MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
https://pubsonline.informs.org/journal/msom
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 1523-4614, EISSN 1526-5498

Submitted to *Manufacturing & Service Operations Management*

The Data-Driven Censored Newsvendor Problem

(Authors' names are not included for peer review)

Abstract. Problem definition: We study a censored variant of the data-driven newsvendor problem, where the decision-maker must select an ordering quantity that minimizes expected overage and underage costs based only on offline *censored* sales data, rather than historical demand realizations. Our goal is to understand how the degree of historical censoring affects the performance of any learning algorithm for this problem. To isolate this impact, we adopt a distributionally robust optimization framework, evaluating policies according to their worst-case regret over an *ambiguity set* of distributions. This set is defined by the largest historical order quantity (the *observable boundary* of the dataset), and contains all distributions matching the true demand distribution up to this boundary, while allowing them to be arbitrary afterwards. **Methodology/results:** We demonstrate a *spectrum of achievability* under demand censoring by deriving a natural necessary and sufficient condition under which vanishing regret is an achievable goal. In regimes in which it is not, we characterize the information loss due to censoring: an insurmountable lower bound on the performance of any policy, *even when the decision-maker has access to infinitely many demand samples*. We then leverage these sharp characterizations to propose a natural *robust* algorithm that adapts to the historical level of demand censoring. We derive finite-sample guarantees for this algorithm across all censoring regimes and show its near-optimality with matching lower bounds (up to polylogarithmic factors). We moreover demonstrate its robust performance via extensive numerical experiments on both synthetic and real-world datasets. **Managerial implications:** Demand censoring is pervasive in practice. Our work provides decision-makers with simple, adaptive policies that are robust to environments that exhibit varying levels of historical demand censoring.

Key words: Newsvendor, censored data, data-driven decision-making, distributionally robust optimization, minimax regret, finite-sample guarantees

1. Introduction

The classical newsvendor problem models settings where a decision-maker makes an inventory decision to fulfill future random demand, with the goal of minimizing expected overage and underage costs once demand is realized. In the most basic version of this problem — wherein the decision-maker makes a single decision and knows the distribution from which demand is drawn — the optimal solution is well-known to be the critical quantile, an ordering quantity that balances the likelihood of overage and underage according to their relative costs (Zipkin 2000). In more realistic settings, however, the decision-maker does not know the true demand distribution; instead, she has access to data (e.g., historical demand) that allows her to learn good ordering policies despite this information gap. This is the so-called *data-driven* newsvendor problem.

In the offline setting, when the decision-maker only needs to make a *single* ordering decision given historical demand realizations, the Sample Average Approximation (SAA) heuristic, which computes the sample critical quantile, has been shown to be near-optimal, with extensive efforts devoted to deriving tight performance guarantees for this natural policy (Chen and Ma 2024). This line of work, however, assumes that demand is *fully observable*, an assumption that fails to hold in many real-world applications, such as brick-and-mortar retail settings. In these settings, more often than not, demand is *censored*. That is, if the decision-maker only had a fixed amount of inventory to give out over a period of time, she would only observe how many units were *sold*, as opposed to how many were actually *desired*. Learning in these settings, then, is likely to be more challenging, especially in highly resource-constrained settings where historical demand consistently dwarfs historical inventory levels (e.g., demand for monocultural events such as the Eras Tour, or for hotel rooms associated with conference venues). When this occurs, the decision-maker has likely *never* observed the true demand realizations.

We conjecture that this challenge is why the offline, data-driven censored newsvendor problem has been explored to a much lesser extent. Indeed, to the best of our knowledge, the design of algorithms with finite-sample guarantees for this problem has been studied in only two fairly recent works (Ban 2020, Fan et al. 2022), both of which rely on the crucial assumption that a constant fraction of historical inventory levels exceed the critical quantile (henceforth referred to as the *optimal newsvendor quantity*). Upon first reflection, the necessity of such an assumption is evident: if all historical inventory levels were zero, no demand would ever be observed. In this case, what can the decision-maker possibly hope to accomplish? Still, this assumption remains inherently unsatisfying, for two reasons. The first reason is practical: there exist many highly constrained real-world settings where there is no reason to expect that historical inventory levels were in fact high enough to capture the optimal newsvendor quantity, as in the aforementioned examples. The second reason is more fundamental; in particular, it is not clear that this assumption is even testable, given that it requires knowledge of the true newsvendor quantity. Faced with this fact, the decision-maker has no option but to deploy the proposed algorithms and hope for the best (a solution we show fails dramatically when the assumption does not hold). Our work is motivated by this important gap in the literature. Namely, we seek to answer the following research questions:

What is the impact of demand censoring on algorithm performance in the offline data-driven newsvendor problem? Can we characterize a lower bound on algorithm performance as a function of the level of historical demand censoring? Do there exist simple algorithms that asymptotically achieve this lower bound?

1.1. Main Contributions

A flexible modeling framework. In an attempt to tackle these questions, one of the main contributions of this work is a modeling framework that allows us to formalize the impact of demand censoring on algorithm performance for the newsvendor problem. In particular, while a standard performance metric for the data-driven newsvendor problem is the additive optimality gap relative to the optimal newsvendor cost under complete information, we argue that this benchmark ceases to be meaningful in highly censored settings. This is most obvious for the pathological instance where all historical inventory levels are zero, in which case there is clearly no hope for the decision-maker to learn the optimal newsvendor quantity. Thus motivated, we turn to the distributionally robust optimization (DRO) framework. Specifically, we consider the *ambiguity set* of a given instance to be the set of all possible distributions sharing the same cumulative distribution function (cdf) as the true (unknown) demand distribution, up until the maximum historical inventory level, referred to as the *observable boundary* of the dataset and denoted by λ. After λ, distributions in the ambiguity set may take an arbitrary shape (Definition 1). Given the ambiguity set, we evaluate the performance of any data-driven policy according to its worst-case additive optimality gap compared to the optimal newsvendor quantity (i.e., its worst-case regret), *over all possible distributions in the ambiguity set* (Definition 2). (Such an ambiguity set was first defined in Bu et al. (2023) for the problem of offline pricing under censored demand.)

Introducing this modeling framework for the data-driven censored newsvendor problem directly addresses the benchmark issue described above. When λ is low, much of the demand will never be observed. In such information-poor settings, we posit that it is more appropriate for a decision-maker to instead view nature as adversarial. This view is reflected by a large ambiguity set of distributions against which she needs to compete. In historically unconstrained settings, on the other hand, when λ is large, existing work hints at the idea that the decision-maker's objective is fundamentally easier (Ban 2020, Fan et al. 2022). This other extreme is reflected in a much smaller ambiguity set of distributions against which to compete.

Impact of censoring on achievable algorithm performance. This flexible framework allows us to formalize the existence of a *spectrum of achievability* determined by the observable boundary of the dataset. Our first main technical contribution is to exactly characterize this spectrum (Theorem 1). In particular, we provide a *necessary and sufficient* condition for any algorithm to achieve vanishing regret in the worst case. This condition turns out to be equivalent to the condition assumed in Ban (2020) and Fan et al. (2022): the observable boundary λ must exceed the optimal newsvendor quantity. We moreover recover that in this *identifiable* regime, the minimax optimal ordering quantity is exactly equal to the optimal newsvendor quantity corresponding to the true demand distribution. This result formalizes that the optimality of the critical quantile is robust to a wide range of censoring levels.

In the *unidentifiable* setting where this condition fails to hold, however, our result implies that the regret of any algorithm is bounded away from zero, even if the decision-maker has access to infinitely many samples. We show this by providing an exact lower bound on the worst-case regret in such settings, which we refer to as the *minimax risk* Δ and which can be interpreted as the information loss due to demand censoring. The minimax risk exhibits the natural property of being decreasing in λ, reflecting the intuition that the decision-maker's problem becomes easier as the data becomes less censored. We moreover produce an explicit expression for the minimax optimal ordering quantity, which similarly depends on λ, and at a high level hedges against a worst-case distribution that places weight exclusively on λ and a known upper bound on the optimal newsvendor quantity.

A robust, near-optimal algorithm. These characterizations prove to be pivotal in demonstrating that a simple algorithm asymptotically achieves the fundamental lower bound due to demand censoring, as the number of samples grows large. Our algorithm, Robust Censored Newsvendor (RCN, Algorithm 1), proceeds in two stages. In the first stage, it tests for identifiability by estimating the fraction of demand that lies below λ. If this estimate exceeds the critical ratio by some appropriately tuned confidence parameter, the algorithm outputs an empirical estimate of the optimal newsvendor quantity. If it falls short of the critical ratio by some confidence term, on the other hand, it classifies the problem as unidentifiable and computes an empirical estimate of the minimax optimal ordering quantity. In between, it is unable to conclusively determine identifiability, and defaults to outputting λ.

This practical algorithm lies on the spectrum of achievability characterized in Theorem 1. Namely, across *all* regimes of identifiability, it guarantees $O(1/\sqrt{N})$ worst-case regret with respect to the minimax risk with probability $1-\delta$, where N is the number of samples associated with ordering quantity λ (henceforth referred to as samples at the boundary), and δ determines the algorithm's confidence level for identifiability (Theorem 2). While seemingly intuitive, the above description sweeps under the rug that our algorithm (necessarily) only uses censored data to compute all of its estimates. Hence, it is a priori unclear that any of these would in fact be unbiased, a prerequisite to our algorithm obtaining its strong guarantees. Herein lies a crucial design choice: our algorithm *only* uses samples at the boundary. This choice is quite subtle: while censored data in general will produce biased estimates, by subsetting amongst samples at the boundary, we are able to recover unbiasedness of all empirical estimates.

We subsequently complement these upper bounds with matching lower bounds (up to polylogarithmic factors) for all possible censoring regimes (Theorem 3). We do so via a unified treatment for all possible instances, reducing the decision-making problem to that of hypothesis testing.

Finally, we demonstrate the strong practical performance of our algorithm via extensive computational experiments, on both synthetic and real-world datasets. We specifically explore the dependence of our algorithm's numerical performance on (i) the number of samples N, (ii) the observable boundary λ, and (iii) the variability of the underlying demand distribution. We observe that our algorithm asymptotically achieves the minimax risk Δ across all censoring regimes, as compared to state-of-the-art benchmarks that assume identifiability (Ban 2020, Fan et al. 2022). We moreover propose a practical modification to RCN (RCN$^+$, Algorithm 2) which leverages the entirety of the dataset, as opposed to only samples at the boundary; we observe that this modification can yield significant performance improvements in the identifiable regime.

Paper organization. We review the related literature in the remainder of this section. We formally present our model in Section 2, and formalize the minimax risk of the censored newsvendor in Section 3. In Section 4 we design and analyze a best-of-both-worlds algorithm which asymptotically achieves the minimax risk. We complement this with a matching lower bound in Section 5. We demonstrate our algorithm's strong numerical performance in Section 6, and conclude in Section 7.

1.2. Related Literature

The newsvendor model is one of the most foundational models in the operations literature, beginning with the work of Scarf (1958). In the past few decades, there has been significant focus on incorporating data-driven methods to solve variants of this canonical model, when the decision-maker either has access to *offline* data or can collect data adaptively (i.e., *online*). In this section we discuss the most closely related literature to the offline censored newsvendor problem we consider. We refer readers to Chen and Ma (2024) for a comprehensive survey of data-driven newsvendor results.

Data-Driven Newsvendor: Non-Parametric Setting. Our work falls within the large body of literature on the data-driven newsvendor in offline, non-parametric settings. Within this line of work, the most common assumption is that the decision-maker has access to *uncensored* demand samples. For this setting, the Sample Average Approximation (SAA) algorithm was first shown to be near-optimal in Levi et al. (2007); their bounds were later refined by Levi et al. (2015) and Lin et al. (2022). Cheung and Simchi-Levi (2019) showed that these upper bounds are tight by providing a lower bound on the number of demand samples required for an algorithm to achieve low relative regret with high probability. While the bounds provided by these latter works are most relevant in the asymptotic regime in which the number of samples is large, Besbes and Mouchtaki (2023) more recently provided an exact analysis of the SAA algorithm across all data sizes, demonstrating that only tens of samples suffice for this algorithm to achieve strong performance. Contrary to this line of work, we assume that the decision-maker only has access to *censored* demand samples, a setting in which the classical SAA algorithm is not implementable, *sans* modification. We note however that, for the special case where the minimum historical ordering quantity goes to infinity, our setting reduces to the uncensored setting, in which case our results recover near-optimality of the SAA algorithm.

While censored demand has been studied in the literature, the vast majority of existing work considers the *online* setting, in which the decision-maker makes sequential ordering decisions over a finite horizon. Early work by Godfrey and Powell (2001) and Huh and Rusmevichientong (2009) demonstrated the strong performance of gradient-based methods in these settings. Huh et al. (2011) showed that adaptively using the well-known Kaplan-Meier (KM) estimator (Kaplan and Meier 1958), which reconstructs the empirical cumulative distribution function by uniformly redistributing the mass of censored observations to uncensored observations, converges almost surely to the set of optimal solutions under discrete demand. Besbes and Muharremoglu (2013) later characterized the implications of demand censoring in repeated newsvendor problems. Less closely related to our work is the design of sequential decision-making algorithms under censored demand, for more general models with inventory carry-over (e.g., with capacity constraints (Shi et al. 2016), setup costs (Yuan et al. 2021, Fan et al. 2024), positive lead times (Agrawal and Jia 2022, Xie et al. 2024), uncertain supply (Chen et al. 2024)), and in nonstationary settings (Lugosi et al. 2024).

The online setting differs fundamentally from the offline setting in the presence of censored demand since, in the online case, the decision-maker can adaptively adjust order quantities based on observed sales, thereby influencing future samples. In contrast, the offline setting offers no such control. This distinction partially explains the dearth of literature on the impact of censored data on single-period newsvendor decisions. Ban (2020) was the first to tackle this question, under the assumption that the maximum historical ordering quantity exceeds the optimal ordering quantity (i.e., in the identifiable regime). They derive an asymptotically consistent estimator for the optimal policy and use this to provide asymptotic confidence intervals. In an early version of their work, Fan et al. (2022) studied the sample complexity of learning the optimal ordering quantity when historical samples are generated from a given data-collecting policy. They show that a variant of the SAA algorithm is near-optimal, assuming that (i) historical inventory levels are independently and identically distributed, and (ii) a constant fraction of historical inventory levels exceed the optimal ordering quantity (an assumption that is analogous to identifiability in our setting).[1] We make two important contributions relative to these recent works. Firstly, contrary to Ban (2020), we are interested in designing algorithms with *finite-sample* guarantees. More importantly, however, the primary motivator behind our work is to characterize the fundamental limits that demand censoring places on algorithm performance, therefore requiring us to relax the identifiability assumptions upon which both of these papers rely. This motivation allows us to design simple algorithms that achieve these limits with strong guarantees *across all censoring regimes* (i.e., irrespective of how the historical ordering quantities are generated, and even if *no* historical ordering quantity exceeds the optimal newsvendor quantity).

Finally, we highlight that while the KM estimator is applicable to our setting, to the best of our knowledge, finite-sample guarantees do not exist for this heuristic. More importantly, however, this estimator only estimates the tail of the distribution at the observable boundary; past the observable boundary, it is not defined. As a result, we do not expect it to perform well in the unidentifiable regime, when the optimal newsvendor quantity is past the maximum historical ordering level. We demonstrate its poor performance in this regime in our numerical experiments (see Section 6).

Data-Driven Newsvendor: Robust Setting. While our work is concerned with the non-parametric setting, we briefly mention the large body of literature on robust and Bayesian newsvendor models, where the decision-maker has access to additional information on the underlying demand distribution (e.g., a parametric form, or its moments).

