*, second moment* $\overline{m}_2$*, and probability of censoring* $\mathbb{P}(\min(d, k) = k) = 1 - \overline{m}_0$.

By incorporating the probability of demand censoring into the model, the worst-case sales distribution becomes a three-point distribution. Comparing to the two-point sales distribution in Theorem 2, the "extra" point in the sales distribution comes from a mass of $\gamma = 1 - \overline{m}_0$ at point k. Furthermore, supposing that $\gamma = 0$, we have that $q_S^* = q_C^*$, and the three-point sales distribution in Theorem 4 restores the two-point sales distribution in Theorem 2.

Remark 2 (Interpretation of the censoring-probability constraint). Under the continuous-demand setting of Section Section 2, the sales variable $y = \min(d, k)$ can be interpreted as a mixed distribution consisting of a continuous part on $[0, k)$ and an atom at k. Section 3.2 incorporates explicit information regarding the magnitude of this atom: the lower bound $\gamma = 1 - \overline{m}_0$ represents the probability mass for $\mathbb{P}(y = k)$. Relative to Problem (2), Problem (3) imposes one additional structural restriction on the ambiguity set, namely a lower bound on the probability mass at k. As a result, the ambiguity set is tighter, the worst-case sales distribution changes from a two-point distribution to a three-point distribution with one support point fixed at k, and the resulting non-trivial robust order quantity weakly decreases, i.e., $q_S^* \leq q_C^*$. When $\gamma = 0$, this additional restriction disappears and Problem (3) reduces to Problem (2).

4 A Data-Driven Maximin Model with Multiple Censoring Levels

An important assumption in the previous sections is that all sales observations are censored at the same censoring threshold k. In other settings, however, historical sales may be censored at multiple historical censoring thresholds, for example because the firm used different order-up-to levels or shelf-space allocations across periods (e.g., during an experimental or exploratory phase to learn demand). In this section, we consider sales data censored at multiple levels and study how these groups of censored observations affect the optimal robust order quantity.

Suppose we have $M > 1$ groups of censored demand. Let the ith group of sales be censored by threshold k_i, where we assume that the indexing is such that $k_i < k_j$ for $i < j$. As in previous sections, assume that, for each group of data, we have confidence intervals for the first two moments of the sales and the probability of demand censoring. Model (3) is extended as follows:

$$\begin{aligned}
\max_{q\geq 0} \quad \min_{F\in\mathcal{D}} \quad & \int_\Omega r\min(x,q)dF(x) - cq \\
\text{s.t.} \quad & \int_\Omega dF(x) = 1,\ \forall F \in \mathcal{D} \\
& \int_\Omega \mathbb{1}\{x < k_i\}dF(x) \in [\underline{m}_{i,0}, \overline{m}_{i,0}],\ \forall F\in\mathcal{D},\ i = 1,\cdots,M \\
& \int_\Omega \min(x,k_i)^j dF(d) \in [\underline{m}_{i,j}, \overline{m}_{i,j}],\ \forall F\in\mathcal{D},\ i=1,\cdots,M,\ j=1,2,
\end{aligned} \tag{4}$$

where $[\underline{m}_{i,0}, \overline{m}_{i,0}]$ is the confidence interval for the probability that demand is *not* censored by threshold k_i and $[\underline{m}_{i,j}, \overline{m}_{i,j}]$ is the confidence interval for the j-th sales moment that is censored by threshold k_i. We assume all confidence intervals satisfy the sufficient conditions in Lemmas 1 and 2.

4.1 A Solution by Semidefinite Programming.

By strong duality (Shapiro, 2001), the inner minimization of Problem (4) can be replaced by its dual. Conditioning on the location of the order quantity q among the sorted censoring levels $k_1 < k_2 < \cdots < k_M$, we can write the dual of the inner minimization problem as the maximum of M semidefinite programs.

Theorem 5. *Given confidence intervals for the probability of demand censoring and the first two sales moments for all M groups of data, $[\underline{m}_{i,j}, \overline{m}_{i,j}]$, where $i = 1,\cdots,M$ and $j = 0,1,2$, if $q \in (k_J, k_{J+1}]$, the dual of the inner minimization problem in* (4) *is equivalent to the following semidefinite optimization* *problem:*

$$
\begin{aligned}
V(q|q\in(k_J,k_{J+1}]) \triangleq \quad \max \quad & \frac{1}{2}\sum_{j=0}^{2}\sum_{i=1}^{M}\left[(\overline{m}_{i,j}+\underline{m}_{i,j})y_{i,j}+(\overline{m}_{i,j}-\underline{m}_{i,j})z_{i,j}+y\right] \\
s.t. \quad & 0=\sum_{i+j=2n-1} x^w_{i,j},\ n=1,2;\ w=0,\cdots,J-1, \\
& \Lambda^w(r,0;k_w,k_{w+1})=\sum_{i+j=2n} x^w_{i,j},\ n=0,1,2;\ w=0,\cdots,J-1, \\
& 0=\sum_{i+j=2n-1} v_{i,j},\ n=1,2, \\
& \Lambda^w(r,0;k_J,q)=\sum_{i+j=2n} v_{i,j},\ n=0,1,2;\ w=J, \\
& 0=\sum_{i+j=2n-1} u_{i,j},\ n=1,2, \\
& \Lambda^w(0,rq;q,k_{J+1})=\sum_{i+j=2n} u_{i,j},\ n=0,1,2;\ w=J, \\
& 0=\sum_{i+j=2n-1} z^w_{i,j},\ n=1,2,\ w=J+1,\cdots,M-1, \\
& \Lambda^w(0,rq;k_w,k_{w+1})=\sum_{i+j=2n} z^w_{i,j},\ n=0,1,2,\ w=J+1,\cdots,M-1, \\
& X^w,Z^w,V,U\succeq 0,\ \forall w, \\
& y_{i,j}\in[z_{i,j},-z_{i,j}],\quad i=1,\cdots,M,\ j=0,1,2, \\
& z_{i,j}\le 0,\quad i=1,\cdots,M,\ j=0,1,2,
\end{aligned}
$$

where $\Lambda^w(\cdot)$ is the SDP-representable polynomial function: $\Lambda^w(A,B;a,b)=\sum_{i=0}^{n}\sum_{j=i}^{2+i-n} Y_j\binom{j}{i}\binom{2-j}{n-i}a^{j-i}b^i$, $Y_0=-\sum_{j=1}^{2}\sum_{i=1}^{w} y_{i,j}k_i^j-\sum_{i=w+1}^{M} y_{i,0}-cq+B$, $Y_1=-\sum_{i=w+1}^{M} y_{i,1}+A$, *and* $Y_2=-\sum_{i=w+1}^{M} y_{i,2}$. *Moreover,* $V(q|q\in(k_J,k_{J+1}])$ *is concave in q.*

The SDP problem $V(q|q\in(k_J,k_{J+1}])$ is concave in q, so we can always find an optimal quantity q^*_J such that $q^*_J\in\arg\max_{q\in(k_J,k_{J+1}]} V(q|q\in(k_J,k_{J+1}])$ using, say, a golden-section search method. Solving the conditional problem $V(q|q\in(k_J,k_{J+1}])$ only provides the optimal solution in the restricted domain $q\in(k_J,k_{J+1}]$. To solve the master problem (4), namely $\max_{0\le q\le k_M} V(q)$, we need to search the best solutions for all subproblems; this is equivalent to solving $\max\left\{\max_{q\in(k_0,k_1]} V(q|q\in(k_0,k_1]),\ \cdots,\ \max_{q\in(k_{M-1},k_M]} V(q|q\in(k_{M-1},k_M])\right\}$, where each is solvable via an SDP, possibly in parallel, per Theorem 5. We conclude by noting that these SDP problems are efficiently solvable (i.e., in seconds to minutes) for reasonable values of M; further details are provided in Section 5.

It is also useful to clarify how our framework relates to, and differs from, existing DRO/moment-problem techniques (e.g., Bertsimas & Popescu (2005)). While we build on similar duality-based ideas, our setting is fundamentally driven by censored observations (i.e., demand data). In the single-censoring model, this censoring enters through moment constraints of the form $\mathbb{E}[\min(d,k)^j]$ (see model (2)), and in the multi-level setting through constraints of the form $\mathbb{E}[\min(d,k_i)^j]$ across multiple thresholds k_i (see model (4)).

At a technical level, the dual of our inner moment problem is closely related in spirit to Bertsimas & Popescu (2005), but censoring induces a different dual structure. In particular, the dual constraints become piecewise in the demand variable, with separate regimes (e.g., $d \in (0, k]$ and $d \in (k, \infty)$ in the single-level case), whereas the uncensored moment problems are defined over a single domain $(0, \infty)$. With multiple censoring levels, the partition becomes finer (e.g., $d \in (k_w, k_{w+1}]$), leading to a more subtle indexing of the dual constraints. As a consequence, in our formulation the resulting computation is not represented by a single SDP in general; rather, it can be expressed as the maximum over a finite collection of SDPs (one for each relevant interval induced by the censoring levels). These structural differences are specific to censored-demand information and are not present in the standard uncensored moment-DRO formulations.

Several extensions are natural. First, higher-order censored moments (i.e., $\ell > 2$) can be incorporated by adding corresponding constraints and applying the same SDP machinery to the inner problem, followed by the same outer search over q. Second, the ambiguity set can be strengthened using richer summary information – e.g., partitioned (segment-wise) moments, asymmetry measures, or other censoring-induced statistics – along the lines of the partitioned-information ideas in Natarajan et al. (2018). These directions can enrich the model while retaining tractability.

4.2 An Illustrative Multi-period Implementation

This subsection presents a stylized multi-period implementation of the multi-threshold model when successive historical order quantities generate successive censoring levels. Note that our proposed implementation is not intended as a general exploration policy. In particular, if the initial censoring level is too small, the accumulated sales data contain limited information about the upper tail of demand, creating an "information bottleneck" (Lariviere & Porteus, 1999) that a purely rolling update may not overcome. For that reason, the procedure below is most informative when the initial threshold is deliberately large, or when exogenous experimentation provides several censoring levels. Our core theoretical results in Sections 3.1–4 do not depend on this subsection.

We revisit our model from a multi-period perspective. In each update period, the firm chooses an order quantity using confidence intervals constructed from sales data generated under previous order decisions, and the current order decision becomes a future censoring level. This repeated-update perspective is related in spirit to Saghafian & Tomlin (2016), which studies Bayesian updating in a repeated newsvendor setting without censoring and under distributional assumptions.