Our modeling framework's reliance on the ambiguity set induced by the true demand distribution and the observable boundary λ is directly inspired by the distributionally robust newsvendor literature. In this body of work, Scarf (1958) first studied the minimax solution of the newsvendor problem when both the mean and variance of the true demand distribution are known; Gallego and Moon (1993) later simplified and extended the analysis to variants of the newsvendor model. Perakis and Roels (2008) consider the minimax regret

[1] Subsequent to our work, Fan et al. (2025) echoed our results, establishing that under i.i.d. data-collecting policies, this condition is necessary to learn the newsvendor solution to a certain accuracy.

objective when the decision-maker knows various distributional quantities such as the range and median. More recent work has assumed knowledge of the distribution's semi-variance, under demand asymmetry (Natarajan et al. 2018), and considered the distributionally robust newsvendor problem under a Wasserstein ambiguity set (Lee et al. 2021).

Contrary to this line of work, however, our ambiguity set construction exists not because of some exogenously given information about the true demand distribution; rather, it arises from the fact that demand is unobservable past λ. Before the observable boundary, we assume no additional information on the demand distribution. In light of this, our work can best be viewed as lying at the intersection of the literature on non-parametric and distributionally robust newsvendor models. Recent work by Xu et al. (2022) and Fu et al. (2024) on the uncensored data-driven newsvendor is philosophically in the same hybrid vein, as they leverage historical samples to construct an ambiguity set defined by the nonparametric characteristics of the true distribution. We note that the notion of ambiguity set we consider is related to the first-order stochastic dominance (FSD) ambiguity set defined in Fu et al. (2024). Our ambiguity set, however, cannot exactly be cast as a FSD ambiguity set, given that we assume that the mean, as opposed to the support, of the underlying demand distribution is bounded. Moreover, Fu et al. (2024) are interested in the absolute cost, as opposed to the notion of regret we consider to characterize the performance of our learning algorithm. Finally, Besbes et al. (2022) similarly consider various distributionally robust optimization formulations for the uncensored newsvendor problem, when samples are drawn from a shifted distribution. Similar to our setting, they are interested in proving the convergence of data-driven policies to the minimax regret (though without finite-sample guarantees).

Impact of Demand Censoring: Beyond the Classical Newsvendor. There exists a large body of work in operations on demand estimation under censored data. A common approach is to assume that the demand distribution has a parametric form (see, e.g., Nahmias (1994), Agrawal and Smith (1996), Vulcano et al. (2012), Mersereau (2015)). In contrast, our focus is on the nonparametric setting, as previously discussed. Most relevant to our work in this regard is Bu et al. (2023), who introduced the concept of problem identifiability for an offline *pricing* problem under censored data, and whose hybrid robustness framework we leverage to characterize the impact of demand censoring on the newsvendor problem.

2. Problem Formulation

Technical notation. In what follows, for $N \in \mathbb{N}_+$, we let $[N] = \{1, 2, \dots, N\}$. We moreover use $\mathbb{P}_G(\cdot)$ and $\mathbb{E}_G[\cdot]$ to respectively denote the probability of an event and the expectation of a random variable when the source of underlying randomness has a cumulative distribution function (cdf) G.

Model primitives. We consider the classical single-period newsvendor problem, in which a decision-maker faced with random demand $D \geq 0$ must decide on the number of units to satisfy this demand. We let G denote the cdf of D, and assume that $\mathbb{E}_G[D] < \infty$.

For any ordering decision $q \geq 0$, once demand is realized and fulfilled to the maximum extent possible, the decision-maker incurs a lost sales penalty $b > 0$ for each unsatisfied unit of demand (also referred to as the *underage cost*). If all demand is satisfied and there is leftover inventory at the end of the period, the decision-maker incurs a per-unit *overage cost* $h > 0$. Given G, the decision-maker's goal is to determine an ordering quantity that minimizes the newsvendor cost, given by:

$$C_G(q) = \mathbb{E}_G\left[b(D-q)^+ + h(q-D)^+\right], \tag{1}$$

where $(\cdot)^+ = \max\{\cdot, 0\}$ is used to denote the positive part. When G is known, the optimal ordering quantity, denoted by $q_G^\star$, is determined by the *critical ratio* $\rho = \frac{b}{b+h} \in (0,1)$ (Zipkin 2000). Formally, the so-called *critical quantile* $q_G^\star$ is the ρ-th quantile of G, i.e.,

$$q_G^\star = \inf\{q \mid G(q) \geq \rho\}. \tag{2}$$

In the data-driven setting we consider, however, the decision-maker does not know G, but has access to historical (equivalently, offline) *sales* data. The historical data is defined by $K \in \mathbb{N}^+$ historical ordering quantities, denoted by q_k^{off}, for $k \in [K]$, with $q_1^{\text{off}} < q_2^{\text{off}} < \ldots < q_K^{\text{off}}$. We assume that the historical ordering quantities are exogenous and fixed.For each ordering quantity q_k^{off}, $k \in [K]$, the decision-maker observes N_k sales samples $s_{ki}^{\text{off}} = \min\{d_{ki}^{\text{off}}, q_k^{\text{off}}\}$, $i \in [N_k]$, where d_{ki}^{off} refers to the true realization of historical demand in period i associated with ordering quantity q_k^{off}, assumed to be drawn independently and identically (i.i.d.) from G. We moreover assume that d_{ki}^{off} and q_k^{off} are independent, for all $k \in [K], i \in [N_k]$. Let $d^{\text{off}} = \left(d_{ki}^{\text{off}}, i \in [N_k], k \in [K]\right)$ and $s^{\text{off}} = \left(s_{ki}^{\text{off}}, i \in [N_k], k \in [K]\right)$. We let $N = N_K$ be the number of samples associated with the largest historical ordering quantity; we moreover denote $\lambda = q_K^{\text{off}}$ to be this largest historical ordering quantity. As we will later see, this latter quantity, henceforth referred to as the *observable boundary* of the dataset, plays a key role in learning the optimal order quantity.

Finally, we assume that the decision-maker has access to a known upper bound M on $q_G^\star$, and let $\mathcal{G} = \left\{F \mid F \text{ has nonnegative support}, \mathbb{E}_F[D] < \infty,\, q_F^\star \leq M\right\}$ be the set of distributions satisfying the above assumptions on the true underlying cdf G.

REMARK 1. The assumption that M is known is common in the newsvendor literature (Huh and Rusmevichientong (2009), Besbes and Muharremoglu (2013), Ban (2020)). For instance, if the decision-maker knows the support of the underlying distribution G, she may take M to be an upper bound on this support. Our main results will show that access to a finite upper bound on $q_G^\star$ is in fact *necessary* to avoid pathologies in which infinite regret is unavoidable for the decision-maker. This fact will become clear in the subsequent analysis; hence, we defer a counterexample establishing necessity of this assumption to Remark 3.

Objective. Let π denote a policy which takes as input the historical sales data s^{off}, and outputs an ordering quantity $q^{\pi} \in [0, M]$. We use Π to denote the set of all such mappings.

In the classical data-driven newsvendor problem, the decision-maker observes the true realizations of historical demand, rather than censored sales data. In this uncensored setting, a standard metric to evaluate the performance of a policy π is the additive optimality gap of q^{π} relative to the optimal ordering quantity $q_G^{\star}$. We refer to this optimality gap as the *vanilla regret* of policy π, formally defined as:

$$\text{Vanilla-Regret}(q^{\pi}) = C_G(q^{\pi}) - C_G(q_G^{\star}). \tag{3}$$

In the censored setting we consider, however, this notion of regret may or may not be meaningful, depending on the level of demand censoring. To see this, consider an extreme case where $\lambda = +\infty$, and demand is bounded. In this case, the decision-maker always observes the true demand when λ was the ordering quantity. This setting reduces to the classical uncensored data-driven newsvendor over these demand samples, in which case competing against $q_G^{\star}$ is indeed a meaningful metric.

On the other hand, consider the pathological case in which $\lambda = 0$ (i.e., $q_k^{\text{off}} = 0 \;\forall\, k \in [K]$). In this case, the decision-maker never observes any demand information! Since the dataset is completely uninformative in this case, it is philosophically more natural to consider an adversarial framework in which the policy should aim to perform well against *any* distribution chosen by nature.

To interpolate between these two pathological cases, we draw from the distributionally robust optimization (DRO) literature, and consider the *ambiguity set* of demand distributions induced by λ.

DEFINITION 1 (AMBIGUITY SET). Given λ, the *ambiguity set* associated with G is the set of all distributions $F \in \mathcal{G}$ that share the same cdf as G, for $x < \lambda$.Formally:

$$\mathcal{F}(\lambda; G) = \{F \in \mathcal{G} \mid F(x) = G(x) \;\forall\, x < \lambda\}. \tag{4}$$

When $\lambda = 0$, $\mathcal{F}(\lambda; G) = \mathcal{G}$, i.e., the ambiguity set consists of all possible cdfs satisfying our mild distributional assumptions. When $\lambda = +\infty$, however, there is no ambiguity surrounding the underlying demand distribution, and $\mathcal{F}(\lambda; G) = \{G\}$. Now, for the non-pathological cases where $\lambda \in (0, +\infty)$, the ambiguity set captures the idea that, prior to λ, there may be hope of reconstructing the true cdf G, required to compute $q_G^{\star}$, by Equation (2). Past λ, however, the decision-maker never observes any demand realizations. Hence, even with infinitely many samples, she will never be able to estimate the tail of the true demand distribution. In this case, we recover the adversarial framework that was motivated when $\lambda = 0$; namely, for all intents and purposes, the tail of the demand distribution past λ can be arbitrary.

Given the ambiguity set $\mathcal{F}(\lambda; G)$, our performance metric for any policy $\pi \in \Pi$ is its worst-case optimality gap against the optimal newsvendor cost of *any* distribution $F \in \mathcal{F}(\lambda; G)$. We refer to this additive optimality gap as the *regret* of a policy, which we formally define below.

Definition 2 (Regret). Given true demand distribution G and dataset s^{off} with observable boundary λ, the *regret* of policy $\pi \in \Pi$ with access to censored demand samples s^{off} is defined as:

$$\text{Regret}(q^{\pi}) = \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^{\pi}) - C_F(q_F^{\star}). \tag{5}$$

We will equivalently refer to the regret of a policy as its *worst-case* regret, in line with the fact that we take the supremum over all $F \in \mathcal{F}(\lambda;G)$.Notice that the regret of a policy, as defined in Equation (5), implicitly depends on (i) the true demand distribution G, and (ii) the observable boundary λ. As a result, our guarantees will be instance-dependent, as a function of these two model primitives, in addition to b and h.

The aim is to design a policy with low regret, for any true underlying distribution G, as we scale N, the number of data samples at the boundary.[2] In the remainder of our work, we refer to this problem as the *data-driven censored newsvendor problem*.

Observe that, since $G \in \mathcal{F}(\lambda;G)$ for all λ, $\text{Regret}(q^{\pi}) \geq \text{Vanilla-Regret}(q^{\pi})$ for any policy π. In the same vein, since $\mathcal{F}(\lambda';G) \subset \mathcal{F}(\lambda;G)$ for $\lambda' > \lambda$, the regret of a policy is (weakly) decreasing in λ. This formalizes the natural idea that the censored setting is indeed at least as hard as the uncensored setting, and only becomes more challenging as more demand realizations are censored. In the following section we formalize the extent to which this is true, and how this depends on the value of λ.

Remark 2. While we make minimal assumptions on the underlying demand distribution for our main set of results, in Appendix B we explore the value the decision-maker can derive from additional distributional information. In particular, we study the case where the distribution is *globally well-separated* (i.e., continuous with a known lower bound $\gamma > 0$ on its probability density function).

3. Cost of Demand Censoring in the Data-Driven Newsvendor

In the uncensored setting, it is well-known that the vanilla regret (defined in Equation (3)) of any policy is lower bounded by

$$\inf_{\pi \in \Pi} \mathbb{E}_G\left[C_G(q^{\pi}) - C_G(q_G^{\star})\right] = \Omega(1/\sqrt{N}).$$

This lower bound is achieved by the classical sample average approximation (SAA) algorithm (Levi et al. 2007), which outputs the ρ-th empirical quantile of G, i.e.,

$$q_{\widehat{G}}^{\star} = \inf\left\{q \mid \frac{1}{\sum_{k\in[K]} N_k} \sum_{k=1}^{K} \sum_{i=1}^{N_k} \mathbb{1}\left\{d_{ki}^{\text{off}} \leq q\right\} \geq \rho\right\}. \tag{6}$$

We first highlight that this solution is not implementable in the censored setting, since d_{ki}^{off} is unobserved. While there exist natural adaptations of $q_{\widehat{G}}^{\star}$ to the censored setting (e.g., replacing $\mathbb{1}\left\{d_{ki}^{\text{off}} \leq q\right\}$ by $\mathbb{1}\left\{s_{ki}^{\text{off}} \leq q\right\}$

[2] The number of samples associated with the largest ordering level is a common choice for the scaling parameter in censored settings (see, e.g., Bu et al. (2023) and Fan et al. (2022)). At a high level, N can be viewed as the "bottleneck" of the dataset, in that it is the limiting factor in accurately estimating the true underlying cdf G.

in Equation (6), or conditioning on uncensored samples), we show in Section 6 that these naive solutions exhibit remarkably poor performance, both on vanilla and worst-case regret. Rather than searching for a policy that exhibits strong performance, however, in this section we ask a more fundamental question: *Is vanishing regret even achievable in the censored setting?*

We return to our extreme cases to answer this question neither in the affirmative nor in the pejorative. To see this, consider first the case where $\lambda = +\infty$ and demand is bounded. In this case, $\mathcal{F}(\lambda; G) = \{G\}$, as argued in Section 2. Moreover, samples for which λ was the ordering quantity are uncensored; computing $q^\star_{\widehat{G}}$ over these samples guarantees $O(1/\sqrt{N})$ regret, by existing results for the uncensored setting (Chen and Ma 2024). Consider now the case where $\lambda = 0$. In this case, the ambiguity set $\mathcal{F}(\lambda; G) = \mathcal{G}$, implying that the quantity output by any policy π must simultaneously compete against the space of *all* possible distributions. For any quantity q^π, however, it is not difficult to construct a distribution such that q^π incurs constant loss relative to the optimal ordering quantity $q^\star_F$, *independent of the number of samples* N. To illustrate this idea, consider the policy that computes the ρ-th empirical quantile with respect to sales. This policy outputs $q^\pi = 0$ when $\lambda = 0$. Consider now the atomic distribution $F \in \mathcal{F}(\lambda; G)$ for which $D = M$ with probability 1. For this distribution, $q^\star_F = M$, which implies that $\mathrm{Regret}(q^\pi) = \Omega(M)$, a constant independent of N.

Thus motivated, in this section we investigate the dependence of the optimal achievable performance of any policy on the observable boundary λ.

3.1. Algorithm Performance and Minimax Risk

Understanding whether or not vanishing regret is achievable, for a fixed λ, is closely related to the concept of problem identifiability, first introduced by Bu et al. (2023).

DEFINITION 3 (PROBLEM IDENTIFIABILITY). Given true demand distribution G and observable boundary λ, the data-driven censored newsvendor problem is *identifiable* if there exists a policy π such that, for any $\epsilon > 0$:

$$\lim_{N\to\infty} \mathbb{P}_G\left[\sup_{F\in\mathcal{F}(\lambda;G)} \left\{C_F(q^\pi) - C_F(q^\star_F)\right\} < \epsilon\right] = 1, \tag{7}$$

where the historical ordering quantities q_k^{off} are fixed, and the probability is taken over the randomness in the censored demand samples $d_{ki}^{\mathrm{off}} \sim G$. If no such policy exists, the problem is *unidentifiable*.

In words, a problem is unidentifiable if, even with infinitely many samples, no policy can achieve zero regret in the worst case. Note the unwieldiness of this definition, however, as it is a statement about the space of *all* policies $\pi \in \Pi$. To gain further insight into identifiability, we introduce the more tractable notion of minimax risk, which will be central to all of our results.

DEFINITION 4 (MINIMAX RISK). The *minimax risk* Δ of a data-driven censored newsvendor problem defined by true demand distribution G and observable boundary λ is the minimum achievable regret of any constant $q \in [0, M]$. Formally:

$$\Delta = \inf_{q\in[0,M]} \sup_{F\in\mathcal{F}(\lambda;G)} \left\{C_F(q) - C_F(q^\star_F)\right\}. \tag{8}$$

We refer to the quantity that achieves Δ, if it exists, as the *minimax optimal ordering quantity* q^{Δ}.

At a high level, the minimax risk can be thought of as the *cost of demand censoring*. It quantifies the information loss due to the decision-maker never being able to observe demand realizations that exceed λ. This concept moreover allows us to distinguish between the two sources of information loss in the data-driven censored newsvendor problem: (i) the information loss due to the fact there are finitely many samples of data, which also exists in the uncensored setting and depends on N, and (ii) the information loss due to censoring, which is dependent on λ, but independent of N.

The following proposition creates a formal connection between the minimax risk and identifiability. It implies that, if $\Delta > 0$, *the regret of any policy is lower bounded by a constant.*

PROPOSITION 1. *The data-driven censored newsvendor problem is identifiable if and only if* $\Delta = 0$.

Proposition 1 thus motivates us to move the goalpost for the success of any given policy to incurring regret that is upper bounded by $\Delta + o(1)$. We say that any policy that achieves this is *near-optimal*.

We defer the proof of Proposition 1 to Appendix D.1.1. While it is easy to establish that identifiability implies $\Delta = 0$, the other direction is more challenging, as it requires exhibiting a policy that achieves vanishing regret when $\Delta = 0$. The proof of this direction actually follows from our main algorithmic contribution in Section 4, in which we design and analyze such a policy.

We highlight that neither Δ nor q^{Δ} is computable by the decision-maker, since the supremum in Equation (8) is taken over the ambiguity set $\mathcal{F}(\lambda; G)$, which itself depends on G, unknown to the decision-maker. As a result, Δ is not an operationalizable quantity. Despite this fact, in the following section, we obtain *explicit characterizations* of Δ and q^{Δ}, depending on the problem primitives. Not only will these be important in understanding the cost of demand censoring in the data-driven newsvendor problem, but they will also be instrumental in designing a simple algorithm that asymptotically achieves the fundamental lower bound Δ.

3.2. Closed-Form Characterization of Minimax Risk

In our first main contribution, we leverage the more tractable notion of minimax risk by providing a closed-form characterization of Δ and q^{Δ}, and in doing so obtain a simple necessary and sufficient condition for identifiability. For ease of notation, we let $G^{-}(\lambda) = \mathbb{P}_G(D < \lambda)$ denote the mass of demand observations that lie before λ. At a high level, $G^{-}(\lambda)$ can be viewed as a proxy for the "censoring level" of the data, since $1 - G^{-}(\lambda)$ is precisely the fraction of demand that the decision-maker will never observe, even with infinitely many samples. Theorem 1 establishes that whether or not a problem is identifiable hinges solely on the relative ordering of $G^{-}(\lambda)$ and ρ.

THEOREM 1. *Consider a data-driven censored newsvendor problem with demand distribution G and observable boundary λ.*

1. *If $G^{-}(\lambda) \geq \rho$:*

$$q^{\Delta} = q^{\star}_G < \lambda \qquad , \qquad \Delta = 0.$$

2. *If* $G^-(\lambda) < \rho$*:*

$$q^\Delta = \frac{bM + h\lambda - (b+h)G^-(\lambda)M}{(b+h)(1-G^-(\lambda))} \qquad , \qquad \Delta = \frac{h(b-(b+h)G^-(\lambda))(M-\lambda)}{(b+h)(1-G^-(\lambda))} \geq 0.$$

For ease of notation, in the unidentifiable regime we refer to the minimax optimal ordering quantity q^Δ as the *unidentifiable ordering quantity*, denoted as:

$$q_G^\dagger = \frac{bM + h\lambda - (b+h)G^-(\lambda)M}{(b+h)(1-G^-(\lambda))}. \tag{9}$$

The following result then emerges as a corollary of Theorem 1 and Proposition 1.

COROLLARY 1. *A data-driven censored newsvendor problem with demand distribution G and observable boundary λ is identifiable if and only if $G^-(\lambda) \geq \rho$, or if $G^-(\lambda) < \rho$ and $M = \lambda$.*

In Appendix C.2 we establish that $G^-(\lambda) < \rho$ implies $\lambda \leq q_G^\star \leq M$. Hence, for $G^-(\lambda) < \rho$ and $M = \lambda$ to simultaneously hold, it must be that $q_G^\star = \lambda = M$, a condition that we do not expect to hold in practice. For instance, if $G(x)$ is continuous for all $x < \lambda$, it is not hard to see that $G^-(\lambda) < \rho$ implies $q_G^\star > \lambda$, and so $\lambda < M$. We henceforth make the following mild assumption in the remainder of the paper.

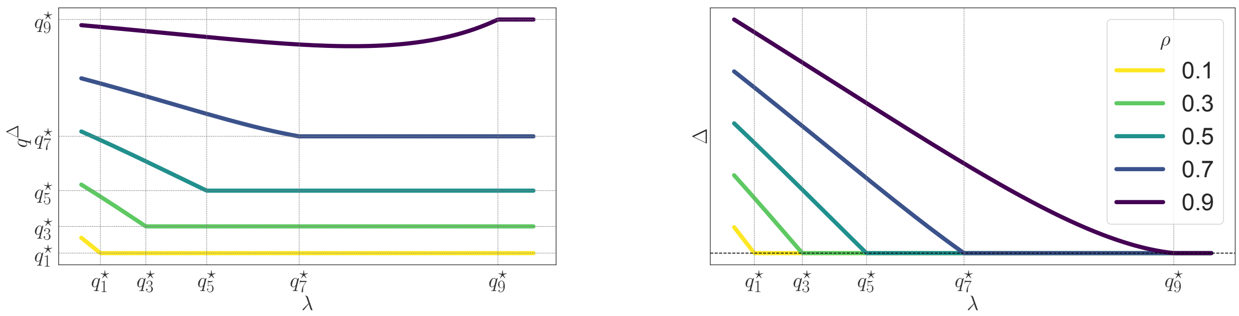
ASSUMPTION 1. *Suppose $G^-(\lambda) < \rho$. Then, $\lambda < M$.*

Under this assumption, we say that we are in the *identifiable regime* if $G^-(\lambda) \geq \rho$, and in the *unidentifiable regime* if $G^-(\lambda) < \rho$.[3]

We discuss the implications of Theorem 1, deferring a proof sketch to the end of the section. At a high level, Theorem 1 can be viewed as characterizing a "spectrum of achievability" for the data-driven censored newsvendor problem, based on the additional distributional information of $G^-(\lambda)$. In particular, when $\lambda = +\infty$ and demand is bounded, we have that $G^-(\lambda) = 1$. Theorem 1 then recovers the fact that the optimal newsvendor cost $C_G(q_G^\star)$ is achievable with infinitely many samples in this case, and that the problem is identifiable as a result. When $\lambda = 0$, on the other hand, $G^-(\lambda) = 0$ Theorem 1 then classifies the problem as unidentifiable, with $\Delta = \frac{bhM}{b+h} > 0$. This recovers the intuition that, for any policy, there exists a worst-case distribution that forces it to incur constant loss.

Figure 1 illustrates the dependence of the minimax risk Δ and the minimax optimal ordering quantity q^Δ on λ, as it interpolates between these two extremes. For $G^-(\lambda) < \rho$, even with infinitely many historical samples, for any policy that outputs q^π, there exists a distribution in the ambiguity set that can make it such that the policy over- (resp., under-) orders, causing constant regret. This can be seen in Figure 1b, where $\Delta > 0$ for all $\lambda < q_G^\star$. As λ increases from 0 (and $G^-(\lambda)$ increases as a result), Δ strictly decreases. This reflects the phenomenon that, as the observable boundary increases, the size of the ambiguity set

[3] We highlight that Assumption 1 does not affect any of our results. It is simply introduced to make clear that the main salient driver of identifiability is the relationship between $G^-(\lambda)$ and ρ, rather than a spurious artifact of the maximum ordering quantity M.



(a) q^Δ vs. λ (b) Δ vs. λ

Figure 1 **Dependence of q^Δ and Δ on λ for $D \sim$ Exponential$(1/80)$, $M = 200$, $h = 1$, and $\rho \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. We abuse notation and let $q^\star_{10\rho}$ denote the optimal newsvendor quantity associated with ρ. By Theorem 1, $\lambda \geq q^\star_{10\rho}$ corresponds to the identifiable regime.**

decreases, which causes a decrease in information loss. Once λ reaches $q^\star_G$ (i.e., $G^-(\lambda) \geq \rho$), we have a phase transition, with $\Delta = 0$. The problem becomes identifiable, with vanishing regret being achievable throughout this region, despite the fact that demand remains censored by the historical ordering quantities. Hence, Theorem 1 highlights that, for a wide range of values of λ, demand censoring is not a barrier to effective decision-making.

We conclude our discussion with an analysis of the minimax optimal ordering quantity q^Δ. Perhaps surprisingly, Theorem 1 establishes that $q^\Delta = q^\star_G$, for *any* value of $G^-(\lambda) \geq \rho$. This can be seen in Figure 1a, where q^Δ plateaus for $\lambda \geq q^\star_G$. So, not only is vanishing regret achievable, but in fact, nature is so limited in this regime that *any* distribution $F \in \mathcal{F}(\lambda; G)$ is such that $q^\star_F = q^\star_G$. When $G^-(\lambda) < \rho$, on the other hand, q^Δ hedges between over- and under-ordering. This can best be seen when $G^-(\lambda) = 0$, in which case q^Δ mixes between the maximum ordering quantity M and the observable boundary λ at a rate of ρ, by Theorem 1. This intuition is also reflected in the fact that, for fixed $G^-(\lambda)$, Δ is increasing in $M - \lambda$, since hedging between over- and under-ordering becomes harder as unseen demand samples take on a larger set of possible values. Finally, Figure 1a and Figure 1b illustrate that both q^Δ and Δ are increasing in b, for fixed values of h and λ, as under-ordering becomes costlier.

We now provide a proof sketch of Theorem 1, highlighting important auxiliary results upon which we rely in the rest of the paper. We defer a formal proof of the theorem to Appendix D.1.4.

Proof sketch. Consider first the case where $G^-(\lambda) \geq \rho$. Lemma 1 first establishes that, for all $F \in \mathcal{F}(\lambda; G)$, the optimal newsvendor ordering quantity $q^\star_F$ coincides with $q^\star_G$. Its proof can be found in Appendix D.1.2.

LEMMA 1. *Suppose $G^-(\lambda) \geq \rho$. Then, for all $F \in \mathcal{F}(\lambda; G)$, $q^\star_F = q^\star_G < \lambda$.*

Importantly, *this fact is independent of the decision-maker's ordering decision q*. Letting $q = q^\star_G$, this then implies that $C_F(q) - C_F(q^\star_F) = 0$, for all $F \in \mathcal{F}(\lambda; G)$. Hence, by definition of the minimax risk, $\Delta = 0$, with $q^\Delta = q^\star_G$.

Consider now the case where $G^{-}(\lambda) < \rho$. Lemma 2 is the main workhorse for our result: it provides closed-form expressions for the worst-case regret over $\mathcal{F}(\lambda; G)$, given q. Its proof can be found in Appendix D.1.3.

LEMMA 2. *Suppose $G^{-}(\lambda) < \rho$. Then, for any $q \in [0, M]$,*

$$\begin{aligned}
\text{Regret}(q) &:= \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) \\
&= \begin{cases} b(M-q) + (b+h)\Big[\mathbb{E}_G\big[(q-D)\mathbb{1}\{D \le q\}\big] - (M-D)\mathbb{1}\{D < \lambda\}\big]\Big] & \text{if } q < \lambda, \\ (b-(b+h)G^{-}(\lambda))(M-q) & \text{if } q \in \left[\lambda, q_G^\dagger\right] \\ h(q-\lambda) & \text{if } q > q_G^\dagger. \end{cases}
\end{aligned}$$

At a high level, when $q < \lambda$ (i.e., the decision-maker orders no more than what she has historically observed), any worst-case demand distribution sets $q_F^\star = M$, ensuring that q incurs high underage costs. This is in line with the intuition that exploration is important under high censoring levels.

When $q \ge \lambda$, on the other hand, the worst-case demand distribution selects the worst of $q_F^\star = M$ (high underage costs) and $q_F^\star = \lambda$ (high overage costs). We show that the former is optimal for nature when $q \in [\lambda, q_G^\dagger]$, while the latter is optimal for $q > q_G^\dagger$.

We finalize the argument by comparing the three regimes of q, and showing that the minimax risk is achieved at $q = q_G^\dagger$. At this point,

$$(b-(b+h)G^{-}(\lambda))(M-q) = h(q-\lambda) = \frac{h(M-\lambda)(b-(b+h)G^{-}(\lambda))}{(b+h)(1-G^{-}(\lambda))}.$$

Therefore, q is precisely set such that nature is indifferent between enforcing high underage and high overage costs, given this ordering quantity. The expression for Δ follows from algebra. Finally, it is easy to argue that, when $G^{-}(\lambda) < \rho$, $q_G^\dagger \in [\lambda, M]$ and $\Delta \ge 0$. □

Characterization of worst-case distributions and behavior of worst-case regret. The above analysis allows us to gain insight into the worst-case distribution that achieves the minimax risk Δ. To formalize this, we define a distribution F_p parametrized by $p \in \{0, 1 - G^{-}(\lambda)\}$:

$$F_p(x) = \begin{cases} G(x) & x < \lambda \\ G^{-}(\lambda) + p & x \in [\lambda, M) \\ 1 & x = M \end{cases}$$

By construction, $F_p \in \mathcal{F}(\lambda; G)$. Moreover, when $p = 0$, F_p places the entirety of the remaining mass on M; when $p = 1 - G^{-}(\lambda)$, on the other hand, the entirety of the remaining mass is placed on λ. We define the *restricted ambiguity set* as the set of all valid distributions F_p, i.e.,:

$$\mathcal{B}(\lambda; G) = \left\{F_p \mid p \in \{0, 1 - G^{-}(\lambda)\}\right\}.$$

Proposition 2 below establishes that $\mathcal{B}(\lambda; G)$ is indeed the worst-case family of distributions, for any ordering quantity q.