We restate the general multi-threshold model (4) as

$$
\begin{aligned}
\max_{q \geq 0} \; f(q, \mathbf{k}, \mathbf{m}) = \max_{q \geq 0} \quad \min_{F \in \mathcal{D}} \quad & \mathbb{E}_F[r \min(x, q) - cq] \\
\text{s.t.} \quad & \int_\Omega dF(x) = 1, \\
& \int_\Omega \mathbb{1}\{x < k_i\} \, dF(x) \in [\underline{m}_{i,0}, \overline{m}_{i,0}], \quad i = 1, \dots, M, \\
& \int_\Omega \min(x, k_i)^j \, dF(x) \in [\underline{m}_{i,j}, \overline{m}_{i,j}], \quad i = 1, \dots, M, \; j = 1, 2.
\end{aligned}
$$

where $\mathbf{k} \triangleq (k_1, \dots, k_M)$ and $\mathbf{m} \triangleq \{\underline{m}_{i,j}, \overline{m}_{i,j}\}_{i,j}$.

Sequential implementation. Our implementation follows the rolling (sequential) manner. In update period t, we set the state to be the collection of previous order quantities, $\mathbf{k}_t = (k_1, \ldots, k_t) = (q_0, \ldots, q_{t-1}^*)$. Given $\mathbf{k}_t$ and the confidence intervals $\mathbf{m}_t$ constructed from all accumulated sales data, we compute the next decision by solving $q_t^* \in \arg\max_{q_t} f(q_t, \mathbf{k}_t, \mathbf{m}_t)$, and then set $k_{t+1} = q_t^*$. This rule is intentionally myopic: it repeatedly re-solves the static DRO problem using the accumulated data and does not separately optimize experimentation. We use it here only to illustrate how the multi-threshold model can be implemented sequentially. In particular, it is not a general exploration policy and is not designed to recover from an initially uninformative low threshold.

Numerical illustration. We implement the rolling procedure in blocks. Starting from an initial order quantity q_0, the firm applies the same quantity for n consecutive periods, then uses the resulting sales observations to update the confidence intervals and compute the next order quantity. At stage t, the update uses all data collected so far under censoring levels $\{q_0, q_1^*, \ldots, q_{t-1}^*\}$. For simplicity, we fix $n = 10$ throughout.[3]

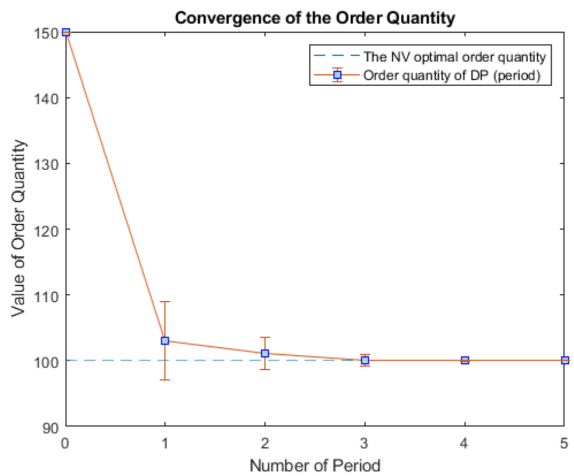


Figure 2: Convergence of order quantities to the optimal newsvendor solution.

5 Numerical Experiments

In our numerical studies, we utilize real sales data downloaded from Kaggle.com. Since there is no evidence of censoring (see Appendix B in E-Companion Appendix), we assume that this sales data set is a demand data set. We generate various sales data sets by censoring the raw demand data by different inventory capacities. To further support conclusions obtained using the real dataset, numerical studies that are based on synthetic demand data are also provided in Appendix B of the E-Companion Appendix; we use well-behaved gamma and normal distributions, a heavy-tailed lognormal

[3]Hall (1992) recommends (page 286) at least ten data points for bootstrap-based procedures.

In the numerical study, demand is normally distributed with $\mu = 100$ and $\sigma = 20$. With $r = 2$ and $c = 1$, the optimal newsvendor quantity is 100. We initialize the procedure at $q_0 = 1.5\mu$, so that the first block of observations is only lightly censored and therefore informative about the upper tail. In each update period, the confidence intervals $\mathbf{m}_t$ are obtained by bootstrap from the accumulated censored sales data, using 10,000 resamples. We repeat the simulation 100 times and summarize the resulting path in Figure 2. In this high-initial-threshold setting, the order quantity moves quickly toward the optimal benchmark and then stabilizes, illustrating the rolling implementation in a high-initial-threshold setting.

distribution, and an ill-behaved beta distribution, which, using parameters $a, b < 1$, is U-shaped. To begin, we provide a brief overview of the Kaggle data.

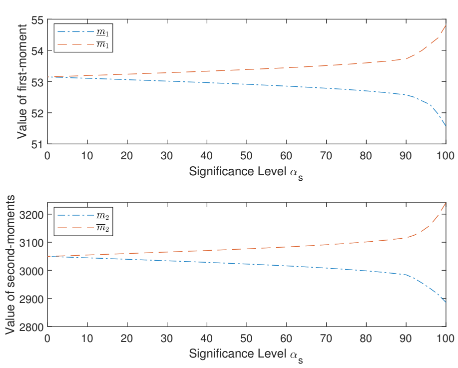
The real sales data set consists of product sales at a set of retail stores, available at `https://www.kaggle.com/c/demand-forecasting-kernels-only`. There are four variables in this data set: date, store, item ID and sales. In our numerical experiments, we select the data for a single product at a single store (randomly selected), from January 1, 2011 to December 31, 2017. There are 1826 observations, d_i for $i = 1, \ldots, 1826$, with mean 53.2 and standard deviation 15.0.

In Section 5.1, we provide a numerical study to identify the values of α_s that admit feasible distributions for any vector of moments in the ambiguity set and we propose finding an appropriate α_s value through $\mathcal{K}$-fold cross validation. In Section 5.2 we provide sensitivity analysis for our model's order quantity and (realized, not worst-case) expected profit. Finally, in Sections 5.3 and 5.4, we compare our model with two benchmarks: (i) a simple parametric de-censoring strategy (normal-based) and (ii) the Kaplan–Meier (KM) myopic policy for censored demand.

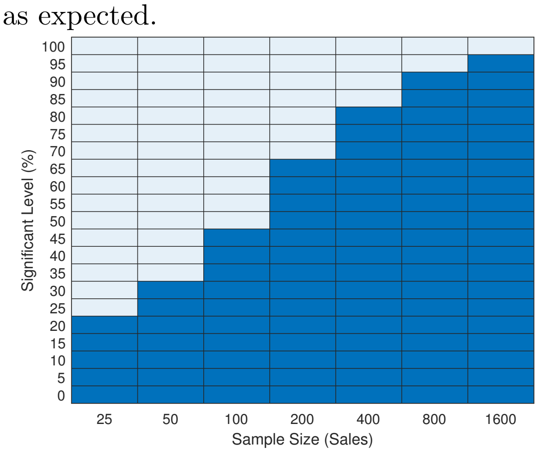
5.1 Selection of Significance Level α_s

In our models, a required preprocessing step is to select a proper significance level α_s to ensure that there exists a valid probability distribution F for any combination of moments contained in the ambiguity sets. A natural question arises: given a sample size N and a censoring threshold k, what values of α_s may we utilize in our models? We answer this question in this section.

In Figure 3-(a), we plot the upper and lower bounds of the first two moments, $\underline{m}_i$ and $\overline{m}_i$ for $i \in \{1, 2\}$, as functions of α_s, using the real data that are censored by censoring threshold $k = 115$, the maximum sales value in the data set. The upper and lower bounds are computed using the percentile Bootstrap method in Algorithm (see Appendix A in the E-Companion Appendix). We observe that the interval widths increase monotonically in α_s, as expected.



(a) Confidence interval.



(b) Real Data.

Figure 3: Impact of significance level α_s on ambiguity sets for sales moments.

Next, we examine this monotonicity property of the ambiguity sets $[\underline{m}_i, \overline{m}_i]$, $i \in \{1, 2\}$, in terms of Lemma 1, which stipulates conditions that ensure a valid distribution F for any combination of moments in the sets. In Figure 3-(b), the x-axis represents sample size $N \in \{25, 50, 100, 200, 400,$

800, 1600} and the y-axis represents significance level $\alpha_s \in \{0\%, 5\%, \ldots, 95\%, 100\%\}$. Given a significance level α_s, we randomly select N samples from the sales data set and re-calculate the ambiguity sets using the percentile Bootstrap method. If the associated ambiguity sets satisfy the necessary and sufficient conditions of Lemma 1, we color it as a dark cell; otherwise it is a light cell. We observe that, for a given (N, k) pair of values, there exists an upper bound $\bar{\alpha}_s$ where, if $\alpha_s \leq \bar{\alpha}_s$, the resulting ambiguity sets satisfy the necessary and sufficient conditions in Lemma 1. The conjecture below summarizes the main insights:

Conjecture 1.

1. *The lower bound of the i-th moment $\underline{m}_i$ is decreasing in α_s, while the upper bound of the i-th moment $\overline{m}_i$ is increasing in α_s.*

2. *Given inputs (N, k), there exists an upper bound $\bar{\alpha}_s$ for α_s, where the ambiguity set of the moments $[\underline{m}_1, \overline{m}_1] \times [\underline{m}_2, \overline{m}_2]$ admits a valid probability distribution F for any element of the set if and only if $\alpha_s \leq \bar{\alpha}_s$.*

3. *The upper bound $\bar{\alpha}_s$ is increasing in the sample size N.*

For example, in Figure 3-(b), given $N = 200$, we have $\bar{\alpha}_s \approx 65\%$, and all values of $\alpha_s \leq 65\%$ induce ambiguity sets that allow a valid distribution F for any element of the sets. In addition, $\bar{\alpha}_s$ increases from 65% to 95% when the sample size increases from $N = 200$ to $N = 1600$. For a given (N, k) pair of inputs, one can numerically find $\bar{\alpha}_s$ through a binary search.

For each candidate α_s, we construct the moment confidence intervals (and hence the ambiguity sets) using only the training data through censored sales observations. We then solve the corresponding model on the training set to obtain the optimal order quantity. Finally, we evaluate this decision on the test set using the raw demand observations via the sample-average realized profit formula. Once the upper bound $\bar{\alpha}_s$ and feasible range of α_s are available, we can find the significance level α_s^* that maximizes the out-of-sample performance using $\mathcal{K}$-fold cross validation, as follows. We first randomly permute the raw demand observations, and then divide the whole data set into $\mathcal{K}$ equal folds, where data in the first $\mathcal{K}$-1 folds are labeled as the training data Ω_{train} and data in the last fold are labeled as the testing data Ω_{test}. For a given α_s in the feasible range, we find the optimal order quantity $q(\alpha_s|\Omega_{train})$ on the training data and calculate the realized profit $\frac{1}{|\Omega_{test}|}\sum_{i=1}^{|\Omega_{test}|}\Pi_{test}$ on the test data, where $\Pi_{test} = \{r \min[d_i, q(\alpha_s|\Omega_{train})] - cq(\alpha_s|\Omega_{train}\}$. We repeat this procedure, letting each fold play the role of the test data set, and average the realized profits. The final score for the given α_s is the averaged profit over the $\mathcal{K}$ folds, and the best significance level α_s^* is the one that provides the highest final score. Finally, we emphasize that the confidence level is a hyper-parameter that is tuned and does not correspond to any probabilistic guarantee of model performance.

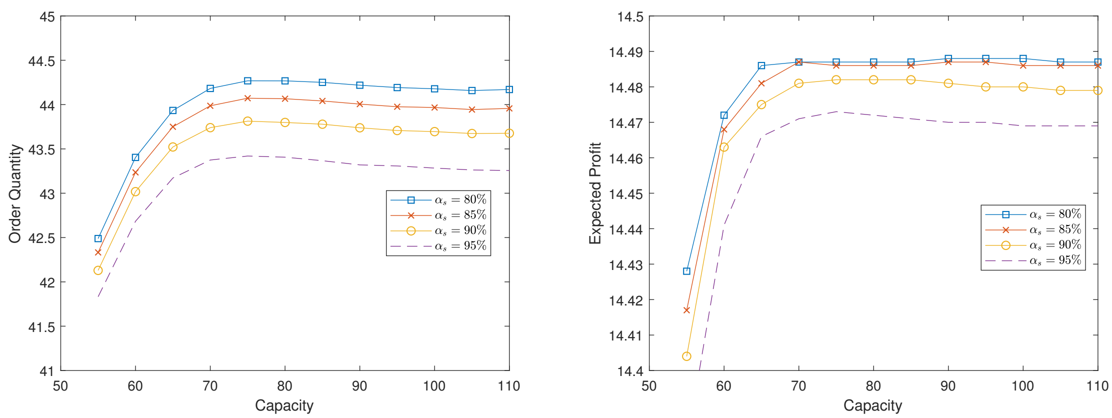
5.2 Sensitivity Analysis

In this section, we examine the order quantity and expected profit's dependence on the censoring levels k_i and the number of censoring levels M, for $\ell = 2$ (i.e., the first two moments of sales). We also performed sensitivity analysis for the significance level α_s and found that, as long as our models are feasible (c.f., Section 5.1) and the order quantity is not trivial, the profit does not depend strongly on the significance level α_s.

In our experiments, the order quantities are obtained via observations of the sales (censored demand), while the expected profit is computed using the raw demand data, rather than the worst-case profit objective of our models. For example, the expected profit is calculated as $\pi_s = \frac{1}{N}\sum_{i=1}^{N}\left[r\min(q, d_i) - cq\right]$, where q is the optimal order quantity in our DRO model.

5.2.1 Impact of Censoring Level.

To study the impact of the censoring level, we fix the unit revenue $r = 1.4$, assume that the demand data are censored by one inventory level $k \in [55, 110]$, and then numerically examine the dependence of the order quantities and profits on k. Figure 4 provides the order quantities and expected profits in the model.



(a) Order Quantity, Maximin Model

(b) Expected Profit, Maximin Model

Figure 4: Impact of censoring threshold k on order quantities and profits.

We examine the cases where $\alpha_s \in \{80\%, 85\%, 90\%, 95\%\}$. In Figure 4, despite minor fluctuations, the recommended order quantity is generally increasing in k, especially when k is relatively small. As expected, both the order quantity and the realized profit converge to their uncensored counterparts as $k \to \infty$. This behavior follows from the definition of censored sales $s = \min(d, k)$: when k is small, censoring is severe and the observed sales contain limited information about the upper tail of demand; increasing k reduces censoring and makes the sales closer to the true demand, which typically tightens the resulting ambiguity set and leads to less conservative order quantities. When k is large, censoring becomes negligible, so the solution progressively approaches that of the baseline newsvendor model with uncensored demand.

Observation 2. *The order quantity is increasing in the censoring threshold k when k is small (i.e., $k \leq 1.5m_1$), and then slightly decreases as k increases further (i.e, $k > 1.5m_1$).*

5.2.2 Impact of the Number of Censoring Levels M.

To study the impact of the number of censoring levels M, we partition the demand dataset into M groups, where $M \in \{1, 2, 3, 4, 5\}$. For each M, we associate each group with a censoring level and construct moment confidence intervals for the censored sales in that group. Specifically, the censoring capacities are $\{90\}$, $\{90, 100\}$, $\{80, 90, 100\}$, $\{80, 90, 100, 110\}$, and $\{70, 80, 90, 100, 110\}$ for $M = 1, 2, 3, 4, 5$, respectively. We fix unit revenue $r = 1.4$ and significance level $\alpha_s = 90\%$. Figure 5 reports the resulting optimal order quantities and realized profits. The histograms display the optimal order quantity (left axis) and the dashed line displays the realized profit (right axis). The total computing time for $M \in \{1, 2, 3, 4, 5\}$ is $\{< 0.1$s, 9.76s, 17.14s, 39.11s, 105.23s$\}$.

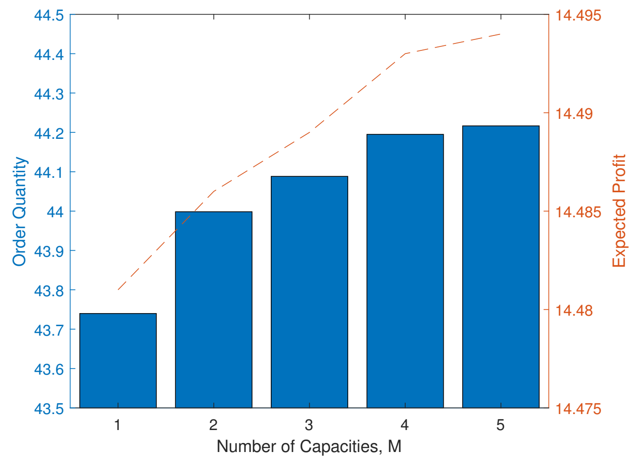


Figure 5: Impact of number of censored groups M on order quantities and profits.

Note that increasing M does not mean that we "consider fewer constraints." Rather, a larger M provides additional moment information about the underlying demand through multiple censoring thresholds, which typically tightens the ambiguity set and can lead to a less conservative worst-case distribution. This additional information can improve the resulting decision and realized profit. In our experiments, the realized profit increases with M for small values of M and then levels off, indicating diminishing returns from adding more censoring levels.

Observation 3. *In our experiments, the realized profit increases with the number of censoring levels M when M is small and then plateaus as M becomes larger.*

5.3 Comparisons with a Simple De-Censoring Strategy

In this section we compare our robust order quantities with a simple parametric benchmark for censored data under an assumed normal demand model. Specifically, we assume demand $D \sim \mathcal{N}(\mu, \sigma^2)$ and observed sales are censored at threshold k. Given the sample mean $\bar{x}$ and variance s^2 of the sales observations, we estimate the underlying demand parameters (μ, σ^2) by inverting the moment relationships implied by censoring (Cohen, 1959; Barr & Sherrill, 1999). Let ϕ and Φ denote the pdf and cdf of the standard normal distribution. We compute ξ^* as the unique solution to $\left(1 - \frac{\phi(\xi)}{\Phi(\xi)}\left(\frac{\phi(\xi)}{\Phi(\xi)} - \xi\right)\right) / \left(\frac{\phi(\xi)}{\Phi(\xi)} - \xi\right)^2 = s^2/(\bar{x} - k)^2$, and set $\hat{\theta} = \frac{\phi(\xi^*)/\Phi(\xi^*)}{\phi(\xi^*)/\Phi(\xi^*) - \xi^*}$. The demand mean and variance are then estimated by $\hat{\mu} = \bar{x} - \hat{\theta}(\bar{x} - k)$ and $\hat{\sigma}^2 = s^2 + \hat{\theta}(\bar{x} - k)^2$. We then apply the classical newsvendor solution under the fitted normal demand distribution $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$.

We systematically compare our robust models and this simple method through numerical studies (based on the Kaggle data). Specifically, we normalize the unit cost $c = 1$, consider different revenues

$r \in \{1.2, 1.4, 1.8, 2.6\}$, and censoring levels $k \in \{70, 80, 90, 100, 110\}$; note that the largest sales value in the Kaggle data set is 115. We compare both the order quantities and expected profits based on different methods. We observe that our model significantly outperforms the simple method when the censoring level is low (i.e., $k = 70$). In particular, we note that the simple method's order quantity q^*_{base} is significantly larger than our robust order quantity q^*_C when $k = 70$, which can be used to explain the poor performance of the simple method. However, when the censoring level is large, the ratio between our robust profits and the simple method's profit becomes approximately 100%, which indicates comparable performance between both methods; this is not surprising since censoring is less impactful. Finally, this simple approach only works if the demand distribution is normally distributed; we are not aware of another simple general-purpose technique that can handle non-normal demand distributions.

5.4 Comparisons with the KM-Myopic Approach

In this section, we compare the realized profit of our DRO policies with the Kaplan–Meier (KM) myopic benchmark of Huh et al. (2011). Given censored sales observations and their associated censoring levels, the KM estimator provides a consistent nonparametric estimate of the demand distribution under standard conditions, and the resulting KM-myopic policy can serve as a strong plug-in benchmark when the sample size is large. Specifically, suppose we observe sales from M censored groups, denoted by $\{(s_{(1,j)}, \delta_{(1,j)}), \ldots, (s_{(N,j)}, \delta_{(N,j)})\}$ for $j = 1, \ldots, M$, where $\delta_{(i,j)} = \mathcal{I}\{s_{(i,j)} < k_j\}$ indicates whether observation (i, j) is uncensored at level k_j. The KM estimator of the complementary cdf is $\bar{F}_{\mathrm{KM}}(x) = \prod_{(i,j):\, s_{(i,j)} \le x} \left(\frac{N-i}{N-i+1}\right)^{\delta_{(i,j)}}$, and the corresponding KM-myopic order quantity is $q^*_{\mathrm{KM}} = \inf\{x : \bar{F}_{\mathrm{KM}}(x) \le c/r\}$.

Given a sample size N, we divide the dataset into M equally sized groups, each censored by a prescribed censoring threshold. We use censoring capacities $\{90\}$, $\{80, 90, 100\}$, and $\{70, 80, 90, 100, 110\}$ for $M = 1, 3$, and 5, respectively, and set $\alpha_s = 90\%$ and $r = 1.4$. In Figure 6, the x-axis shows the sample size and the y-axis reports the profit ratios $\pi^*_C/\pi^*_{\mathrm{KM}}$ (left) and $\pi^*_{CR}/\pi^*_{\mathrm{KM}}$ (right). For our full dataset ($N = 1826$), we obtain $\pi^*_C/\pi^*_{\mathrm{KM}} \approx 100\%$, indicating that our DRO policy does not materially outperform the KM-myopic benchmark when data are abundant (especially for smaller M). We observe a similar pattern in additional stress tests across synthetic demand distributions, critical fractiles $1 - c/r$, and censoring levels: while the resulting order quantities may differ, the realized profits are typically very close (with an average profit ratio of 98.9% across the cases considered). In contrast, when the sample size is smaller and/or censoring is more pronounced (e.g., larger M), our DRO policies can yield substantially higher realized profit than the KM-myopic benchmark (e.g., $\pi^*_C/\pi^*_{\mathrm{KM}} \approx 123\%$ at $N = 400$ for $M = 5$ in Figure 6). As N increases, this advantage diminishes and the two approaches become comparable. A natural explanation is that our DRO construction explicitly accounts for estimation uncertainty through bootstrap-based moment confidence intervals, which can be particularly valuable in finite-sample settings with censored observations; by contrast, the KM-myopic policy is a plug-in approach whose performance improves as the nonparametric estimate stabilizes with more data. Overall, these results suggest that our DRO approach is most beneficial in data-limited regimes, while with abundant data it tends to match the KM-myopic benchmark.