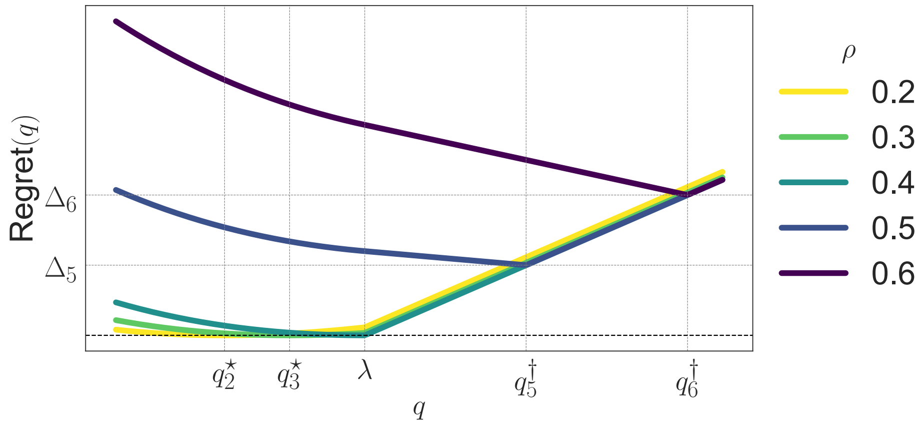


Figure 2 **Regret(q) vs. q for $D \sim$ Exponential$(1/80)$, $\lambda = G^{-1}(0.4)$, $M = 200$, $h = 1$, and $\rho \in \{0.2, 0.3, 0.4, 0.5, 0.6\}$. We abuse notation and let $q^\star_{10\rho}$ and $q^\dagger_{10\rho}$ respectively denote the optimal newsvendor and unidentifiable ordering quantities associated with ρ. Similarly, $\Delta_{10\rho}$ denotes the value of Δ under parameter ρ. When $\rho = 0.4$, $G^-(\lambda) = \rho$ by construction. In this case, we have $q_4^\dagger = q_4^\star = \lambda$. Finally, the bottom-most dashed line corresponds to Regret$(q) = 0$.**

PROPOSITION 2. *Fix $q \in [0, M]$. Then,*

$$\sup_{F \in \mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) = \sup_{F \in \mathcal{B}(\lambda;G)} C_F(q) - C_F(q_F^\star).$$

We defer the proof of the fact to Appendix D.2. Proposition 2 provides useful intuition as to the drivers of regret, as previously discussed. Namely, it precisely reflects the intuition of nature adversarially generating a high- or low-demand distribution to ensure that the decision-maker incurs high underage or overage costs, respectively. [4]

REMARK 3. Theorem 1 and Proposition 2 together help to establish why access to an upper bound M on $q_G^\star$ is in fact necessary for the minimax risk Δ to be bounded in the unidentifiable regime. In particular, Part 2 of Theorem 1 implies that $\Delta \to +\infty$ as $M \to +\infty$, achieved by a worst-case distribution in $\mathcal{B}(\lambda; G)$ setting $q_F^\star = M \to +\infty$.

4. A Near-Optimal Algorithm

As discussed in Section 3, our goal is to design an algorithm whose regret asymptotically achieves the minimax risk Δ, *across both regimes of identifiability*. Theorem 1, however, established that even determining identifiability requires knowledge of $G^-(\lambda)$, which the decision-maker does not have. Therefore, performing well across all regimes seems to be a challenging task. In this section, we design a robust algorithm that achieves the optimal order of regret (up to polylogarithmic factors) in both the identifiable and unidentifiable regimes, in spite of this challenge.

[4] Similar insights have been derived in the uncensored setting (see, e.g., Besbes and Mouchtaki (2023), Perakis and Roels (2008)).

4.1. Algorithm Description

At a high level, our algorithm, Robust Censored Newsvendor (RCN), proceeds in a hierarchical fashion. Motivated by the above challenge, we first test for identifiability, i.e., whether $G^-(\lambda) \geq \rho$. While doing so would be trivial in the uncensored setting, recall that our algorithm only has access to *sales*, rather than true demand data. The following fact, however, will allow us to design a test that correctly determines whether $G^-(\lambda) \geq \rho$, with high probability.

FACT 1. *Let* $s^{\text{off},\lambda} = \left(s^{\text{off}}_{Ki} \mid i \in [N]\right)$. *For all* $i \in [N]$,

$$\mathbb{1}\left\{s_i^{\text{off},\lambda} < \lambda\right\} = \mathbb{1}\left\{\min\{\lambda, d_i^{\text{off}}\} < \lambda\right\} = \mathbb{1}\left\{d_i^{\text{off}} < \lambda\right\}.$$

Fact 1 therefore implies that we can compute an unbiased estimate of $G^-(\lambda)$ by restricting our attention to the samples at the boundary. Thus motivated, our algorithm computes the fraction of samples that lie strictly below λ, i.e.,

$$\widehat{G}^-(\lambda) = \frac{1}{N}\sum_{i\in[N]} \mathbb{1}\left\{s_i^{\text{off},\lambda} < \lambda\right\}.$$

We refer to this quantity as the *censored* estimate of $G^-(\lambda)$ since it uses the censored sales data, rather than the true (unobserved) demand data that would be required to compute the SAA of $G^-(\lambda)$.

Once computed, the censored estimate $\widehat{G}^-(\lambda)$ may fall into one of three cases, depending on the confidence parameter ζ that our algorithm takes as input. In particular, if $\widehat{G}^-(\lambda) \geq \rho + \zeta$, our algorithm classifies the problem as *likely identifiable*. In this case, motivated by Theorem 1, it outputs a censored estimate of $q^\star_G$, denoted by $q^\star_{\widehat{G}}$, defined as the ρ-th sample quantile of $s^{\text{off},\lambda}$. Formally:

$$q^\star_{\widehat{G}} = \inf\left\{x \mid \frac{1}{N}\sum_{i\in[N]} \mathbb{1}\left\{s_i^{\text{off},\lambda} \leq x\right\} \geq \rho\right\}.$$

For ease of notation, we let $\widehat{G}(x) = \frac{1}{N}\sum_{i\in[N]} \mathbb{1}\left\{s_i^{\text{off},\lambda} \leq x\right\}$. Note again here that $q^\star_{\widehat{G}}$ relies only on the samples for which the historical ordering quantity was λ, as was the case for $\widehat{G}^-(\lambda)$.

If, on the other hand $\widehat{G}^-(\lambda) < \rho - \zeta$, the algorithm classifies the problem as *likely unidentifiable*, and outputs a censored estimate of $q^\dagger_G$, denoted by $q^\dagger_{\widehat{G}}$ and defined as:

$$q^\dagger_{\widehat{G}} = \frac{bM + h\lambda - (b+h)\widehat{G}^-(\lambda)M}{(b+h)(1-\widehat{G}^-(\lambda))}.$$

Finally, if $\widehat{G}^-(\lambda) \in [\rho - \zeta, \rho + \zeta)$, our algorithm cannot determine problem identifiability with confidence, in which case it outputs λ. We present a formal description of our algorithm in Algorithm 1.

ALGORITHM 1: Robust Censored Newsvendor (RCN)

Input: Observable boundary λ, censored demand samples $s^{\mathrm{off},\lambda}$, confidence term ζ

Output: Ordering quantity q^{alg}

Compute censored SAA of $G^-(\lambda)$:

$$\widehat{G}^-(\lambda) = \frac{1}{N}\sum_{i\in[N]} \mathbb{1}\left\{s_i^{\mathrm{off},\lambda} < \lambda\right\}. \tag{10}$$

if $\widehat{G}^-(\lambda) \geq \rho + \zeta$ **then** `// Likely identifiable`

Compute censored SAA of $q_G^\star$:

$$q_{\widehat{G}}^\star = \inf\left\{x \mid \widehat{G}(x) \geq \rho\right\}, \tag{11}$$

where $\widehat{G}(x) = \frac{1}{N}\sum_{i\in[N]} \mathbb{1}\left\{s_i^{\mathrm{off},\lambda} \leq x\right\}$. Let $q^{\mathrm{alg}} = q_{\widehat{G}}^\star$.

else if $\widehat{G}^-(\lambda) < \rho - \zeta$ **then** `// Likely unidentifiable`

Compute empirical estimate of $q_G^\dagger$:

$$q_{\widehat{G}}^\dagger = \frac{bM + h\lambda - (b+h)\widehat{G}^-(\lambda)M}{(b+h)(1-\widehat{G}^-(\lambda))}. \tag{12}$$

Let $q^{\mathrm{alg}} = q_{\widehat{G}}^\dagger$.

else `// Knife-edge case`

Let $q^{\mathrm{alg}} = \lambda$.

return q^{alg}

4.2. Algorithm Analysis

Theorem 2 establishes that this intuitive algorithm achieves vanishing minimax regret relative to Δ with constant probability, for an appropriately tuned confidence parameter ζ. Due to space constraints, interested readers can find the proofs of all subsequent results at the following link: `https://anonymous.4open.science/r/rcn_msom_sig`.

THEOREM 2. *Fix $\delta \in (0,1)$, and let $\zeta = \sqrt{\frac{\log(2/\delta)}{2N}}$. With probability at least $1-2\delta$, Algorithm 1 outputs an ordering quantity q^{alg} such that:*

(i) if $G^-(\lambda) \geq \rho + 2\zeta$:

$$\mathrm{Regret}(q^{\mathrm{alg}}) \leq \Delta + \lambda(b+h)\zeta = \lambda(b+h)\zeta$$

(ii) if $G^-(\lambda) \in [\rho, \rho + 2\zeta)$:

$$\mathrm{Regret}(q^{\mathrm{alg}}) \leq \Delta + 2\lambda(b+h)\zeta = 2\lambda(b+h)\zeta$$

(iii) if $G^-(\lambda) \in [\rho - 2\zeta, \rho)$:

$$\mathrm{Regret}(q^{\mathrm{alg}}) \leq \Delta + 2\max\left\{\frac{b}{h}, 1\right\}(M-\lambda)(b+h)\zeta$$

(iv) if $G^-(\lambda) < \rho - 2\zeta$:

$$\mathrm{Regret}(q^{\mathrm{alg}}) \leq \Delta + \max\left\{\frac{b}{h}, 1\right\}(M-\lambda)(b+h)\zeta.$$

Letting $\delta = O(1/\sqrt{N})$, we obtain the following bound on the expected worst-case regret of Algorithm 1.

COROLLARY 2. *Let $\delta = c/\sqrt{N}$, for some constant $c > 0$. Then, there exists a constant $c' > 0$ (independent of M, λ, ρ, and N) such that:*

$$\mathbb{E}_G\left[\text{Regret}(q^{\text{alg}})\right] \leq \begin{cases} c'(b+h)(\lambda + \max\{\rho\lambda, (1-\rho)M\})\sqrt{\log N/N} & \text{if } G^-(\lambda) \geq \rho \\ \Delta + c'(b+h)\left(\max\left\{\frac{\rho}{1-\rho}, 1\right\}(M-\lambda) + \max\{\rho M, (1-\rho)(M-\lambda)\}\right)\sqrt{\log N/N} & \text{if } G^-(\lambda) < \rho. \end{cases}$$

We make a few observations on our algorithm's guarantees before providing a proof sketch of the theorem at the end of this section. (We defer a proof of Corollary 2 to Appendix D.4.) First of all, in the identifiable regime ($G^-(\lambda) \geq \rho$), since $q^\Delta = q^\star_G$ by Theorem 1, our algorithm recovers existing regret guarantees (up to constant factors) for the uncensored setting (Chen and Ma 2024). We emphasize that this result holds for *any* value of λ such that $G^-(\lambda) \geq \rho$, and for *any* set of historical ordering quantities. This further implies that demand censoring is effectively immaterial in the identifiable regime.

In the unidentifiable regime ($G^-(\lambda) < \rho$), it is also near-optimal, incurring an expected worst-case regret of $\Delta + O\left(\sqrt{\frac{\log N}{N}}\right)$. Observe that our upper bound on the algorithm's regret in this regime is linear in $M - \lambda$. This is in line with the construction of the worst-case distribution described in Section 3, for any quantity q: if q is close to λ, the worst-case distribution $F \in \mathcal{F}(\lambda; G)$ sets the optimal newsvendor quantity $q^\star_F$ to M; if q is close to M, on the other hand, F sets $q^\star_F$ to λ. As the gap between λ and M increases, it becomes more difficult for the algorithm to hedge against nature, which is reflected in the regret guarantee. We defer a formal proof to Appendix D.3.

We conclude the section by reframing Theorem 2 in terms of sample complexity, which ties the problem back to the question of identifiability. (Recall, the proof of the fact that $\Delta = 0$ implies identifiability in Proposition 1 relied on the ability to provide a policy π satisfying Definition 3.)

COROLLARY 3. *For any $\epsilon > 0$ and $\delta \in (0,1)$, if $N \geq N(\epsilon, \delta)$, then with probability at least $1 - 2\delta$, the worst-case regret of Algorithm 1 is no greater than $\Delta + \epsilon$, i.e.,*

$$\mathbb{P}_G\left[\sup_{F \in \mathcal{F}(\lambda; G)} \left\{C_F(q^{\text{alg}}) - C_F(q^\star_F) - \Delta\right\} < \epsilon\right] \geq 1 - 2\delta,$$

where $N(\epsilon, \delta) = \frac{2(b+h)^2 \log(2/\delta)}{\epsilon^2} \max\left\{\lambda^2, \max\{(b/h)^2, 1\}(M-\lambda)^2\right\}$.

Letting $\delta = O(1/\sqrt{N})$, Corollary 3 implies that, when $G^-(\lambda) \geq \rho$,

$$\lim_{N \to \infty} \mathbb{P}_G\left[\sup_{F \in \mathcal{F}(\lambda; G)} \left\{C_F(q^{\text{alg}}) - C_F(q^\star_F)\right\} < \epsilon\right] = 1,$$

which gives us identifiability, as promised in the proof of Proposition 1.

5. Lower Bound

In this section we show that the bounds established in Theorem 2 are tight, i.e., that any algorithm necessarily incurs regret lower bounded by $\Delta + \Omega(1/\sqrt{N})$, in expectation. While existing lower bounds apply to our problem (i.e., by taking $\lambda = +\infty$), we provide *instance-dependent* lower bounds, as a function of the regime of identifiability, which we formally define below.

DEFINITION 5 (IDENTIFIABILITY REGIMES). The three identifiability regimes of the data-driven censored newsvendor problem are defined as follows:

(i) the *strictly identifiable* regime: the set of all distributions G such that

$$G^{-}(\lambda) > \rho + \frac{c}{\sqrt{N}} \quad \text{for some } \ c > 0.$$

We let $\mathcal{G}^{\mathrm{id}} \subset \mathcal{G}$ denote the set of all such distributions.

(ii) the *knife-edge* regime: the set of all distributions G such that

$$|G^{-}(\lambda) - \rho| \leq \frac{c}{\sqrt{N}} \quad \text{for some } c > 0.$$

We let $\mathcal{G}^{\mathrm{ke}} \subset \mathcal{G}$ denote the set of all such distributions.

(iii) the *strictly unidentifiable* regime: the set of all distributions G such that

$$G^{-}(\lambda) < \rho - \frac{c}{\sqrt{N}} \quad \text{for some } \ c > 0.$$

We let $\mathcal{G}^{\mathrm{ui}} \subset \mathcal{G}$ denote the set of all such distributions.

Theorem 3 establishes that our algorithm achieves the optimal regret guarantee (up to polylogarithmic factors) across the three identifiability regimes. We refer the reader to Appendix D.5 for the formal proof of this result.

THEOREM 3. *Fix b, h, λ, and M, with $\lambda < M$. For any policy $\pi \in \Pi$ with access to N censored demand samples $s^{\mathrm{off},\lambda}$:*

(i) in the strictly identifiable *regime:*

$$\sup_{G \in \mathcal{G}^{\mathrm{id}}} \mathbb{E}_G\left[\mathrm{Regret}(q^{\pi}) - \Delta\right] \geq \frac{\lambda(b+h)\sqrt{1-\rho}\,\min\{\rho, 1-\rho\}e^{-1/2}}{64\sqrt{N}}$$

(ii) in the knife-edge *regime:*

$$\sup_{G \in \mathcal{G}^{\mathrm{ke}}} \mathbb{E}_G\left[\mathrm{Regret}(q^{\pi}) - \Delta\right] \geq \frac{\lambda(b+h)\sqrt{1-\rho}\,\min\{\rho, 1-\rho\}e^{-1/2}}{32\sqrt{N}}$$

(iii) in the strictly unidentifiable *regime:*

$$\sup_{G \in \mathcal{G}^{\mathrm{ui}}} \mathbb{E}_G\left[\mathrm{Regret}(q^{\pi}) - \Delta\right] \geq \frac{h(M-\lambda)\sqrt{1-\rho}\,\min\{\rho, 1-\rho\}\min\{\rho, 3\rho-1\}e^{-1/2}}{64\sqrt{N}}$$

In the strictly identifiable regime, our bound reflects that the censored setting suffers from the same finite-sample information loss as the uncensored setting (up to constant factors) (Chen and Ma 2024). Moreover, our lower bounds have the same dependence on λ and $M-\lambda$, respectively, for the strictly identifiable and unidentifiable regimes, as the upper bounds derived in Theorem 2. Additionally, the lower bound in the knife-edge regime is exactly twice the lower bound for the strictly identifiable regime. Such a phenomenon also appears in the upper bound, and highlights an additional unavoidable source of loss when $G^{-}(\lambda) \approx \rho$, the regime in which identifiability is the hardest to test.