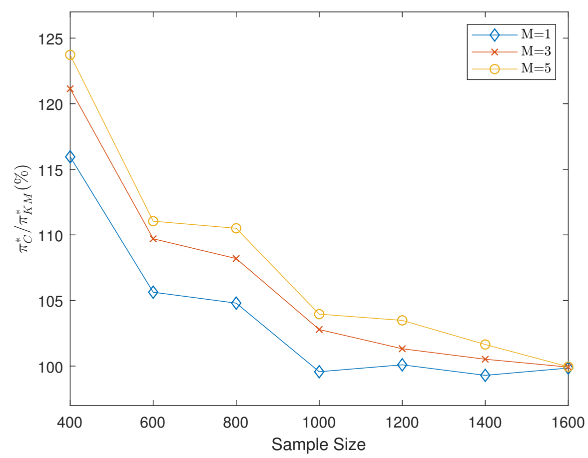


Figure 6: Ratio of Profits With Different Sample Sizes.

6 Conclusion

The primary contribution of our paper is to extend the distributionally robust newsvendor model literature in two important directions: 1) providing a *data-driven* framework and 2) assuming that *demand is censored*, so that only sales data are available. We propose such a model that generalizes the maximin model by Scarf (1958).

We provide a complete framework in our paper to solve such data-driven demand-censored distributionally robust newsvendor models. We first compute confidence intervals for sales moments for a given confidence level α_s, which serve as our ambiguity sets for the sales moments. We then show how to appropriately select a significance level α_s, so that these ambiguity sets allow a valid demand distribution to exist for any vector of moments in the ambiguity set; this ensures our models are well posed. We then solve the models. When considering only the first two sales moments, we derive an explicit closed-form expression for the optimal order quantity for the model. We also analyze two main model extensions: 1) incorporating the probability of demand censoring and 2) allowing sales data to be censored by multiple inventory capacities; these generalizations typically require semidefinite programming to solve.

There are many ways our paper can be extended. Our study is based on a single-item newsvendor problem. One interesting extension can be to study multiple products, possibly correlated, and possibly censored by different inventory levels. Another interesting problem would be to extend our models to allow covariate-driven demand, where the demand in each period may depend upon external features. Finally, we could enrich our models by considering price as a variable, with stochastic demand a function of this price.

References

Barr, D. R. & Sherrill, E. T. (1999). Mean and variance of truncated normal distributions. The American Statistician, 53(4), 357–361.

Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., & Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. Management Science, 59(2), 341–357.

Ben-Tal, A. & Hochman, E. (1976). Stochastic programs with incomplete information. Operations Research, 24(2), 336–347.

Bertsimas, D. & Popescu, I. (2005). Optimal inequalities in probability theory: A convex optimization approach. SIAM Journal on Optimization, 15(3), 780–804.

Bertsimas, D. & Sethuraman, J. (2000). Moment problems and semidefinite optimization. Handbook of semidefinite programming, 469–509. Springer.

Burnetas, A. N. & Smith, C. E. (2000). Adaptive ordering and pricing for perishable products. Operations Research, 48(3), 436–443.

Chandra, T. K. (2012). The Borel-Cantelli Lemma. Springer Science & Business Media.

Cohen, A. C. (1959). Simplified estimators for the normal distribution when samples are singly censored or truncated. Technometrics, 1(3), 217–237.

Delage, E. & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. Operations research, 58(3), 595–612.

DiCiccio, T. J. & Efron, B. (1996). Bootstrap confidence intervals. Statistical science, 189–212.

Ding, X., Puterman, M. L., & Bisi, A. (2002). The censored newsvendor and the optimal acquisition of information. Operations Research, 50(3), 517–527.

Efron, B. (1987). Better bootstrap confidence intervals. Journal of the American statistical Association, 82(397), 171–185.

Efron, B. & Tibshirani, R. J. (1994). An introduction to the bootstrap. CRC press.

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

Godfrey, G. A. & Powell, W. B. (2001). An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. Management Science, 47(8), 1101–1112.

Hall, P. (1992). The bootstrap and edgeworth expansion.

Huh, W. T., Levi, R., Rusmevichientong, P., & Orlin, J. B. (2011). Adaptive data-driven inventory control with censored demand based on kaplan-meier estimator. Operations Research, 59(4), 929–941.

Huh, W. T. & Rusmevichientong, P. (2009). A nonparametric asymptotic analysis of inventory planning with censored demand. Mathematics of Operations Research, 34(1), 103–123.

Lariviere, M. A. & Porteus, E. L. (1999). Stalking information: Bayesian inventory management with unobserved lost sales. Management Science, 45(3), 346–363.

Lu, M. & Shen, Z.-J. M. (2021). A review of robust operations management under model uncertainty. Production and Operations Management, 30(6), 1927–1943.

Lu, X., Song, J.-S., & Zhu, K. (2008). Analysis of perishable-inventory systems with censored demand data. Operations Research, 56(4), 1034–1038.

Natarajan, K., Sim, M., & Uichanco, J. (2018). Asymmetry and ambiguity in newsvendor models. Management Science, 64(7), 3146–3167.

Reiser, M., Yao, L., Wang, X., Wilcox, J., & Gray, S. (2017). A comparison of bootstrap confidence intervals for multi-level longitudinal data using monte-carlo simulation. Monte-Carlo simulation-based statistical modeling, 367–403. Springer.

Saghafian, S. & Tomlin, B. (2016). The newsvendor under demand ambiguity: Combining data with moment and tail information. Operations Research, 64(1), 167–185.

Scarf, H. (1958). A min-max solution of an inventory problem. Studies in the mathematical theory of inventory and production.

Shapiro, A. (2001). On duality theory of conic linear problems. Semi-Infinite Programming: Recent Advances, 135–165. Kluwer.

Yue, J., Chen, B., & Wang, M.-C. (2006). Expected value of distribution information for the newsvendor problem. Operations Research, 54(6), 1128–1136.

Zymler, S., Kuhn, D., & Rustem, B. (2013). Distributionally robust joint chance constraints with second-order moment information. Mathematical Programming, 137(1-2), 167–198.

Proof of the Main Results

Proof of Lemma 1. Our proof is motivated by Theorem 1 in Bertsimas & Sethuraman (2000), and effectively extends it to the case of censoring. Recall that $s = \min(d, k)$ denotes sales, so $s \in [0, k]$. The lemma requires that for every pair $(\mu_1, \mu_2) \in [\underline{m}_1, \overline{m}_1] \times [\underline{m}_2, \overline{m}_2]$, there exists a valid probability distribution of s with $\mathbb{E}[s] = \mu_1$ and $\mathbb{E}[s^2] = \mu_2$.

Necessary conditions for a fixed pair (μ_1, μ_2). For any random variable s, we must have $\mu_2 - \mu_1^2 = \mathrm{Var}(s) \geq 0$, i.e., $\mu_2 \geq \mu_1^2$, equivalently the moment matrix $\mathcal{M}(\mu_1, \mu_2) := \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \succeq 0$. Moreover, since $0 \leq s \leq k$, we have $s^2 \leq ks$ almost surely, which implies $\mu_2 = \mathbb{E}[s^2] \leq k\mathbb{E}[s] = k\mu_1$.

Convert to conditions that hold for all pairs in the box. Requiring $\mu_2 \geq \mu_1^2$ for all (μ_1, μ_2) in the rectangle is equivalent to checking the worst case $\mu_1 = \overline{m}_1$ and $\mu_2 = \underline{m}_2$, hence we must have $\underline{m}_2 \geq (\overline{m}_1)^2$. Similarly, requiring $\mu_2 \leq k\mu_1$ for all (μ_1, μ_2) in the rectangle is equivalent to checking the worst case $\mu_2 = \overline{m}_2$ and $\mu_1 = \underline{m}_1$, hence we must have $\overline{m}_2 \leq k\,\underline{m}_1$.

Sufficiency via explicit construction. Assume the box-level conditions $\underline{m}_2 \geq (\overline{m}_1)^2$ and $\overline{m}_2 \leq k\,\underline{m}_1$ hold. Take any $(\mu_1, \mu_2) \in [\underline{m}_1, \overline{m}_1] \times [\underline{m}_2, \overline{m}_2]$. Then automatically $\mu_2 \geq \underline{m}_2 \geq (\overline{m}_1)^2 \geq \mu_1^2$ and $\mu_2 \leq \overline{m}_2 \leq k\,\underline{m}_1 \leq k\mu_1$, so the pair satisfies $\mu_2 \geq \mu_1^2$ and $\mu_2 \leq k\mu_1$.

We now construct a sales distribution with these moments. If $\mu_1 = 0$, then $\mu_2 \leq k\mu_1$ implies $\mu_2 = 0$, and we take $s \equiv 0$. If $\mu_1 > 0$, define $x_1 := 0$, $x_2 := \frac{\mu_2}{\mu_1} \leq k$, $p := 1 - \frac{\mu_1^2}{\mu_2} \in [0, 1]$. Consider the two-point distribution $\mathbb{P}(s = x_1) = p$ and $\mathbb{P}(s = x_2) = 1 - p$. A direct calculation gives $\mathbb{E}[s] = (1 - p)x_2 = \mu_1$ and $\mathbb{E}[s^2] = (1 - p)x_2^2 = \mu_2$. Thus, for every pair (μ_1, μ_2) in the box, a valid sales distribution exists. Finally, defining $d := s$ yields a valid demand distribution supported on $\mathbb{R}_+$ satisfying $s = \min(d, k)$, completing the proof. $\square$

Proof. Proof of Theorem 1. By strong duality (see Shapiro (2001)), we obtain the dual of the inner minimization problem, redefining $\underline{m}_i = m_i - \Delta_i$ and $\overline{m}_i = m_i + \Delta_i$, where $i = 1, 2$:

$$
\begin{aligned}
\max_{q \geq 0} \max_{\vec{y}, \vec{z}} \quad & m_2 y_2 + m_1 y_1 + \Delta_2 z_2 + \Delta_1 z_1 + y_0 \\
\text{subject to} \quad & x^2 y_2 + x y_1 + y_0 \leq r \min(x, q) - cq, && \forall x \in (0, k], \\
& k^2 y_2 + k y_1 + y_0 \leq r \min(x, q) - cq, && \forall x \in (k, \infty], \\
& z_i \leq y_i \leq -z_i, && i = 1, 2, \\
& z_i \leq 0, && i = 1, 2.
\end{aligned}
\tag{5}
$$

The second constraint is redundant because it is implicitly induced by the first constraint. Thus, we focus on analysis of the first constraint and the case $q \in [0, k]$. Define the left-hand side of the first constraint as $f(x) := x^2 y_2 + xy_1 + y_0$. Since the objective is linear in y_0, y_1, and y_2, we will analyze all cases when $f(x)$ is just bounded by $r \min(x, q) - cq$, for all $x \leq k$. By doing this, any ϵ increase/decrease of y_i, where $i = 0, 1, 2$, will violates the boundary conditions. To facilitate our proof, we provide a figure for each case in the analysis.

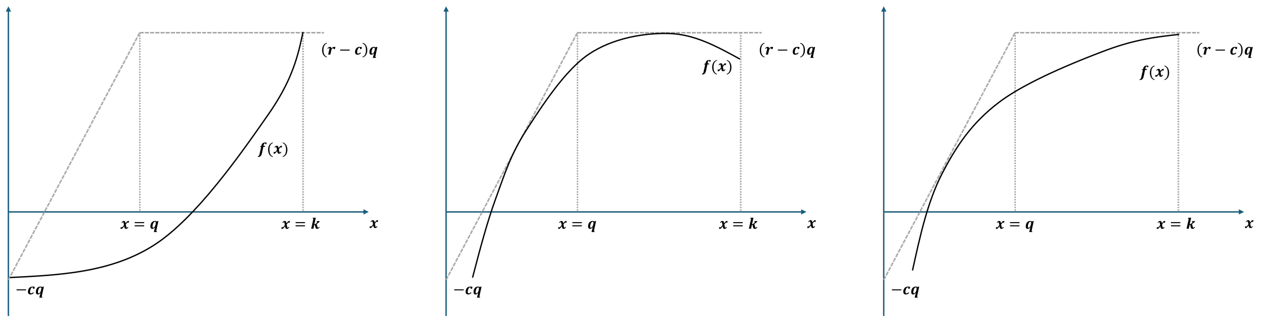


Figure 7: Proof of Theorem 1.

Case (I): $y_2 \geq 0$ and $f(x)$ crosses $(0, -cq)$ and $(k, (r-c)q)$ (see Figure7-1). In this case, we have: (1) $f(0) = -cq$ and (2) $f(k) = (r-c)q$. Accordingly, we obtain (1) $y_0 = -cq$ and (2) $y_1 = \frac{rq}{k} - ky_2$. Plug y_0 and y_1 into the objective function, we obtain $\pi(y_2, z_1, z_2, q) = (m_2 - km_1)y_2 + \left(m_1 \frac{r}{k} - c\right) q + \Delta_2 z_2 + \Delta_1 z_1$, with constraints $z_2 \leq y_2 \leq -z_2$, $z_1 \leq \frac{rq}{k} - ky_2 \leq -z_1$ and $z_i \leq 0$ where $i = 1, 2$. Recall $y_2 \geq 0$ and $z_i \leq 0$, we have (1) $z_2 \leq -y_2$, (2) $z_1 \leq \frac{rq}{k} - ky_2$ if $y_2 \geq \frac{rq}{k^2}$ and (3) $z_1 \leq -\frac{rq}{k} + ky_2$ if $y_2 \leq \frac{rq}{k^2}$. The objective function is increasing in z_1 and z_2.

(a) If $y_2 \geq \frac{rq}{k^2}$, we have $\pi(y_2, q) = (m_2 - km_1 - k\Delta_1 - \Delta_2)y_2 + \left(m_1 \frac{r}{k} + \Delta_1 \frac{r}{k} - c\right) q$. The objective function $\pi(y_2, q)$ is decreasing in y_2 (note that $m_2 - km_1 - k\Delta_1 - \Delta_2 \leq 0$), hence we have $y_2^* = \frac{rq}{k^2}$ to maximize the objective. Therefore, we obtain $\pi(q) = \left(\frac{r}{k^2}(m_2 - \Delta_2) - c\right) q$. To further maximize the objective via q, we have $q^* = 0$ and $\pi^* = 0$ if $\frac{c}{r} \geq \frac{m_2 - \Delta_2}{k^2}$; otherwise $q^* = k$ and $\pi^* = \left[\frac{r}{k^2}(m_2 - \Delta_2) - c\right] k$.

(b) If $y_2 \leq \frac{rq}{k^2}$, we have $\pi(y_2, q) = (m_2 - km_1 + k\Delta_1 - \Delta_2)y_2 + \left(m_1 \frac{r}{k} - \Delta_1 \frac{r}{k} - c\right) q$. To maximize the objective, we have $y_2^* = 0$ (note $m_2 - km_1 + k\Delta_1 - \Delta_2 = \underline{m}_2 - k\underline{m}_1 < 0$) and objective is $\pi(q) = \left(\frac{r}{k}(m_1 - \Delta_1) - c\right) q$. Similarly, we maximize the objective through q, we have $q^* = 0$ when $\frac{c}{r} \geq \frac{m_1 - \Delta_1}{k}$ and $\pi^* = 0$. While when $\frac{c}{r} < \frac{m_1 - \Delta_1}{k}$, $q^* = k$ and $\pi^* = \left(\frac{r}{k}(m_1 - \Delta_1) - c\right) k$. To summarize, the solution of this case is obtained by choosing the larger profit between case (a) and case (b). Hence, we have $q^* = k$ if $\frac{c}{r} \leq \frac{m_1 - \Delta_1}{k} = \frac{\underline{m}_1}{k}$ with profit is $r\underline{m}_1 - kc$; and $q^* = 0$ otherwise with profit equaling 0.

Case (II) $f(x)$ is tangent to $rx - cq$ and $(r-c)q$ (see Figure7-2). In this case, $y_2 < 0$ and $f(x)$ is tangent to $r\min(x, q) - cq$. We have: (1) $f(\frac{r - y_1}{2y_2}) = r\frac{r - y_1}{2y_2} - cq$ and (2) $f(-\frac{y_1}{2y_2}) = (r-c)q$. Thus, the two constraints give $y_1 = \frac{r}{2} - 2qy_2 > 0$ and $y_0 = \frac{(r - 4qy_2)^2}{16y_2} + (r-c)q$. Plug y_1 and y_0 into the objective function, we obtain $\pi(y_2, z_2, z_1, q) = [m_2 - m_1^2 + (m_1 - q)^2]y_2 + \frac{r^2}{16y_2} + \left(\frac{r}{2} - c\right) q + \frac{m_1 r}{2} + \Delta_2 z_2 + \Delta_1 z_1$.

Recall that $y_2 < 0$ and $z_i \leq 0$, so we have (1) $z_2 \leq y_2$ and (2) $z_1 \leq 2qy_2 - \frac{r}{2} < 0$. Note that the objective function $\pi(y_2, z_2, z_1, q)$ is increasing in z_1 and z_2, to maximize the objective, we have $z_1^* = 2qy_2 - \frac{r}{2}$ and $z_2^* = y_2$. Thus, we have $\pi^*(y_2, q) = [\nu + (q + \Delta_1 - m_1)^2]y_2 + \frac{r^2}{16y_2} + \left(\frac{r}{2} - c\right) q + \frac{(m_1 - \Delta_1)r}{2}$, where $\nu = \overline{m}_2 - \underline{m}_1^2$. Taking the first derivative, we have $\frac{d\pi^*(y_2,q)}{dy_2} = \nu + (q + \Delta_1 - m_1)^2 - \frac{r^2}{16y_2^2}$; and it is concave in y_2. Therefore, by first order condition, we obtain $y_2^* = -\frac{r}{4\sqrt{\nu+(q+\Delta_1-m_1)^2}}$. Accordingly, we have $\pi(q) = -\frac{r}{2}\sqrt{\nu + (q + \Delta_1 - m_1)^2} + \left(\frac{r}{2} - c\right) q + \frac{(m_1-\Delta_1)r}{2}$; and the outer layer problem provides $q^* = \underline{m}_1 + \frac{\sqrt{\nu}}{2}\left(\sqrt{\frac{r-c}{c}} - \sqrt{\frac{c}{r-c}}\right)$, where $\underline{m}_1 = m_1 - \Delta_1$. Finally, by checking the two boundary conditions: (1) $\frac{r-y_1}{2y_2} \geq 0$ and (2) $k \geq \frac{-y_1}{2y_2}$, we obtain $\frac{\underline{m}_1}{\sqrt{\nu}} \geq \sqrt{\frac{c}{r-c}}$ and $k \geq \underline{m}_1 + \sqrt{\nu}\sqrt{\frac{r-c}{c}}$.

Case (III): $f(x)$ **crosses** $(k,(r-c)q)$ **and is tangent to** $rx-cq$ **(see Figure7-3).** In this case, we have: (1) $f(\frac{r-y_1}{2y_2}) = r\frac{r-y_1}{2y_2} - cq$ and (2) $f(k) = (r-c)q$. Hence, we obtain $y_1 = r - 2ky_2 - 2\sqrt{ry_2(q-k)} > 0$ and $y_0 = (r-c)q - k(r-2ky_2) + 2k\sqrt{ry_2(q-k)} - k^2y_2$. Plug y_0 and y_1 into the objective, we obtain $\pi(y_2, z_2, z_1, q) = [m_2 - m_1^2 + (k-m_1)^2]y_2 + 2(k-m_1)\sqrt{ry_2(q-k)} + \Delta_1 z_1 + \Delta_2 z_2 + (m_1-k)r + (r-c)q$, with constraints: (1) $z_2 \leq y_2 \leq -z_2$ and $z_1 \leq y_1 \leq -z_1$. Note that the objective function is increasing in z_2 and z_1, to maximize the objective, we have $z_1^* = 2ky_2 + 2\sqrt{ry_2(q-k)} - r$ and $z_2^* = y_2$. Hence, we have $\pi(y_2, q) = [\nu + (k - m_1 + \Delta_1)^2]y_2 + 2[k + \Delta_1 - m_1]\sqrt{ry_2(q-k)} - (k + \Delta_1 - m_1)r + (r-c)q$, where $\nu = \overline{m}_2 - \underline{m}_1^2$. The maximum is obtained at the first-order condition: $y_2^* = -r(k-q)\left(\frac{m_1-\Delta_1-k}{\nu+(k-m_1+\Delta_1)^2}\right)^2$. Now, plug y_2^* into the objective function, we obtain $\pi(q) = r(k-q)\frac{(m_1-\Delta_1-k)^2}{\nu+(k-m_1+\Delta_1)^2} - (k + \Delta_1 - m_1)r + (r-c)q$. Finally, we maximize the objective through q. Taking first derivative of $\pi(q)$ regarding to q yields $\frac{d\pi(q)}{dq} = -r\frac{(m_1-\Delta_1-k)^2}{\nu+(k-m_1+\Delta_1)^2} + r - c = r\frac{\nu}{\nu+(k-m_1+\Delta_1)^2} - c$. As seen, (a) if $k \leq \underline{m}_1 + \sqrt{\frac{r-c}{c}\nu}$, the objective function is increasing in q so $q^* = k$; (b) otherwise, the objective function is decreasing in q so $q^* = 0$. Since the q^* are achieved at boundaries, so this case has no contribution to the final results.