6. Computational Experiments

In this section we demonstrate the practical efficacy of our algorithm via extensive computational experiments. We test all algorithms on synthetic data in Section 6.2, and consider a real-world retail dataset in Section 6.3.

6.1. Description of Metrics and Benchmark Policies

We evaluate our algorithm, RCN, using $\delta = 0.3$, and benchmark it against five other policies from the literature, for all sets of experiments:

- Naive SAA: This algorithm ignores the potential impact of demand censoring and outputs the ρ-th sample quantile, using all samples in the dataset, i.e.:

$$q^{\mathsf{naive}} = \inf\left\{ q \mid \frac{1}{\sum_k N_k} \sum_k \sum_{i\in[N_k]} \mathbb{1}\left\{s_{ki}^{\text{off}} \le q\right\} \ge \rho \right\}.$$

- Subsample SAA: This algorithm subsamples the dataset to only retain uncensored samples (k,i) such that $d_{ki}^{\text{off}} < q_k^{\text{off}}$, and outputs the ρ-th sample quantile for this subset, i.e.:

$$q^{\mathsf{subsample}} = \inf\left\{ q \mid \frac{1}{|\{(k,i) \mid d_{ki}^{\text{off}} < q_k^{\text{off}}\}|} \sum_k \sum_{i:d_{ki}^{\text{off}}<q_k^{\text{off}}} \mathbb{1}\left\{d_{ki}^{\text{off}} \le q\right\} \ge \rho \right\}.$$

  If $|\{(k,i) \mid d_{ki}^{\text{off}} < q_k^{\text{off}}\}| = 0$, the algorithm outputs λ.

- Kaplan-Meier: This algorithm outputs the ρ-quantile of the Kaplan-Meier (KM) estimator of the cdf (Kaplan and Meier 1958), using the lifelines package (Davidson-Pilon 2024).

- Censored SAA: This is the algorithm designed in Fan et al. (2022) for the identifiable regime, when historical inventory levels are i.i.d. samples from a given distribution.

- True SAA: This is the classical SAA heuristic, which outputs the ρ-th sample quantile with respect to the *true* demand data, i.e.:

$$q^{\mathsf{SAA}} = \inf\left\{ q \mid \frac{1}{\sum_k N_k} \sum_k \sum_{i\in[N_k]} \mathbb{1}\left\{d_{ki}^{\text{off}} \le q\right\} \ge \rho \right\}.$$

  Though this algorithm is not implementable, it serves as a useful benchmark in the identifiable regime (where $q_G^\star$ is learnable), to numerically isolate the impact of demand censoring.

In Section 6.3 we consider a practically motivated variant of RCN, referred to as RCN$^+$, which improves the sample efficiency of RCN by incorporating *all* sales data.

We report the following metrics, depending on the regime of identifiability:

- *Unidentifiable regime*: When $G^-(\lambda) < \rho$, we are interested in all policies' performance relative to the unidentifiable ordering quantity $q_G^\dagger$, which is minimax optimal in this regime. Specifically, we report $\text{Regret}(q^\pi) - \Delta$, and its relative counterpart:

$$\mathcal{R}^{ui}(q^\pi) = \frac{\text{Regret}(q^\pi) - \Delta}{\Delta}. \tag{13}$$

  In this case, reporting $\text{Regret}(q^\pi) - \Delta$ instead of $\text{Regret}(q^\pi)$ allows us to set aside the unavoidable loss that *any* policy incurs in the unidentifiable regime. This metric therefore allows us to quantify various policies' ability to learn the minimax optimal ordering quantity.

- *Identifiable regime*: When $G^-(\lambda) \geq \rho$, we are interested in all policies' performance relative to the optimal newsvendor quantity $q_G^\star$. Specifically, we report the vanilla regret $C_G(q^\pi) - C_G(q_G^\star)$, and its relative counterpart:

$$\mathcal{R}^{id}(q^\pi) = \frac{C_G(q^\pi) - C_G(q_G^\star)}{C_G(q_G^\star)}. \tag{14}$$

REMARK 4. Reporting different metrics depending on the identifiability regime allows us to investigate the robustness of various policies, with the desideratum being low regret both when the optimal newsvendor cost is a meaningful benchmark, as well as when it's not.

6.2. Synthetic Experiments

In this section we investigate how our policy's performance is impacted as we vary (i) the number of samples N, and (ii) the observable boundary λ.

6.2.1. Experimental setup. Across all experiments, we let $h = 1$ and vary $b \in \{3, 9, 49\}$, such that $\rho \in \{0.75, 0.9, 0.98\}$.[5] We moreover consider the following demand distributions:

- *Uniform*: Demand is drawn from a discrete uniform distribution with support $\{0, 1, \ldots, 100\}$.
- *Exponential*: Demand is exponentially distributed with mean 80.
- *Poisson*: Demand is drawn from a Poisson distribution with mean 80.
- *Normal*: Demand $D = \max\{0, X\}$ where $X \in \mathcal{N}(80, \sigma)$ is a Gaussian random variable with mean 80 and standard deviation $\sigma \in \{20, 25, \ldots, 40\}$.

We let $M = 320$, an upper bound on $q_G^\star$ across all of the above distributions.

We consider $K = 2$ historical ordering quantities. To generate these, we first fix the observable boundary λ, and define the larger ordering quantity to be $q_2^{\text{off}} = \lambda$. For each of the above instances, we sample N demand

[5] In most practical settings we expect the per-unit underage cost to be significantly higher than the overage cost.

samples from G, and censor them at λ. We generate the other historical ordering quantity by sampling $q_1^{\text{off}} \sim \text{Uniform}[\frac{\lambda}{4}, \frac{3\lambda}{4}]$. Once we have fixed q_1^{off}, we sample N true demand samples from G, and similarly censor them at q_1^{off}. For each distribution, we vary $\lambda \in \left\{\left(\frac{1}{2} + \frac{k}{7}\right)q_G^\star, k \in [7]\right\}$ (i.e., eight equally spaced points between $\frac{q_G^\star}{2}$ and $\frac{3q_G^\star}{2}$), where $q_G^\star$ is the optimal newsvendor quantity under $\rho = 0.9$. Varying λ in this way allows us to generate enough instances for each regime of identifiability. In particular, we will see that our algorithm's performance depends on whether the problem is *easily identifiable* ($G^-(\lambda) \gg \rho$), *easily unidentifiable* ($G^-(\lambda) \ll \rho$), or in a *difficult regime* ($G^-(\lambda) \approx \rho$).

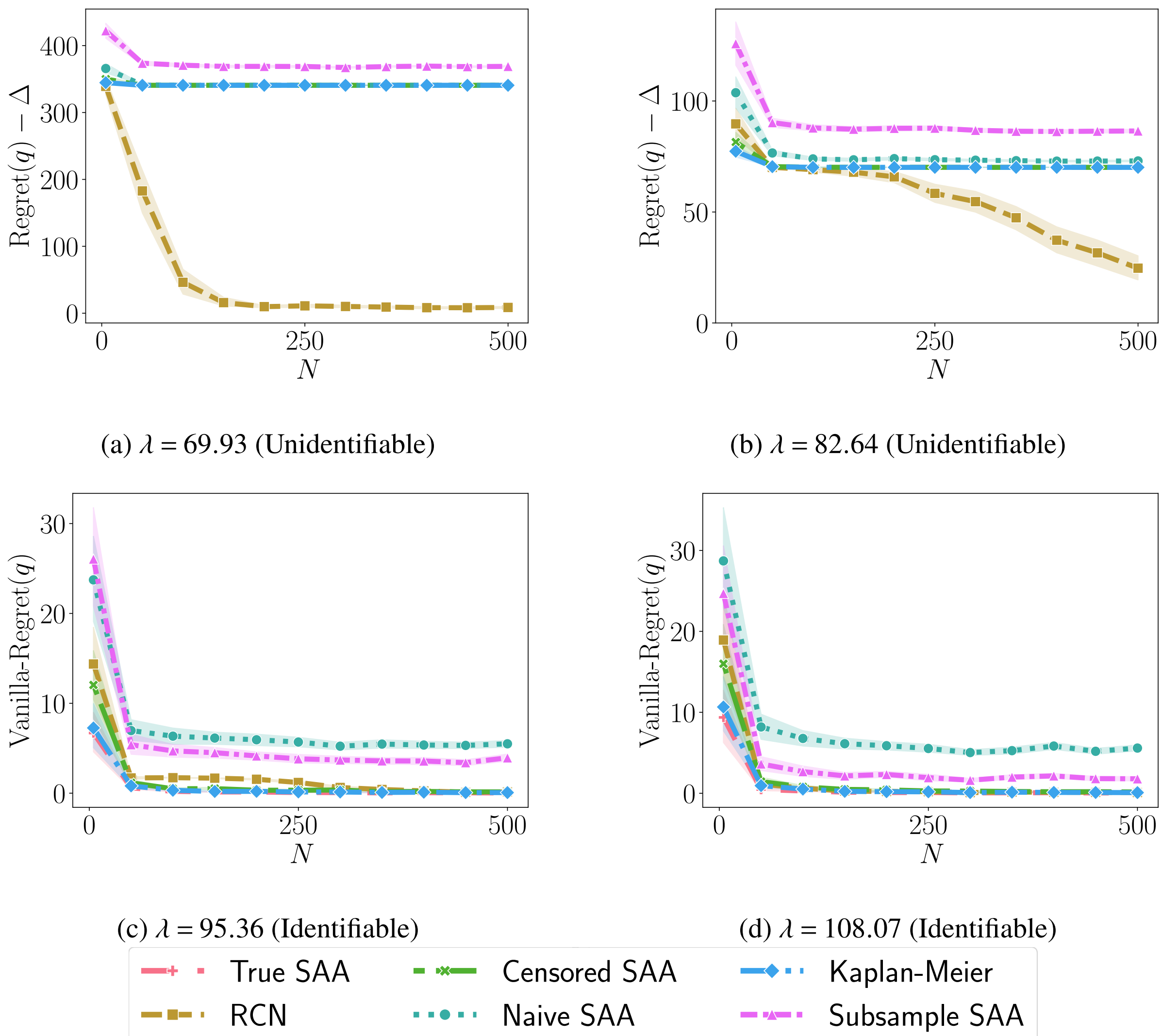
Algorithm evaluation. We run 100 replications for each experiment. To evaluate the algorithms in terms of their additive and relative regret performance we use Monte-Carlo estimation over 10^7 samples from the true underlying demand distribution G.

6.2.2. Impact of N. We first investigate the performance of all policies of interest as N grows large. Here, the underlying demand distribution $D \sim \text{Uniform}\{0, 100\}$ and $\rho = 0.9$, for various values of λ. For this value of ρ, $q_G^\star = 89$; therefore, the problem is unidentifiable for $\lambda \in \{69.93, 82.64\}$ and identifiable for $\lambda \in \{95.36, 108.07\}$. (The results for all other sets of parameters and distributions are entirely analogous. We omit them as such.)

When the problem is unidentifiable, we plot $\text{Regret}(q^\pi) - \Delta$ versus N (see Figures 3a and 3b). RCN is the only algorithm whose worst-case regret asymptotically converges to the minimax risk Δ. We observe that $\text{Regret}(q^\pi) - \Delta$ converges to a constant for all other policies. This is expected, as these latter policies aim to estimate $q_G^\star$, as opposed to attempting to hedge against nature in the unidentifiable regime. Hence, additional samples do not help their performance. Due to censoring at λ, these algorithms often converge to outputting λ, which is why Censored SAA and Kaplan-Meier achieve similar performance.

Notice moreover that our algorithm's convergence to Δ is much slower when $\lambda = 82.64$ compared to when $\lambda = 69.93$. This is because the problem is much closer to identifiability (which would occur at $\lambda \geq q_G^\star = 89$), meaning that RCN requires significantly more samples to test whether or not the problem is identifiable. Specifically, when N is small, we are in the difficult regime where RCN does not have confidence about whether or not the problem is identifiable, and defaults to outputting λ, instead of an estimate of $q_G^\dagger$. With more samples, however, RCN correctly classifies the problem as likely unidentifiable, and outputs $q^{\text{alg}} = q_{\widehat{G}}^\dagger$. Despite this burn-in period, across all values of λ in the unidentifiable regime RCN ensures a 5% relative optimality gap $\mathcal{R}^{ui}(q^{\text{alg}})$ in roughly $N \sim 150$ samples.

When the problem is identifiable, given that the optimal newsvendor cost is indeed achievable in the worst case, we plot Vanilla-Regret versus N (see Figures 3c and 3d). We observe that RCN, Kaplan-Meier, Censored SAA, and True SAA all quickly converge to the optimal newsvendor quantity $q_G^\star$, resulting in vanishing vanilla regret. On the other hand, the performance of Censored SAA and Naive SAA plateaus as N grows large. This again is expected since these policies learn biased estimates of the underlying demand



(a) $\lambda = 69.93$ (Unidentifiable)

(b) $\lambda = 82.64$ (Unidentifiable)

(c) $\lambda = 95.36$ (Identifiable)

(d) $\lambda = 108.07$ (Identifiable)

Figure 3 **Regret$(q) - \Delta$ vs. N (when unidentifiable) and Vanilla-Regret(q) vs. N (when identifiable) for the different benchmark policies, for $\lambda \in \{69.93, 82.64, 95.36, 108.07\}$. Here, $D \sim$ Uniform$\{0, 100\}$ and $\rho = 0.9$, with $q^\star_G = 89$.**

distribution. Note that, for $\lambda = 95.36$, while our algorithm's regret is very low, it lies slightly above that of the three other policies for small values of N. This is the same phenomenon as the one observed when $\lambda = 82.64$: our policy requires a burn-in period when λ is close to $q^\star_G$ to determine identifiability.

6.2.3. Impact of observable boundary. We next study the impact of the observable boundary λ on policy performance. Intuitively, since Regret$(q^\pi) - \Delta$ represents the loss due to errors in estimating q^Δ, RCN should exhibit the best performance for extreme values of λ, when the problem is *easily unidentifiable* ($\lambda \ll q^\star_G$) or *easily identifiable* ($\lambda \gg q^\star_G$), since these regimes are where the algorithm will make the fewest identifiability errors. As λ approaches $q^\star_G$ from either direction, however, one would expect the frequency with which RCN either misclassifies the regime or defaults to λ to increase, leading to higher values of $\text{Regret}(q^\pi) - \Delta$.

To test this hypothesis, we fix $N = 500$ and $\rho = 0.9$, and consider the uniform, exponential, and Poisson distributions described in Section 6.1. Our results are summarized in Table 1; the thick vertical line in each table marks the transition from the unidentifiable regime, where we report $\text{Regret}(q^\pi) - \Delta$ and $\mathcal{R}^{ui}(q^\pi)$ in parentheses, to the identifiable regime, where we report Vanilla-Regret(q^π) and $\mathcal{R}^{id}(q^\pi)$ in parentheses. Results for other values of ρ are provided in Appendix A.2.1.