Summary The optimal order quantity q^* is obtained from Case I and Case II. We compare the associated objectives (profits) in the Case I and Case II and select the maximum of two. □

Proof. Proof of Theorem 2. In this proof, we construct a demand/sale distribution such that the DRO order quantity characterized in Theorem 1 is obtained. Let $y = \min(d,k)$ represents the sales, which is demand that right truncated by censoring threshold k. The underlying distribution of the sales, y, is a two-point distribution with a mass p at point $a \geq 0$ and a mass $1-p$ at point $b \in (a,k]$. The two-point distribution satisfies the conditions: (1) $\mathbb{E}(y) = ap + b(1-p) = \underline{m}_1$,and (2) $\mathbb{E}(y^2) = a^2p + b^2(1-p) = \overline{m}_2$. Hence, we have $b = \underline{m}_1 + \frac{\overline{m}_2 - \underline{m}_1^2}{\underline{m}_1 - a} < k$ and $p = \frac{(\overline{m}_2 - \underline{m}_1^2)}{(\underline{m}_1-a)^2 + (\overline{m}_2 - \underline{m}_1^2)}$. That is to say, all the two-point distributions with a first moment $\underline{m}_1$ and a second moment $\overline{m}_2$ can be indexed by parameter $a \geq 0$ (if $b = \underline{m}_1 + \frac{\overline{m}_2 - \underline{m}_1^2}{\underline{m}_1 - a} < k$).

Case I. Suppose $a > 0$ and $b < k$, we have profit function $\Pi = r\min(d,q) - cq$. The profit function is increasing in q if $q \leq a$ and is decreasing in q if $q \geq b$. Therefore, the optimal quantity q maximizing the profit should be chose in $(a,b]$. Now, let $q \in (a,b]$ and plug b and p back, we obtain: $\Pi(a,q) = r(a-q)\frac{\overline{m}_2 - \underline{m}_1^2}{(\underline{m}_1-a)^2 + (\overline{m}_2 - \underline{m}_1^2)} + (r-c)q$. Our objective is $\max_q \min_a \Pi(a,q)$. Given q, we first find a^*. Taking first derivative of the profit function regarding to a provides first-order condition $\frac{\partial \Pi(a,q)}{\partial a} = r(\overline{m}_2 - \underline{m}_1^2)\frac{(\underline{m}_1-q)^2 + (\overline{m}_2 - \underline{m}_1^2) - (a-q)^2}{\left[(\underline{m}_1-a)^2 + (\overline{m}_2 - \underline{m}_1^2)\right]^2} = 0$. As seen, the profit function is convex in a, so a^* satisfies the first-order condition $a^* = q - \sqrt{(\underline{m}_1 - q)^2 + (\overline{m}_2 - \underline{m}_1^2)}$. Next, plug a^* back to the profit, we have $\Pi(q) = \frac{r}{2}\left\{\underline{m}_1 + q - \sqrt{(\underline{m}_1 - q)^2 + (\overline{m}_2 - \underline{m}_1^2)}\right\} - cq$. Taking first derivative of the profit function regarding to q yields: $\frac{d\Pi(q)}{dq} = \frac{r}{2}\left\{1 + \frac{(\underline{m}_1 - q)}{\sqrt{(\underline{m}_1-q)^2 + (\overline{m}_2 - \underline{m}_1^2)}}\right\} - c = 0$, which provides

$q^* = \underline{m}_1 + \frac{\sqrt{\nu}}{2}\left(\sqrt{\frac{r-c}{c}} + \sqrt{\frac{c}{r-c}}\right)$, where $\nu = \overline{m}_2 - \underline{m}_1^2$. Plug q^* back, we have $a = \underline{m}_1 - \sqrt{\nu}\sqrt{\frac{c}{r-c}}$, $b = \underline{m}_1 + \sqrt{\nu}\sqrt{\frac{r-c}{c}}$. As can be seen, the results restore the non-trivial case in theorem 1.

Case II. In this case, the first-order condition in case I cannot be satisfied. That is, for a given q, the parameter a minimizing the profit function is negative: $a^* = q - \sqrt{(\underline{m}_1 - q)^2 + (\overline{m}_2 - \underline{m}_1^2)} < 0$. This condition provides $q \leq \frac{\overline{m}_2}{2\underline{m}_1}$. Therefore, we have $a = 0$ and $b < k$. Hence, we have $b = \frac{\overline{m}_2}{\underline{m}_1}$ and $p = 1 - \frac{\underline{m}_1^2}{\overline{m}_2}$.

Case III. In this case, the first-order condition in case I cannot be satisfied. That is, for a given q, the parameter a minimizing the profit function is $a^* = q - \sqrt{(\underline{m}_1 - q)^2 + (\overline{m}_2 - \underline{m}_1^2)}$ and the corresponding parameter b^* is $b^* = \underline{m}_1 + \frac{\overline{m}_2 - \underline{m}_1^2}{\underline{m}_1 - a^*} > k$, which means $q \geq \frac{k^2 - \overline{m}_2}{2(k - \underline{m}_1)}$. Accordingly, we have $a = \frac{\underline{m}_1 k - \overline{m}_2}{k - \underline{m}_1}$ and $p = \frac{(k-\underline{m}_1)^2}{(k-\underline{m}_1)^2 + (\overline{m}_2 - \underline{m}_1^2)}$. □

Proof. Proof of Lemma 2. We prove the first statement of the lemma: by definition, we have $p = \int_0^k dF(x)$ and $1 - \frac{\mu_1}{k} = 1 - \int_0^\infty \frac{\min(x,k)}{k} dF(x)$. Since $1 - \int_0^\infty \frac{\min(x,k)}{k} dF(x) - \int_0^k dF(x) = -\int_0^k \frac{x}{k} dF(x) \leq 0$, so we have $1 - \frac{\mu_1}{k} \leq p$. Next, we prove the necessary and sufficient condition in the second statement of the lemma. If $1 - \frac{\mu_1}{k} \leq p$, $\forall (p, \mu_1) \in [\underline{m}_0, \overline{m}_0] \times [\underline{m}_1, \overline{m}_1]$, choosing $\mu_1 = \underline{m}_1$ and $p = \underline{m}_0$, we obtain $1 - \frac{\underline{m}_1}{k} \leq \underline{m}_0$, thus proving necessity. Next, supposing we have $1 - \frac{\underline{m}_1}{k} \leq \underline{m}_0$, the set inclusions $p \in [\underline{m}_0, \overline{m}_0]$ and $\mu_1 \in [\underline{m}_1, \overline{m}_1]$ imply that $1 - \frac{\mu_1}{k} \leq 1 - \frac{\underline{m}_1}{k} \leq \underline{m}_0 \leq p$, $\forall (p, \mu_1) \in [\underline{m}_0, \overline{m}_0] \times [\underline{m}_1, \overline{m}_1]$, thus showing sufficiency.

Finally, to make the finite-support construction explicit, let $\gamma := 1 - p = \mathbb{P}(s = k)$ denote the mass at the censoring point. For any feasible triple (p, μ_1, μ_2), define the conditional moments below k by

$$\tilde{\mu}_1 = \frac{\mu_1 - \gamma k}{1 - \gamma}, \qquad \tilde{\mu}_2 = \frac{\mu_2 - \gamma k^2}{1 - \gamma}.$$

These are the first two moments of the subdistribution of s on $[0, k)$. If $\tilde{\mu}_1 = 0$, then this subdistribution is degenerate at 0. Otherwise, feasibility implies that $(\tilde{\mu}_1, \tilde{\mu}_2)$ satisfies Lemma 1, so we may take a two-point subdistribution on $\{0, \tilde{\mu}_2/\tilde{\mu}_1\}$ with conditional mass $1 - \tilde{\mu}_1^2/\tilde{\mu}_2$ at 0. Equivalently, one explicit three-point sales distribution is obtained by placing mass

$$\rho = (1-\gamma)\Big(1 - \frac{\tilde{\mu}_1^2}{\tilde{\mu}_2}\Big), \qquad 1 - \gamma - \rho = (1-\gamma)\frac{\tilde{\mu}_1^2}{\tilde{\mu}_2}, \qquad \gamma$$

at the points $x_1 = 0$, $x_2 = \frac{\tilde{\mu}_2}{\tilde{\mu}_1}$,and k, respectively, with the obvious degenerate interpretation when $\tilde{\mu}_1 = 0$. □

Proof. Proof of Theorem 3. By strong duality, we have dual problem as follows:

$$\begin{aligned}
\max_{\vec{y}, \vec{z}} \quad & m_2 y_2 + m_1 y_1 + m_0 y_0 + \Delta_2 z_2 + \Delta_1 z_1 + \Delta_0 z_0 + y & \\
\text{subject to} \quad & x^2 y_2 + x y_1 + y_0 + y \leq r \min(x, q) - cq, & \forall x \in (0, k], \\
& k^2 y_2 + k y_1 + y \leq r \min(x, q) - cq, & \forall x \in (k, \infty), \\
& z_i \leq y_i \leq -z_i, & i = 0, 1, 2, \\
& z_i \leq 0, & i = 0, 1, 2,
\end{aligned} \tag{6}$$

where $\underline{m}_i = m_i - \Delta_i$ and $\overline{m}_i = m_i + \Delta_i$, $\Delta_i > 0$ and $i = 0, 1, 2$.

Similar to the proof of Theorem 1, we shall focus on analysis round function $f(x) := x^2 y_2 + x y_1 + y_0 + y$. Figure 8 illustrates the analysis of each case.

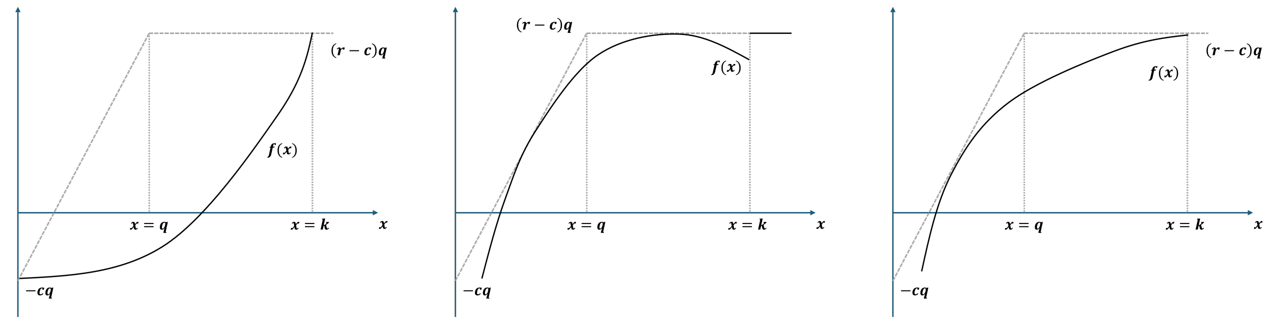


Figure 8: Proof of Theorem 3.