λ	44.50	57.21	69.93	82.64	95.36	108.07	120.79	133.50
True SAA					0.0 (0.1%)	0.0 (0.1%)	0.0 (0.0%)	0.0 (0.0%)
RCN	4.0 (1.7%)	6.7 (3.3%)	7.7 (4.5%)	27 (27%)	0.1 (0.2%)	0.1 (0.2%)	0.1 (0.1%)	0.1 (0.2%)
Censored SAA	1033 (450%)	653 (320%)	341 (200%)	70 (70%)	0.1 (0.3%)	0.1 (0.2%)	0.1 (0.2%)	0.1 (0.3%)
Kaplan-Meier	1033 (450%)	653 (320%)	341 (200%)	70 (70%)	0.1 (0.1%)	0.1 (0.1%)	0.0 (0.1%)	0.1 (0.1%)
Naive SAA	1033 (450%)	653 (320%)	341 (200%)	73 (73%)	5.0 (11%)	5.4 (12%)	4.5 (9.9%)	4.3 (9.6%)
Subsample SAA	1067 (465%)	684 (335%)	369 (217%)	87 (87%)	3.5 (7.7%)	1.9 (4.2%)	2.0 (4.5%)	1.8 (4.1%)

(a) Uniform, $q_G^\star = 89$

λ	92.07	118.38	144.68	170.99	197.30	223.60	249.91	276.22
True SAA					0.4 (0.2%)	0.4 (0.2%)	0.3 (0.2%)	0.3 (0.2%)
RCN	6.7 (4.2%)	6.9 (6.0%)	21 (29%)	4.3 (19%)	1.0 (0.5%)	7.4 (4.0%)	3.9 (2.1%)	0.8 (0.5%)
Censored SAA	345 (216%)	148 (128%)	45 (64%)	4.6 (20%)	0.6 (0.3%)	0.9 (0.5%)	0.8 (0.4%)	0.6 (0.3%)
Kaplan-Meier	345 (216%)	148 (128%)	45 (64%)	4.5 (19%)	0.6 (0.3%)	0.6 (0.3%)	0.5 (0.3%)	0.5 (0.3%)
Naive SAA	345 (216%)	148 (128%)	57 (82%)	29 (125%)	25 (13%)	20 (11%)	18 (9.8%)	15 (8.1%)
Subsample SAA	410 (257%)	214 (185%)	108 (153%)	57 (243%)	40 (22%)	31 (17%)	25 (14%)	20 (11%)

(b) Exponential, $q_G^\star = 184.21$

λ	46	59.14	72.29	85.43	98.57	111.71	124.86	138
True SAA					0.0 (0.2%)	0.0 (0.1%)	0.0 (0.2%)	0.0 (0.2%)
RCN	0.0 (0.0%)	0.5 (0.2%)	2.4 (1.1%)	7.1 (4.7%)	0.1 (0.3%)	0.0 (0.3%)	0.1 (0.3%)	0.1 (0.4%)
Censored SAA	2260 (900%)	2131 (891%)	1542 (697%)	247 (165%)	0.1 (0.4%)	0.1 (0.3%)	0.1 (0.4%)	0.1 (0.4%)
Kaplan-Meier	2260 (900%)	2131 (891%)	1542 (697%)	247 (165%)	0.1 (0.3%)	0.0 (0.3%)	0.1 (0.3%)	0.1 (0.4%)
Naive SAA	2260 (900%)	2131 (891%)	1542 (697%)	247 (165%)	2.0 (12%)	2.0 (12%)	1.9 (12%)	1.6 (9.8%)
Subsample SAA	2260 (900%)	2138 (894%)	1544 (699%)	250 (168%)	0.2 (1.2%)	0.1 (0.7%)	0.3 (1.7%)	0.3 (1.7%)

(c) Poisson, $q_G^\star = 92$

Table 1 Impact of λ on policy performance. Values to the left of the thick vertical line correspond to the unidentifiable regime, where we report Regret$(q^\pi) - \Delta$ and $\mathcal{R}^{ui}(q^\pi)\%$; values to the right of the thick vertical line correspond to the identifiable regime, where we report Vanilla-Regret(q^π) and $\mathcal{R}^{id}(q^\pi)\%$.

When the problem is *easily unidentifiable* (i.e., $\lambda \ll q_G^\star$), RCN is the only near-optimal algorithm, incurring regret of at most 6% relative to the minimax risk Δ. (This occurs in the case of the exponential distribution, for $\lambda = 118.38$; see Table 1b.) While our algorithm approximates the unidentifiable quantile $q_G^\dagger$, the non-robust algorithms incur regret that is at least twice as high as Δ in the best case, and 10 times as high in the worst case. Observe that the performance of these algorithms is near-identical across many instances. This is because, for small values of λ, any SAA-style solution outputs a quantity close to λ, since most demand realizations are censored.

When the problem is *easily identifiable* (i.e. $\lambda \gg q^\star_G$), RCN remains near-optimal: in the worst case, its cost is within 4% of the true newsvendor optimum (occurring at $\lambda = 223.60$ in Table 1b). Censored SAA and Kaplan-Meier similarly perform exceptionally well, with a relative vanilla regret of at most 1% in all cases. Again, this highlights that censoring is essentially immaterial in this regime. Finally, Subsample SAA and Naive SAA retain their poor performance across most instances. This performance slightly improves for larger values of λ, since fewer demand samples are censored as λ increases.

Finally, we note that across all three tables, as λ approaches $q^\star_G$ from below, RCN incurs higher regret. This is in line with our intuition that determining identifiability becomes more challenging close to the identifiability boundary $\lambda = q^\star_G$, with RCN making more misclassification errors. The exception to this trend is the exponential distribution (Table 1b), where we observe the algorithm's regret steeply decreasing at $\lambda = 170.99$. For this instance, RCN classifies the problem as "knife-edge" in 99% of replications, thereby defaulting to outputting λ. However, for this value of λ, $q^\dagger_G = 184.97$; this explains the algorithm's strong performance in the knife-edge regime.

Overall, our results illustrate the robustness of RCN to varying values of λ, with strong performance relative to the true newsvendor cost when the problem is identifiable. Moreover, our experiments show that the benefits of our algorithm are the largest in highly censored settings.

6.3. Real-World Dataset

Finally, we complement our synthetic experiments by demonstrating our algorithm's robust performance on a real-world grocery retail dataset (Wang et al. 2025).

6.3.1. Dataset description and pre-processing. The original dataset consists of three months of detailed hourly sales data for 863 distinct perishable products across 898 grocery stores in China.[6] While the dataset does not specify the number of units in stock at the beginning of each day, it indicates whether an out-of-stock event occurred for each (product, hour) tuple. We use this out-of-stock indicator to reconstruct historical inventory levels as follows.

For each product p in the dataset, we first filter out all days that began with an out-of-stock event. For each remaining day t, we reconstruct the initial inventory level of the product, denoted by I^{off}_{pt}, by first considering the set of all days in which a stockout event occurred. We use $\mathcal{T}_p$ to denote this set. Then, for all $t \in \mathcal{T}_p$, I^{off}_{pt} corresponds to the total sales recorded up until the stockout event. Now, for all $t \in \mathcal{T}_p^c$, we

[6] In the remainder of this section, we abuse terminology and refer to each (product, store) tuple as a product.

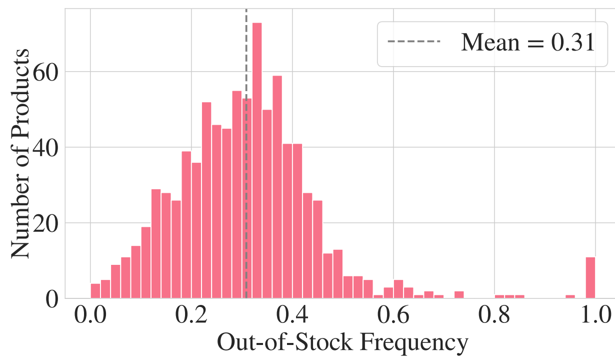


Figure 4 **Distribution of out-of-stock frequency.**

let $I_{pt}^{\text{off}} = \max\left\{ \min_{\substack{t' \in \mathcal{T}_p: \\ I_{pt'}^{\text{off}} > s_{pt}^{\text{off}}}} I_{pt'}^{\text{off}}, s_{pt}^{\text{off}} + 1 \right\}$, where s_{pt}^{off} is used to denote the recorded sales on day t. In words, on non-stockout days we let the inventory level be the smallest known inventory level (i.e., across all stockout days); if the recorded sales exceeds all known inventory levels, we assume the inventory level exceeded the recorded sales by 1. We use this data to construct K_p selling seasons for product p, where each selling season corresponds to days t for which the initial inventory levels where identical. With this setup in hand, in the remainder of the section we adopt the same notation as before, letting N_{pk} denote the number of days associated with selling season k of product p, with s_{pki}^{off} and q_{pki}^{off} respectively denoting the observed sales and historical ordering quantities for day $i \in [N_{pk}]$. Finally, let $\mathcal{P}$ be the set of all products in our pre-processed dataset.

Figure 4 shows the distribution of out-of-stock frequency (i.e., the percentage of days in which demand was censored) in our dataset. On average, products stocked out on 31% of days, with 25% of products stocking out at least 40% of the time. These results highlight that *demand censoring is a common occurrence in practice*, and underscores the importance of algorithms that are robust to high levels of censoring.

Training and testing datasets. The dataset provided by Wang et al. (2025) is randomly partitioned into training and evaluation sets, for each product $p \in \mathcal{P}$. These sets respectively comprise 2,605,058 and 215,978 samples (i.e., a 92-8 split), where each sample corresponds to a $(s_{pki}^{\text{off}}, q_{pki}^{\text{off}})$ pair. Only the training set is used by the algorithms; the testing set is exclusively used for evaluation. We highlight that there is no notion of *true* demand distribution G_p that the decision-maker could use to evaluate the benchmark algorithms, given that there is no data past the observable boundary of the testing set. Indeed, this is precisely the obstacle that motivates the introduction of the distributionally robust optimization framework for this problem. Given this issue, we use our testing set to construct a partial demand distribution G_p *up until the observable boundary* λ_p. This partial distribution, constructed using the Kaplan-Meier estimator, is used to determine identifiability for each tested value of ρ; we also report the worst-case regret of each algorithm using this partial cdf. After this construction, we observe that the average value of $G^{-}(\lambda)$ across products is 0.83, with only 66% of products such that $G^-(\lambda) \geq 0.9$. Given this, we expect non-robust algorithms to perform particularly poorly in settings where b is much larger than h, as is frequently the case in practice, and assumed in the literature (Huh and Rusmevichientong 2009)).

6.3.2. Algorithms. Despite the strong theoretical guarantees of RCN, our algorithm discards potentially valuable data by exclusively using sales data where the historical ordering quantity is equal to λ. This issue is particularly salient in our dataset, where the vast majority of products have fewer than 10 samples at the boundary, with the average sample size at the boundary less than 4, and only 1% of products having $N_K \geq 20$. In contrast, the sample size increases significantly as we move away from the boundary. In particular, by the 10th-highest historical selling season, the number of samples exceeds 75; it reaches 200 by the 15th-highest historical selling season. These observations underscore the potential value of using the entirety of the offline dataset.

We address this sample inefficiency by proposing RCN$^+$, an extension of RCN that leverages *all* sales data. Before presenting the modified algorithm, we provide some intuition behind our approach.[7] Suppose there exists historical ordering quantity $q_k^{\text{off}} < \lambda$ such that $G^-(q_k^{\text{off}}) \geq \rho$. Then, $G^-(\lambda) \geq \rho$, which implies that the underlying problem is identifiable, with minimax optimal ordering quantity $q^\Delta = q_G^\star$, by Theorem 1. Our modified algorithm leverages this insight to improve the sample efficiency of RCN. Specifically, the first stage of our algorithm can be viewed through the lens of multiple hypothesis testing. For each $k \in [K]$, we test whether $G^-(q_k^{\text{off}}) \geq \rho$. If any k passes this test, the problem is classified as *likely identifiable*, and we compute the censored SAA of $q_G^\star$ using *all* of the data that has satisfied this test. If there doesn't exist a value of k for which we estimate that $G^-(q_k^{\text{off}}) \geq \rho$, then we may be in the unidentifiable regime, and we proceed as in RCN. We provide a formal description of this practical extension in Algorithm 2, and defer its theoretical analysis to Appendix E.

Finally, given the poor performance of Naive SAA and Subsample SAA in our synthetic experiments, for this set of experiments we only consider Kaplan-Meier and Censored SAA as benchmark algorithms.

6.3.3. Results. In all experiments we let $M_p = 2.5\lambda_p$ for all $p \in \mathcal{P}$. We moreover let $h = 1$ and vary $b \in \{3, 9, 49\}$. For these respective values of b, 83%, 66%, and 49% of products are identifiable.

In Figure 5 we plot $\text{Regret}(q) - \Delta$ for all four algorithms and each value of b, across both identifiable and unidentifiable instances. (Recall, since $\Delta = 0$ in the identifiable regime, this reduces to reporting the worst-case regret $\text{Regret}(q)$ for these instances.) As observed in our synthetic experiments, in the identifiable regime (Figure 5a) Censored SAA and Kaplan-Meier are the best-performing algorithms, with $\text{Regret}(q) \leq 1.3$ on average, across all values of b. RCN, however, achieves slightly higher regret, with $\text{Regret}(q) \in \{3.12, 4.15, 5.47\}$ on average. The reason for this is two-fold. On the one hand, if RCN classifies the instance as likely identifiable, it only uses samples at the boundary, of which there are very

[7] For notational simplicity we omit all quantities' dependence on the product p in the remainder of this subsection.

ALGORITHM 2: Robust Censored Newsvendor$^+$ (RCN$^+$)

Input: Dataset s^{off}, confidence terms ζ_k for $k \in [K]$, $\delta > 0$

Output: Ordering quantity q^{alg}

```
// Conduct multiple hypothesis test of identifiability
```

For all $k \in [K]$, compute censored SAA of $G^-(q_k^{\text{off}})$:

$$\widehat{G}_k^-(q_k^{\text{off}}) = \frac{1}{N_k} \sum_{i \in [N_k]} \mathbb{1}\left\{s_{ki}^{\text{off}} < q_k^{\text{off}}\right\}. \tag{15}$$

Construct "likely identifiable" set $\mathcal{U}_{\text{est}}$, defined as:

$$\mathcal{U}_{\text{est}} = \left\{k \in [K] \mid \widehat{G}_k^-(q_k^{\text{off}}) \geq \rho + \zeta_k\right\} \tag{16}$$

if $\mathcal{U}_{est} \neq \emptyset$ **then** `// Likely identifiable`

Compute censored SAA of $q_G^\star$ using all samples associated with $\mathcal{U}_{\text{est}}$:

$$q^{\text{alg}} = \inf\left\{x \mid \frac{1}{\sum_{k \in \mathcal{U}_{\text{est}}} N_k} \sum_{k \in \mathcal{U}_{\text{est}}} \sum_{i \in [N_k]} \mathbb{1}\left\{s_{ki}^{\text{off}} \leq x\right\} \geq \rho\right\}. \tag{17}$$

else if $\widehat{G}_k^-(q_k^{\text{off}}) < \rho - \zeta_k$ *for all* $k \in [K]$ **then** `// Likely unidentifiable`

Compute empirical estimate of $q_G^\dagger$:

$$q^{\text{alg}} = \frac{bM + h\lambda - (b+h)\widehat{G}_K^-(\lambda)M}{(b+h)(1 - \widehat{G}_K^-(\lambda))}. \tag{18}$$

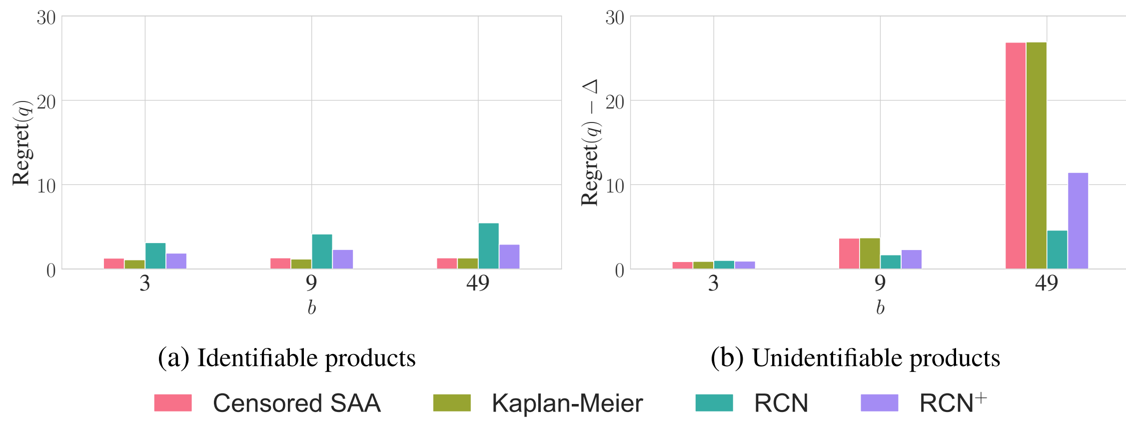
else `// Knife-edge case`

$q^{\text{alg}} = \lambda$.

return q^{alg}

few; this results in a higher estimation error of $q_G^\star$. On the other hand, RCN struggles to determine identifiability with confidence as b grows larger and the gap between ρ and $G^-(\lambda)$ decreases; this results in more misclassification errors, where it estimates either $q_G^\dagger$ or simply outputs λ. Our results show that RCN$^+$ mitigates these issues, with $\text{Regret}(q) \in \{1.87, 2.31, 2.93\}$, on average. (We highlight that one cannot expect RCN$^+$ to achieve the same guarantees as Censored SAA and Kaplan-Meier, given that it too incurs misclassification errors in testing for identifiability.)