Case (I): When $y_2 \geq 0$ and $f(k^-) \leq (r-c)q$ (See Figure 8-1). This case corresponds to case (I) in the proof of Theorem 1. we have: (1) $f(0) = -cq$ and (2) $f(k^+) = (r-c)q$. Accordingly, the objective function becomes $\pi(y_2, y_0, z_0, z_1, z_2, q) = (m_2 - k m_1) y_2 + \left(m_1 \frac{r}{k} - c\right) q + \left(\frac{m_1}{k} + m_0 - 1\right) y_0 + \Delta_2 z_2 + \Delta_1 z_1 + \Delta_0 z_0$, with constraints $z_2 \leq y_2 \leq -z_2$, $z_1 \leq \frac{rq+y_0}{k} - k y_2 \leq -z_1$, $z_0 \leq y_0 \leq -z_0$ and $z_i \leq 0$ where $i = 0, 1, 2$. Simplifying the constraints, we have (1) $z_2 \leq -y_2$, (2) $z_1 \leq \frac{rq+y_0}{k} - k y_2$ if $y_2 \geq \frac{rq+y_0}{k^2}$, (3) $z_1 \leq -\frac{rq+y_0}{k} + k y_2$ if $y_2 \leq \frac{rq+y_0}{k^2}$ and (4) $z_0 \leq y_0$.

(a) When $y_2 \geq \frac{rq+y_0}{k^2}$, we have $\pi(y_2, y_0, q) = (\underline{m}_2 - k\overline{m}_1) y_2 + \left(\frac{\overline{m}_1}{k} - 1 + \overline{m}_0\right) y_0 + \left(\overline{m}_1 \frac{r}{k} - c\right) q$. The objective function is decreasing in y_2 and is increasing in y_0 (because $\frac{\overline{m}_1}{k} - 1 + \overline{m}_0 \geq 0$). To maximize the objective, we have $y_2^* = \frac{rq}{k^2}$ and $y_0^* = 0$ (by plugging $y_2^* = \frac{rq+y_0}{k^2}$ back to the objective, the objective is increasing in y_0 as $\frac{m_2}{k^2} - 1 + \overline{m}_0 \geq 0$). From here, our case restores the case when $y_2 \geq \frac{rq}{k^2}$ in the proof of theorem 2 Case I.

(b) If $y_2 \leq \frac{rq+y_0}{k^2}$, we have $\pi(y_2, q) = (\underline{m}_2 - k\overline{m}_1) y_2 + \left(\frac{m_1}{k} - 1 + \overline{m}_0\right) y_0 + \left(\underline{m}_1 \frac{r}{k} - c\right) q$. To maximize the objective, we have $y_2^* = 0$ and $y_0^* = 0$; and thus $\pi(q) = \left(\underline{m}_1 \frac{r}{k} - c\right) q$. From here, this case restores the case when $y_2 \leq \frac{rq}{k^2}$ in the proof of theorem 2 Case I. Therefore, the optimal q should be chosen from either 0 or k.

Case (II) When $y_2 < 0$ and $f(k^-) < (r-c)q$ (See Figure 8-2). In this case, $y_2 < 0$ and $f(x)$ is tangent to $r\min(x, q) - cq$. We have: (1) $f(\frac{r-y_1}{2y_2}) = r\frac{r-y_1}{2y_2} - cq$. (2) $f(-\frac{y_1}{2y_2}) = (r-c)q$. (3) $f(k^+) = (r-c)q$. Plug y_1, y_0 and y in to the objective function, we obtain $\pi = \left[m_2 - 2m_1 q - (k-q)^2 + q^2 + m_0(k-q)^2\right] y_2 + \frac{(m_1-k)r}{2} + m_0\left(\frac{r(k-q)}{2} + \frac{r^2}{16y_2}\right) + (r-c)q + \sum_{i=0}^{2} \Delta_i z_i$, with constraints $z_2 \leq y_2 \leq -z_2$, $z_1 \leq \frac{r}{2} - 2q y_2 \leq -z_1$, $z_0 \leq y_0 \leq -z_0$ and $z_i \leq 0$. Note that the objective function is increasing in z_i, to maximize the objective, we have $z_2^* = y_2$, $z_1^* = 2q y_2 - \frac{r}{2}$ and $z_0^* = y_0$. Plug z_1^* and z_2^* back to the objective, we have $\pi^*(y_2, q) = \left[\nu + (\underline{m}_1 - q)^2 - (1-\overline{m}_0)(k-q)^2\right] y_2 + \overline{m}_0 \frac{r^2}{16y_2} + (r-c)q + (\underline{m}_1 - k)\frac{r}{2} + \overline{m}_0 \frac{r(k-q)}{2}$, where $\nu = \overline{m}_2 - \underline{m}_1^2$.

To maximize the objective, we take first derivative with regarding to y_2 and yield the first order condition (the objective is concave in y_2): $y_2^* = -\frac{r}{4}\sqrt{\frac{\overline{m}_0}{\nu + (\underline{m}_1 - q)^2 - (1-\overline{m}_0)(k-q)^2}}$. Next, we plug y_2^* into the objective and maximize $\pi(q)$ via q. By taking first derivative, we have the first order condition: $q^* = \frac{\underline{m}_1 - k(1-\overline{m}_0)}{\overline{m}_0} + \frac{\sqrt{\overline{m}_0 \nu - (1-\overline{m}_0)(k-\underline{m}_1)^2}}{2\overline{m}_0}\left(\sqrt{\frac{r-c}{c-(1-\overline{m}_0)r}} - \sqrt{\frac{c-(1-\overline{m}_0)r}{r-c}}\right)$. Finally, we check the two boundary conditions: (1) $\frac{r-y_1}{2y_2} \geq 0$ and (2) $k \geq \frac{-y_1}{2y_2}$. We obtain $\frac{m_1}{\sqrt{\nu}} \geq \theta\sqrt{\frac{c}{r-c}}$, where $\theta = \sqrt{1 - \frac{1-\overline{m}_0}{c^2-(1-\overline{m}_0)rc}\frac{(ck-r\underline{m}_1)^2}{\nu}}$ and $k \geq \underline{m}_1 + \sqrt{\frac{r-c}{c}\nu}$. Therefore, if the two boundary conditions are satisfied, we have

$$q^* = \frac{\underline{m}_1 - (1-\gamma)k}{\gamma} + \frac{\sqrt{\gamma\nu - (1-\gamma)(k-\underline{m}_1)^2}}{2\gamma}\left(\sqrt{\frac{r-c}{c-(1-\gamma)r}} - \sqrt{\frac{c-(1-\gamma)r}{r-c}}\right),$$
$$\pi^* = \frac{\sqrt{r-c}}{\gamma}\left\{\sqrt{r-c}[\underline{m}_1 - (1-\gamma)k] - \sqrt{\gamma\nu - (1-\gamma)(k-\underline{m}_1)^2}\sqrt{c-(1-\gamma)r}\right\},$$

where $\nu = m_2 + \Delta_2 - (m_1 - \Delta_1)^2$ and $\gamma = m_0 + \Delta_0$

Case III: $f(x)$ **passes through** $(x, f(x)) = (k, (r-c)q)$ **(see Figure 8-3).** This case restores the case III in the proof of theorem 2; it won't contribute to our final results.. □

Proof. Proof of Lemma 3. We start by writing the two non-trivial order quantities as: $q_S^* = \frac{\beta-\gamma k}{1-\gamma} + \frac{\sqrt{\nu(1-\gamma)-\gamma(k-\beta)^2}}{2(1-\gamma)}\left(\sqrt{\frac{r-c}{c-\gamma r}} - \sqrt{\frac{c-\gamma r}{r-c}}\right)$, and $q_C^* = \beta + \frac{\sqrt{\nu}}{2}\left(\sqrt{\frac{r-c}{c}} - \sqrt{\frac{c}{r-c}}\right)$, where $\nu = m_2 + \Delta_2 - (m_1 - \Delta_1)^2$, $\beta = m_1 - \Delta_1$ and $\gamma = m_0 + \Delta_0$.

We first show $\beta - \frac{\beta-\gamma k}{1-\gamma} > 0$. That is because $\beta - \frac{\beta-\gamma k}{1-\gamma} = \frac{1}{1-\gamma}[(1-\gamma)\beta - (\beta-\gamma k)] = \frac{\gamma}{1-\gamma}(k-\beta) > 0$. Next, we compare the second terms in the two order quantities. Note that we have $\frac{\sqrt{\nu(1-\gamma)-\gamma(k-\beta)^2}}{2(1-\gamma)}\left(\sqrt{\frac{r-c}{c-\gamma r}} - \sqrt{\frac{c-\gamma r}{r-c}}\right) \leq \frac{\sqrt{\nu}}{2(1-\gamma)}\sqrt{\frac{c-\gamma r}{c}}\left(\sqrt{\frac{r-c}{c-\gamma r}} - \sqrt{\frac{c-\gamma r}{r-c}}\right)$, because the non-trivial order quantity exists under condition $k \geq \beta + \sqrt{\frac{r-c}{c}\nu}$. Hence, we have

$$\frac{\sqrt{\nu}}{2}\left(\sqrt{\frac{r-c}{c}} - \sqrt{\frac{c}{r-c}}\right) - \frac{\sqrt{\nu(1-\gamma)-\gamma(k-\beta)^2}}{2(1-\gamma)}\left(\sqrt{\frac{r-c}{c-\gamma r}} - \sqrt{\frac{c-\gamma r}{r-c}}\right)$$
$$\geq \frac{\sqrt{\nu}}{2}\left(\sqrt{\frac{r-c}{c}} - \sqrt{\frac{c}{r-c}}\right) - \frac{\sqrt{\nu}}{2(1-\gamma)}\sqrt{\frac{c-\gamma r}{c}}\left(\sqrt{\frac{r-c}{c-\gamma r}} - \sqrt{\frac{c-\gamma r}{r-c}}\right) = -\frac{\gamma\sqrt{\nu}}{1-\gamma}\sqrt{\frac{r-c}{c}}.$$

Therefore, We have $q_C^* - q_S^* \geq \frac{\gamma}{1-\gamma}(k-\beta) - \frac{\gamma}{1-\gamma}\sqrt{\frac{\nu(r-c)}{c}}\frac{\gamma}{1-\gamma}\left(k - \beta - \sqrt{\frac{r-c}{c}\nu}\right) \geq 0$. □

Proof. Proof of Theorem 4 Similar to the proof of Theorem 2, we construct a sales distribution (a three-point distribution) that corresponds to the result characterized in Theorem 3. Let $y = \min(d, k)$ to represent the sales. The underlying distribution of the sales, y, can be a three-point distribution with a mass p at point $a \geq 0$, a mass $1-\gamma-p$ at point $b \in (a, k)$ and a mass γ at k, where $1-\gamma = \overline{m}_0$. Suppose $a > 0$ and given first moment $\underline{m}_1$ and second moment $\overline{m}_2$, we have $b = \frac{\underline{m}_1 - \gamma k - ap}{1-\gamma-p} < k$ and $p = \frac{(1-\gamma)(\overline{m}_2 - \underline{m}_1^2) - \gamma(k-\underline{m}_1)^2}{(\underline{m}_1-a)^2 - \gamma(k-a)^2 + (\overline{m}_2 - \underline{m}_1^2)}$.