Figure 5b, on the other hand, demonstrates that both of our algorithms generate substantial gains relative to non-robust algorithms in the unidentifiable regime, as the lost-sales penalty grows large. In the worst case ($b = 49$), Censored SAA and Kaplan-Meier incur $\text{Regret}(q) - \Delta \geq 26$ on average. In contrast, RCN and RCN$^+$ incur $\text{Regret}(q) - \Delta$ of 4.6 and 11.4, respectively. This latter observation is a priori surprising, given that RCN$^+$ and RCN both output an estimate of $q_G^\dagger$ which only uses the samples at the boundary. The reason this occurs is that, for each product, the confidence terms ζ_k used by RCN$^+$ are scaled by an additional factor of $\sqrt{\log K}$ in order to guarantee $|\widehat{G}_k^-(q_k^{\text{off}}) - G^-(q_k^{\text{off}})| \leq \zeta_k$ with high probability (we refer the reader to Appendix E for these arguments). Hence, the "likely unidentifiable" condition in Algorithm 2



(a) Identifiable products

(b) Unidentifiable products

Figure 5 **Regret(q) − Δ as a function of b, for both identifiable and unidentifiable products.**

b	Kaplan-Meier	Censored SAA	RCN	RCN$^+$
3	0.038	-0.028	-0.050	0.028
9	0.196 (**)	0.171 (**)	-0.091 (**)	0.042
49	0.257 (**)	0.257 (**)	-0.013	0.057

Table 2 **Correlation between Regret(q) − Δ and historical out-of-stock frequency. Here, (**) represents statistically significant results (p-value < 0.05).**

is more difficult to satisfy, and RCN$^+$ outputs λ more frequently. We conjecture however that these artifacts of our theoretical analysis can be eliminated with appropriate tuning of the confidence terms. Finally, we observe that when b is small, even the non-robust algorithms incur low regret. The reason for this is that, even though $\lambda = 1.68 < q_G^\dagger = 2.55$ on average, lost sales are not as costly. Hence, this regime is "easy" for any reasonable (non-robust) algorithm.

Finally, we explore the robustness of our algorithms to historical out-of-stock frequency, a practical, interpretable proxy for regime difficulty. Specifically, in Table 2 we report the Pearson correlation between Regret(q) − Δ and out-of-stock frequency. Our findings corroborate Figure 5. Namely, when $b = 3$, all algorithms perform well regardless of whether the product is identifiable; hence, there is no significant correlation between out-of-stock frequency and performance. However, for $b \in \{9, 49\}$, Kaplan-Meier and Censored SAA are the only algorithms whose regret is positively correlated with out-of-stock frequency. In contrast, RCN$^+$ does not exhibit a significant correlation, highlighting its robustness to demand censoring. Notably, the performance of RCN is significantly *negatively* correlated with out-of-stock frequency: this is due to the decrease in Regret(q) − Δ observed between Figure 5a and Figure 5b, as misclassification errors (where RCN outputs λ instead of an estimate of $q_G^\dagger$) are less costly in this regime.

7. Conclusion

In this work we considered the offline data-driven censored newsvendor problem. Specifically, we leveraged the distributionally robust optimization framework to isolate the impact of demand censoring on newsvendor decision-making. Through this framework we established crisp characterizations of the information loss due to censoring, and used these to design a robust algorithm with near-optimal performance guarantees across all possible demand censoring levels. Our algorithm is practical and intuitive, and demonstrates strong numerical performance on both synthetic and real-world datasets.

Several future directions emerge from our work. Most immediately, we conjecture that all of our analytical and algorithmic techniques go through under common assumptions on the demand distribution, as noted in Remark 3. It would moreover be interesting to see if our framework and insights port over to the *feature-based* setting, in which the decision-maker also has access to uncensored contextual information that may help mitigate the effect of demand censoring. Lastly, it would be interesting to extend our modeling framework to the setting with inventory carry-over, where demand may be endogenous to previous inventory levels.

References

Agrawal N, Smith SA (1996) Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics (NRL)* 43(6):839–861.

Agrawal S, Jia R (2022) Learning in structured mdps with convex cost functions: Improved regret bounds for inventory management. *Operations Research* 70(3):1646–1664.

Ban GY (2020) Confidence intervals for data-driven inventory policies with demand censoring. *Operations Research* 68(2):309–326.

Besbes O, Ma W, Mouchtaki O (2022) Beyond iid: data-driven decision-making in heterogeneous environments. *Advances in Neural Information Processing Systems* 35:23979–23991.

Besbes O, Mouchtaki O (2023) How big should your data really be? data-driven newsvendor: Learning one sample at a time. *Management Science* 69(10):5848–5865.

Besbes O, Muharremoglu A (2013) On implications of demand censoring in the newsvendor problem. *Management Science* 59(6):1407–1424.

Boucheron S, Lugosi G, Massart P (2013) *Concentration Inequalities: A Nonasymptotic Theory of Independence* (Oxford University Press).

Bu J, Simchi-Levi D, Wang L (2023) Offline pricing and demand learning with censored data. *Management Science* 69(2):885–903.

Chen B, Jiang J, Zhang J, Zhou Z (2024) Learning to order for inventory systems with lost sales and uncertain supplies. *Management Science* .

Chen Z, Ma W (2024) Survey of data-driven newsvendor: Unified analysis and spectrum of achievable regrets. *arXiv preprint arXiv:2409.03505* .

Cheung WC, Simchi-Levi D (2019) Sampling-based approximation schemes for capacitated stochastic inventory control models. *Mathematics of Operations Research* 44(2):668–692.

Davidson-Pilon C (2024) lifelines, survival analysis in python.

Duchi J (2024) Lecture notes on statistics and information theory.

Fan X, Chen B, Lennon Olsen T, Pinedo M, Qin H, Zhou Z (2024) Don't follow RL blindly: Lower sample complexity of learning optimal inventory control policies with fixed ordering costs. *Available at SSRN 4828001* .

Fan X, Chen B, Zhou Z (2022) Sample complexity of policy learning for inventory control with censored demand. *Available at SSRN 4178567* .

Fan X, Chen B, Zhou Z (2025) Sample complexity of policy learning for inventory control with censored demand. *Available at SSRN 4178567* .

Fu M, Li X, Zhang L (2024) Distributionally robust newsvendor under stochastic dominance with a feature-based application. *Manufacturing & Service Operations Management* 26(5):1962–1977.

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

Godfrey GA, Powell WB (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.

Huh WT, Levi R, Rusmevichientong P, Orlin JB (2011) Adaptive data-driven inventory control with censored demand based on kaplan-meier estimator. *Operations Research* 59(4):929–941.

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

Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53(282):457–481.

Lattimore T, Szepesvári C (2020) *Bandit algorithms* (Cambridge University Press).

Lee S, Kim H, Moon I (2021) A data-driven distributionally robust newsvendor model with a wasserstein ambiguity set. *Journal of the Operational Research Society* 72(8):1879–1897.

Levi R, Perakis G, Uichanco J (2015) The data-driven newsvendor problem: New bounds and insights. *Operations Research* 63(6):1294–1306.

Levi R, Roundy RO, Shmoys DB (2007) Provably near-optimal sampling-based policies for stochastic inventory control models. *Mathematics of Operations Research* 32(4):821–839.

Lin M, Huh WT, Krishnan H, Uichanco J (2022) Data-driven newsvendor problem: Performance of the sample average approximation. *Operations Research* 70(4):1996–2012.

Lugosi G, Markakis MG, Neu G (2024) On the hardness of learning from censored and nonstationary demand. *INFORMS Journal on Optimization* 6(2):63–83.

Massart P (1990) The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *The Annals of Probability* 1269–1283.

Mersereau AJ (2015) Demand estimation from censored observations with inventory record inaccuracy. *Manufacturing & Service Operations Management* 17(3):335–349.

Nahmias S (1994) Demand estimation in lost sales inventory systems. *Naval Research Logistics (NRL)* 41(6):739–757.

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

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

Sason I (2015) On reverse pinsker inequalities. *arXiv preprint arXiv:1503.07118* .

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

Shi C, Chen W, Duenyas I (2016) Nonparametric data-driven algorithms for multiproduct inventory systems with censored demand. *Operations Research* 64(2):362–370.

Vulcano G, Van Ryzin G, Ratliff R (2012) Estimating primary demand for substitutable products from sales transaction data. *Operations Research* 60(2):313–334.

Wang Y, Gu J, Long L, Li X, Shen L, Fu Z, Zhou X, Jiang X (2025) Freshretailnet-50k: A stockout-annotated censored demand dataset for latent demand recovery and forecasting in fresh retail. *arXiv preprint arXiv:2505.16319* .

Xie Y, Ma W, Xin L (2024) Vc theory for inventory policies. *arXiv preprint arXiv:2404.11509* .

Xu L, Zheng Y, Jiang L (2022) A robust data-driven approach for the newsvendor problem with nonparametric information. *Manufacturing & Service Operations Management* 24(1):504–523.

Yuan H, Luo Q, Shi C (2021) Marrying stochastic gradient descent with bandits: Learning algorithms for inventory systems with fixed costs. *Management Science* 67(10):6089–6115.

Zipkin P (2000) *Foundations of inventory management* (Maidenhead, England: Irwin Professional Publishing).

λ	44.50	57.21	69.93	82.64	95.36	108.07	120.79	133.50
True SAA				0.0 (0.0%)	0.0 (0.1%)	0.0 (0.1%)	0.0 (0.1%)	0.0 (0.1%)
RCN	4.5 (3.0%)	5.4 (5.0%)	11 (26%)	0.2 (0.4%)	0.1 (0.2%)	0.1 (0.2%)	0.1 (0.1%)	0.1 (0.2%)
Censored SAA	184 (120%)	74 (68%)	8.8 (21%)	0.1 (0.2%)	0.1 (0.1%)	0.1 (0.2%)	0.1 (0.1%)	0.0 (0.1%)
Kaplan-Meier	184 (120%)	74 (68%)	8.8 (21%)	0.0 (0.1%)	0.0 (0.1%)	0.1 (0.1%)	0.0 (0.1%)	0.0 (0.1%)
Naive SAA	184 (120%)	80 (74%)	21 (49%)	12 (31%)	9.0 (24%)	7.3 (19%)	5.6 (15%)	4.8 (13%)
Subsample SAA	209 (137%)	97 (89%)	27 (63%)	10.0 (27%)	3.9 (10%)	2.7 (7.2%)	2.3 (6.2%)	2.3 (6.2%)

(a) Uniform, $q_G^\star = 74$

λ	92.07	118.38	144.68	170.99	197.30	223.60	249.91	276.22
True SAA		0.1 (0.1%)	0.1 (0.1%)	0.1 (0.1%)	0.1 (0.1%)	0.1 (0.1%)	0.1 (0.1%)	0.1 (0.1%)
RCN	11 (23%)	0.4 (0.3%)	0.7 (0.7%)	0.3 (0.3%)	0.2 (0.2%)	0.3 (0.3%)	0.2 (0.2%)	0.2 (0.2%)
Censored SAA	13 (27%)	0.2 (0.2%)	0.2 (0.2%)	0.2 (0.2%)	0.2 (0.2%)	0.1 (0.1%)	0.2 (0.2%)	0.1 (0.1%)
Naive SAA	34 (70%)	19 (17%)	13 (12%)	7.6 (6.9%)	7.5 (6.7%)	5.5 (5.0%)	3.3 (2.9%)	1.4 (1.3%)
Kaplan-Meier	13 (27%)	0.2 (0.2%)	0.2 (0.2%)	0.2 (0.2%)	0.2 (0.1%)	0.2 (0.1%)	0.2 (0.1%)	0.1 (0.1%)
Subsample SAA	46 (95%)	26 (23%)	18 (17%)	13 (12%)	11 (9.7%)	8.3 (7.5%)	6.4 (5.7%)	4.5 (4.1%)

(b) Exponential, $q_G^\star = 110.90$

λ	46	59.14	72.29	85.43	98.57	111.71	124.86	138
True SAA					0.0 (0.1%)	0.0 (0.2%)	0.0 (0.1%)	0.0 (0.2%)
RCN	0.0 (0.0%)	0.5 (0.2%)	2.4 (1.4%)	1.0 (6.8%)	0.0 (0.2%)	0.0 (0.3%)	0.0 (0.3%)	0.0 (0.3%)
Censored SAA	628 (300%)	590 (297%)	380 (219%)	0.6 (4.4%)	0.0 (0.2%)	0.0 (0.3%)	0.0 (0.2%)	0.0 (0.2%)
Kaplan-Meier	628 (300%)	590 (297%)	380 (219%)	0.6 (4.4%)	0.0 (0.2%)	0.0 (0.3%)	0.0 (0.2%)	0.0 (0.3%)
Naive SAA	628 (300%)	590 (297%)	380 (219%)	3.6 (26%)	3.0 (26%)	2.8 (25%)	2.2 (19%)	2.1 (18%)
Subsample SAA	628 (300%)	593 (298%)	383 (221%)	2.6 (18%)	0.1 (0.5%)	0.1 (1.1%)	0.2 (1.6%)	0.2 (1.5%)

(c) Poisson, $q_G^\star = 86$

Table 3 Impact of λ on policy performance. Here, $\rho = 0.75$. Values to the left of the thick vertical line correspond to the unidentifiable regime, where we report Regret$(q^\pi) - \Delta$ and $\mathcal{R}^{ui}(q^\pi)$%; values to the right of the thick vertical line correspond to the identifiable regime, where we report Vanilla-Regret(q^π) and $\mathcal{R}^{id}(q^\pi)$%.

Appendix A: Computational Experiments: Additional Details

A.1. Simulation Information

Computing Infrastructure: The experiments were conducted on a personal computer with an Apple M2, 8-core processor and 16.0GB of RAM.

A.2. Synthetic Experiments: Supplemental Results

A.2.1. Impact of the Observable Boundary. Tables 3 and 4 show the impact of λ on policy performance, for $\rho \in \{0.75, 0.98\}$. Trends are identical to those identified for $\rho = 0.9$. Note that, for $\rho = 0.98$, all non-robust benchmarks exhibit extremely high relative regret for small values of λ, as the worst-case distribution forces these policies to incur exorbitant underage costs in the unidentifiable regime.