Case I. Suppose $a > 0$ and $b < k$, we have profit function $\Pi = r\min(d,q) - cq$. The profit is increasing in q if $q \leq a$ and is decreasing in q if $q \geq k$. Therefore, the optimal quantity q maximizing the profit should be chose in $(a, b]$. Let $q \in (a, b]$ and plug b and p back to the profit function, we obtain: $\Pi(a,q) = r(a-q)\frac{(\overline{m}_2 - \underline{m}_1^2) - \frac{\gamma}{(1-\gamma)}(k-\underline{m}_1)^2}{\left(\frac{\underline{m}_1-\gamma k}{(1-\gamma)} - a\right)^2 - \frac{\gamma}{(1-\gamma)^2}(\underline{m}_1-k)^2 + \frac{\overline{m}_2 - \underline{m}_1^2}{1-\gamma}} + (r-c)q$. Our objective is $\max_q \min_a \Pi(a,q)$. For any given quantity q, taking first derivative of the profit function regarding to

a provides $\left(\frac{\underline{m}_1-\gamma k}{(1-\gamma)}-a\right)^2-\frac{\gamma}{(1-\gamma)^2}(\underline{m}_1-k)^2+\frac{\overline{m}_2-\underline{m}_1^2}{1-\gamma}-2(a-q)\left(\frac{\underline{m}_1-\gamma k}{(1-\gamma)}-a\right)=0$. By checking the second derivative, the profit function is convex in a and the optimal a minimizing the profit can be achieved by the first-order condition above. Next, plug a^* back to the profit, we have $\Pi(q)=-\frac{r}{2}(1-\gamma)\left\{\frac{\underline{m}_1-\gamma k}{(1-\gamma)}-q-\sqrt{\left(\frac{\underline{m}_1-\gamma k}{(1-\gamma)}-q\right)^2+\frac{1}{(1-\gamma)}\left(-\frac{\gamma}{1-\gamma}(\underline{m}_1-k)^2+\overline{m}_2-\underline{m}_1^2\right)}\right\}+(r-c)q$. The optimal q^* is set to maximize the above profit function: $q^*=\frac{\underline{m}_1-\gamma k}{1-\gamma}+\frac{\sqrt{\overline{\nu}(1-\gamma)-\gamma(k-\underline{m}_1)^2}}{2(1-\gamma)}\left(\sqrt{\frac{r-c}{c-r\gamma}}-\sqrt{\frac{c-\gamma r}{r-c}}\right)$. Plug q^* back, we have $a=\frac{\underline{m}_1-\gamma k}{1-\gamma}+\frac{\sqrt{\overline{\nu}(1-\gamma)-\gamma(k-\underline{m}_1)^2}}{(1-\gamma)}\sqrt{\frac{r-c}{c-r\gamma}}$.

Case II. In this case, the first-order condition in case I cannot be satisfied: the parameter a minimizing the profit function is negative $a^*<0$. In this case, we have $a=0$ and $b<k$.

Case III. In this case, the first-order condition in case I cannot be satisfied: the parameter b^* that minimizes the profit function greater than k: $b^*=\frac{\underline{m}_1-\gamma k-a^*p}{1-\gamma-p}>k$. □

Proof. Proof of Theorem 5. By strong duality (Shapiro, 2001), the inner minimization problem can be replaced by its dual, and we may write the model as (where $\ell=2$)

$$\begin{aligned} V(q)\triangleq\quad \max\quad & \frac{1}{2}\sum_{j=0}^{2}\sum_{i=1}^{M}\left[(\overline{m}_{i,j}+\underline{m}_{i,j})y_{i,j}+(\overline{m}_{i,j}-\underline{m}_{i,j})z_{i,j}\right]+y \\ \text{s.t.}\quad & \sum_{j=1}^{2}\sum_{i=w+1}^{M}y_{i,j}x^j+\sum_{j=1}^{2}\sum_{i=0}^{w}y_{i,j}k_i^j+\sum_{i=w+1}^{M}y_{i,0}+y\le r\min(x,q)-cq,\ x\in(k_w,k_{w+1}],\ w=0,\cdots,M, \\ & z_{i,j}\le y_{i,j}\le -z_{i,j},\quad i=1,\cdots,M,\ j=0,1,2, \\ & z_{i,j}\le 0,\quad i=1,\cdots,M,\ j=0,1,2 \end{aligned}\tag{7}$$

where $k_1<k_2<\cdots<k_M$ and we set $k_0=0$ and $k_{M+1}=\infty$. Formulation (7) has $2\times(M+1)$ constraints.

We next replicate the relevant portions of Proposition 3.1 in Bertsimas & Popescu (2005).

1. The polynomial $g(x)=\sum_{n=0}^{\ell}y_nx^n$ satisfies $g(x)\ge 0$ for all $x\in[0,a]$ if and only if there exists a positive semidefinite matrix $X=[x_{i,j}]_{i,j=0,\dots,\ell}$, such that

$$\begin{aligned} 0=&\sum_{i+j=2n-1}x_{i,j}, && n=1,\dots,\ell; \\ \sum_{m=0}^{n}y_m\binom{\ell-m}{n-m}a^m=&\sum_{i,j:i+j=2n}x_{i,j}, && n=0,\dots,\ell; \\ X\succeq&0. \end{aligned}$$

2. The polynomial $g(x)=\sum_{n=0}^{\ell}y_nx^n$ satisfies $g(x)\ge 0$ for all $x\in[a,b]$ if and only if there exists

a positive semidefinite matrix $X = [x_{i,j}]_{i,j=0,\ldots,\ell}$, such that

$$\begin{aligned} 0 = & \sum_{i+j=2n-1} x_{i,j}, & n = 1,\ldots,\ell; \\ \sum_{m=0}^{n}\sum_{t=m}^{\ell+m-n} y_t \binom{t}{m}\binom{\ell-t}{n-m} a^{t-m} b^m = & \sum_{i+j=2n} x_{i,j}, & n = 0,\ldots,\ell; \\ X \succeq & 0. & \end{aligned}$$

By applying parts (c) and (f) of the above proposition, the formulation in the theorem can be obtained.

Finally, we show that the function $V(q|q \in (k_J, k_{J+1}))$ is concave in q. We utilize the original formulation of $V(q)$ (here, we drop condition $q \in (k_J, k_{J+1})$ in the expression for the ease of presentation). Suppose that we have $\bar{q} = \lambda q_1 + (1-\lambda) q_2$, where $\lambda \in (0,1)$ and $q_1 < q_2$ ($q_i \in (k_J, k_{J+1})$, for $i = 1, 2$). Now, define $y^*_{i,j}(q)$, $z^*_{i,j}(q)$, and $y^*(q)$ as the solution to $V(q)$. In other words, we have $V(q_1) = \frac{1}{2}\sum_{j=0}^{2}\sum_{i=1}^{M}\left[(\overline{m}_{i,j} + \underline{m}_{i,j}) y^*_{i,j}(q_1) + (\overline{m}_{i,j} - \underline{m}_{i,j}) z^*_{i,j}(q_1)\right] + y(q_1)$, where $\sum_{j=1}^{2}\sum_{i=w+1}^{M} y^*_{i,j}(q_1) x^j + \sum_{j=1}^{2}\sum_{i=0}^{w} y^*_{i,j}(q_1) k_i^j + \sum_{i=w+1}^{M} y_{i,0} + y^*(q_1) \le r\min(x, q_1) - cq_1$, $x \in (k_J, k_{J+1}]$, $y^*_{i,j}(q_1) \in [z^*_{i,j}(q_1), -z^*_{i,j}(q_1)]$ and $z^*_{i,j}(q_1) \le 0$.

Similarly, we have $V(q_2) = \frac{1}{2}\sum_{j=0}^{2}\sum_{i=1}^{M}\left[(\overline{m}_{i,j} + \underline{m}_{i,j}) y^*_{i,j}(q_2) + (\overline{m}_{i,j} - \underline{m}_{i,j}) z^*_{i,j}(q_2)\right] + y(q_2)$, where $\sum_{j=1}^{2}\sum_{i=w+1}^{M} y^*_{i,j}(q_2) x^j + \sum_{j=1}^{2}\sum_{i=0}^{w} y^*_{i,j}(q_2) k_i^j + \sum_{i=w+1}^{M} y_{i,0} + y^*(q_2) \le r\min(x, q_2) - cq_2$, $x \in (k_J, k_{J+1}]$, $y^*_{i,j}(q_2) \in [z^*_{i,j}(q_2), -z^*_{i,j}(q_2)]$ and $z^*_{i,j}(q_2) \le 0$.

Now, we take a convex combination of the left-hand sides of the two constraints and obtain

$$\begin{aligned} \lambda LHS(q_1) + (1-\lambda) LHS(q_2) & \le \lambda[\min r(x, q_1) - cq_1] + (1-\lambda)[r\min(x, q_2) - cq_2] \\ & \le r\min(x, \bar{q}) - c\bar{q}, \end{aligned}$$

where $LHS(q) := \displaystyle\sum_{j=1}^{2}\sum_{i=w+1}^{M} y^*_{i,j}(q) x^j + \sum_{j=1}^{2}\sum_{i=0}^{w} y^*_{i,j}(q) k_i^j + \sum_{i=w+1}^{M} y_{i,0} + y^*(q)$. The last inequality is due to the concavity of the function $f(q) = \min(x, q)$. In other words, $(\lambda y^*_{i,j}(q_1) + (1-\lambda) y^*_{i,j}(q_2), \lambda z^*_{i,j}(q_1) + (1-\lambda) z^*_{i,j}(q_2), \lambda y^*(q_1) + (1-\lambda) y^*(q_2))$ is a feasible solution for $V(\bar{q})$, which implies

$$\begin{aligned} V(\bar{q}) & \ge \lambda\left(\frac{1}{2}\sum_{j=0}^{2}\sum_{i=1}^{M}\left[(\overline{m}_{i,j} + \underline{m}_{i,j}) y^*_{i,j}(q_1) + (\overline{m}_{i,j} - \underline{m}_{i,j}) z^*_{i,j}(q_1)\right] + y(q_1)\right) \\ & \quad + (1-\lambda)\left(\frac{1}{2}\sum_{j=0}^{2}\sum_{i=1}^{M}\left[(\overline{m}_{i,j} + \underline{m}_{i,j}) y^*_{i,j}(q_2) + (\overline{m}_{i,j} - \underline{m}_{i,j}) z^*_{i,j}(q_2)\right] + y(q_2)\right) \\ & = \lambda V(q_1) + (1-\lambda) V(q_2). \end{aligned}$$

Therefore, we conclude that $V(q)$ is concave in q by definition.

□