A.2.2. Vanilla Regret in the Unidentifiable Regime. While RCN achieves the optimal worst-case regret guarantee in the unidentifiable regime, a natural question is whether its strong performance extends to *vanilla regret* (see Eq. (3)) in this regime, evaluated with respect to the true distribution G. In this section we provide numerical support for the idea that vanilla regret is not the meaningful regret metric in the unidentifiable regime. In particular, our results will illustrate the intuitive fact that the performance of all

λ	44.50	57.21	69.93	82.64	95.36	108.07	120.79	133.50
True SAA						0.0 (0.0%)	0.0 (0.0%)	0.0 (0.1%)
RCN	4.7 (1.7%)	6.9 (2.7%)	6.9 (2.9%)	11 (5.4%)	115 (100%)	0.0 (0.1%)	0.1 (0.1%)	0.1 (0.3%)
Censored SAA	7165 (2651%)	5102 (2001%)	3335 (1401%)	1605 (751%)	116 (101%)	1.2 (2.4%)	0.9 (1.9%)	2.5 (5.1%)
Kaplan-Meier	7165 (2651%)	5102 (2001%)	3335 (1401%)	1605 (751%)	116 (101%)	0.0 (0.1%)	0.0 (0.1%)	0.1 (0.2%)
Naive SAA	7165 (2651%)	5102 (2001%)	3335 (1401%)	1605 (751%)	116 (101%)	1.2 (2.4%)	1.3 (2.7%)	1.3 (2.7%)
Subsample SAA	7207 (2666%)	5135 (2014%)	3377 (1419%)	1629 (762%)	121 (105%)	0.4 (0.9%)	0.7 (1.3%)	0.7 (1.5%)

(a) Uniform, $q_G^\star = 97$

λ	92.07	118.38	144.68	170.99	197.30	223.60	249.91	276.22
True SAA								
RCN	6.2 (2.8%)	7.6 (4.0%)	7.6 (4.8%)	7.7 (6.0%)	69 (71%)	111 (163%)	48 (118%)	11 (59%)
Censored SAA	3232 (1482%)	1956 (1038%)	1138 (720%)	626 (490%)	316 (324%)	140 (206%)	51 (124%)	13 (71%)
Naive SAA	3232 (1482%)	1956 (1038%)	1138 (720%)	626 (490%)	316 (324%)	140 (206%)	55 (136%)	30 (169%)
Kaplan-Meier	3232 (1482%)	1956 (1038%)	1138 (720%)	626 (490%)	316 (324%)	140 (206%)	49 (120%)	11 (61%)
Subsample SAA	3326 (1525%)	2053 (1090%)	1246 (788%)	737 (576%)	424 (435%)	246 (362%)	144 (352%)	101 (564%)

(b) Exponential, $q_G^\star = 312.96$

λ	46	59.14	72.29	85.43	98.57	111.71	124.86	138
True SAA						0.1 (0.6%)	0.2 (0.7%)	0.1 (0.5%)
RCN	0.0 (0.0%)	0.5 (0.2%)	2.5 (1.0%)	8.1 (3.7%)	3.1 (16%)	0.2 (1.0%)	0.2 (1.1%)	0.2 (1.0%)
Censored SAA	13397 (4900%)	12652 (4857%)	9578 (3887%)	2717 (1226%)	3.3 (16%)	0.6 (2.5%)	0.9 (4.2%)	1.9 (8.4%)
Kaplan-Meier	13397 (4900%)	12652 (4857%)	9578 (3887%)	2717 (1226%)	3.2 (16%)	0.2 (1.0%)	0.2 (1.1%)	0.2 (0.9%)
Naive SAA	13397 (4900%)	12652 (4857%)	9578 (3887%)	2717 (1226%)	4.8 (24%)	1.7 (7.8%)	1.7 (7.5%)	1.5 (6.7%)
Subsample SAA	13397 (4900%)	12680 (4868%)	9589 (3892%)	2722 (1229%)	5.2 (26%)	0.3 (1.4%)	0.4 (1.7%)	0.5 (2.4%)

(c) Poisson, $q_G^\star = 99$

Table 4 Impact of λ on policy performance. Here, $\rho = 0.98$. Values to the left of the thick vertical line correspond to the unidentifiable regime, where we report Regret$(q^\pi) - \Delta$ and $\mathcal{R}^{ui}(q^\pi)\%$; values to the right of the thick vertical line correspond to the identifiable regime, where we report Vanilla-Regret(q^π) and $\mathcal{R}^{id}(q^\pi)\%$.

non-robust algorithms is poor in the unidentifiable regime, but improves as λ increases. By contrast, the performance of RCN is highly dependent on the gap between $q_G^\star$ and $q_G^\dagger$, which in turn depends on λ and the tightness of the upper bound M.

Setup. We consider the uniform, exponential, and Poisson distributions described in Section 6.2. Across all experiments, we let $h = 1$ and $b = 9$ (i.e., $\rho = 0.9$). We moreover let $M = 100$ in the case of the uniform distribution, and $M = 325$ for the exponential and Poisson distributions. We report our results in Table 5.

Results. For the uniform distribution (Table 5a), it is easy to show that $q_G^\star = q_G^\dagger$ across all values of λ. Since RCN outputs an empirical estimate of $q_G^\dagger$ in this setting, it achieves near-zero vanilla regret across all values of λ. In the worst case, its additive vanilla regret is 0.9 (corresponding to a relative vanilla regret of 2.1%) in the knife-edge case when $\lambda = 82.64$. In contrast, the other policies perform very poorly for small values of λ, with a relative vanilla regret of 300% in the worst case, achieved by Subsample SAA. Since Censored SAA, Kaplan-Meier and Naive SAA all output λ in the unidentifiable regime, their performance improves as λ increases, thereby approaching $q_G^\star$. Similarly, the performance of Subsample SAA improves, since the fraction of samples for which $d_{ki} < q_k^{\text{off}}$ increases with λ.

Under the exponential distribution (Table 5b), RCN continues to perform well in the unidentifiable regime, despite the fact that $q_G^\star \neq q_G^\dagger$ ($q_G^\star = 184.21$ and $q_G^\dagger \in \{251.32, 234.17, 214.81, 194.24\}$ for increasing unidentifiable values of λ). Its cost exceeds the optimal newsvendor cost by 12% in the worst case, with improved performance as λ increases and $q_G^\dagger$ approaches $q_G^\star$. By contrast, the remaining policies exhibit significantly higher regret (up to 79% for Subsample SAA), as for the uniform distribution.

Under the Poisson distribution (Table 5c), the difference between $q_G^\star$ and $q_G^\dagger$ is significantly higher than in the previous two cases ($q_G^\star = 92$ versus $q_G^\dagger \in \{297.10, 298.14, 293.32, 234.82\}$ for increasing unidentifiable

λ). The gap between $q_G^\dagger$ and $q_G^\star$ — explained by the very loose upper bound M on $q_G^\star$ — results in the poor performance of RCN; in the best case ($\lambda = 85.43$), its cost is over 9 times the optimal newsvendor cost. Finally, we note that when $\lambda = 46$, our algorithm — which outputs an estimate of $q_G^\dagger$ — outperforms all baselines. This is despite the fact that $q_G^\dagger$ is over three times as high as $q_G^\star$ for this value of λ, and can be explained by the fact that $\rho = 0.9$, meaning that lost sales are far more heavily penalized than overage. Consequently, over-ordering relative to $q_G^\star$ mitigates expected cost more effectively than under-ordering by outputting λ, which the baselines do. As λ increases and approaches $q_G^\star$, this advantage diminishes, since λ approaches $q_G^\star$ more rapidly than $q_G^\dagger$.

λ	44.50	57.21	69.93	82.64
True SAA	0.1 (0.2%)	0.1 (0.2%)	0.1 (0.2%)	0.1 (0.2%)
RCN	0.0 (0.1%)	0.1 (0.1%)	0.1 (0.1%)	0.9 (2.1%)
Censored SAA	101 (225%)	52 (115%)	19 (41%)	2.1 (4.7%)
Kaplan-Meier	101 (225%)	52 (115%)	19 (41%)	2.1 (4.7%)
Naive SAA	101 (225%)	52 (115%)	19 (41%)	5.2 (12%)
Subsample SAA	135 (300%)	85 (188%)	45 (100%)	19 (42%)

(a) Uniform, $q_G^\star = 89$, $q_G^\dagger = 89$

λ	92.07	118.38	144.68	170.99
True SAA	0.6 (0.3%)	0.6 (0.3%)	0.5 (0.3%)	0.5 (0.3%)
RCN	21 (12%)	13 (7.0%)	9.6 (5.2%)	1.3 (0.7%)
Censored SAA	81 (44%)	37 (20%)	12 (6.4%)	1.6 (0.9%)
Kaplan-Meier	81 (44%)	37 (20%)	12 (6.4%)	1.5 (0.8%)
Naive SAA	81 (44%)	37 (20%)	24 (13%)	26 (14%)
Subsample SAA	146 (79%)	103 (56%)	74 (40%)	54 (29%)

(b) Exponential, $q_G^\star = 184.21$

λ	46	59.14	72.29	85.43
True SAA	0.0 (0.2%)	0.0 (0.1%)	0.0 (0.2%)	0.0 (0.2%)
RCN	201 (1251%)	202 (1258%)	197 (1228%)	138 (857%)
Censored SAA	290 (1805%)	172 (1069%)	63 (389%)	4.6 (29%)
Kaplan-Meier	290 (1805%)	172 (1069%)	63 (389%)	4.6 (29%)
Naive SAA	290 (1805%)	172 (1069%)	63 (389%)	4.6 (29%)
Subsample SAA	290 (1806%)	179 (1114%)	65 (405%)	8.5 (53%)

(c) Poisson, $q_G^\star = 92$

Table 5 Comparison of policies' vanilla regret in the unidentifiable regime. Here, $\rho = 0.9$.

Appendix B: On The Value of Information: The Case of Globally Well-Separated Distributions

In this section we study how to leverage additional distributional information to improve decision-making in the face of censored data. We do so by considering the setting where the demand distribution is *globally well-separated*, and the decision-maker knows a lower bound $\gamma > 0$ on its pdf. For this setting, our goal is to quantify (i) the *value of information*, measured as the reduction in minimax risk when additional well-separatedness information is available, and (ii) the resulting *sample complexity improvements* of a modified version of RCN that explicitly exploits this information to detect whether the problem instance is identifiable or not. In the remainder of this section, we let Δ^{ws} and $q^{\Delta^{\mathsf{ws}}}$ respectively denote the minimax risk and minimax optimal ordering quantity under the well-separatedness condition.

B.1. Characterizing the Minimax Risk

We first characterize Δ^{ws} and $q^{\Delta^{\mathsf{ws}}}$ in this setting. Under this additional distributional information, the ambiguity set is given by:

$$\mathcal{F}(\lambda; G) = \left\{F \in \mathcal{G} \mid F(x) = G(x)\ \forall\, x < \lambda,\ f(x) \geq \gamma \text{ on its support}, q_F^\star \leq M\right\}. \tag{19}$$

The following lemma will be useful in our derivation of $q^{\Delta^{\mathsf{ws}}}$. We defer its proof to Appendix B.1.1.

LEMMA 3. *For all $F \in \mathcal{F}(\lambda; G)$, F is a distribution with bounded support, with $f(x) = 0$ for all $x > \frac{1}{\gamma}$. Moreover, $q_F^\star \leq \lambda + \frac{\rho - G^-(\lambda)}{\gamma}$.*

Noting that $G^-(\lambda) \geq \lambda\gamma \implies \lambda + \frac{\rho - G^-(\lambda)}{\gamma} \leq \frac{\rho}{\gamma} \leq \frac{1}{\gamma}$, we let $\widetilde{M} = \min\left\{M, \lambda + \frac{\rho - G^-(\lambda)}{\gamma}\right\}$ and modify the definition of $\mathcal{F}(\lambda; G)$ in (19) to be such that $q_F^\star \leq \widetilde{M}$.

THEOREM 4. *Consider a data-driven censored newsvendor problem with well-separated demand distribution G and observable boundary λ. Then, given $\gamma > 0$:*

1. *If $G^-(\lambda) \geq \rho$, $q^{\Delta^{\mathsf{ws}}} = q_G^\star < \lambda$, and $\Delta^{\mathsf{ws}} = 0$.*
2. *If $G^-(\lambda) < \rho$,*

$$q^{\Delta^{\mathsf{ws}}} = \lambda + \frac{1 - G^-(\lambda) - \sqrt{(1 - G^-(\lambda))^2 + (\rho - G^-(\lambda) - \gamma(\widetilde{M} - \lambda))^2 - (\rho - G^-(\lambda))^2}}{\gamma} \geq \lambda.$$

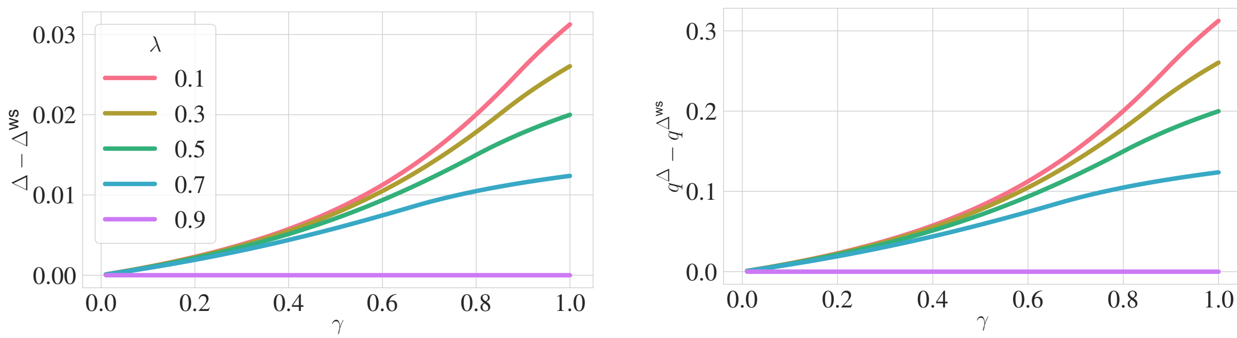
Moreover,

$$\Delta^{\mathsf{ws}} = (b + h)(1 - \rho) \cdot \frac{1 - G^-(\lambda) - \sqrt{(1 - G^-(\lambda))^2 + (\rho - G^-(\lambda) - \gamma(\widetilde{M} - \lambda))^2 - (\rho - G^-(\lambda))^2}}{\gamma}.$$

To gain further insight into the value of distributional information in this setting, in Figure 6a we plot $\Delta - \Delta^{\mathsf{ws}}$ as a function of γ, when $D \sim \text{Unif}[0,1]$ and $b = 9, h = 1$ so $\rho = 0.9$. We observe that $\Delta - \Delta^{\mathsf{ws}}$ is increasing in γ, for fixed λ. This phenomenon is intuitive, since the size of the ambiguity set decreases as γ increases; this supports the interpretation of γ as the quality of the decision-maker's additional information. We additionally observe that, for fixed γ, $\Delta - \Delta^{\mathsf{ws}}$ is concave decreasing in λ. This shows that the decision-maker has the most to gain from additional distributional information when the dataset is highly censored and therefore relatively uninformative.

Figure 6b further illustrates that γ not only affects the worst-case regret of the decision-maker, but also her minimax optimal ordering quantity. In particular, we observe that $q^{\Delta^{\mathsf{ws}}}$ is *lower* than q^Δ for all values of λ and γ. Moreover, for fixed λ, the difference between q^Δ and $q^{\Delta^{\mathsf{ws}}}$ is increasing in γ. The reason for this is also intuitive: as γ increases, the decision-maker is guaranteed that a larger amount of mass resides close to λ, thereby reducing the likelihood of large realizations of D and lowering the minimax optimal ordering quantity as a result.

Proof of Theorem 4. The proof of the first fact is identical to the more general setting we consider in Theorem 1; we omit it as such. In the remainder of the proof we focus on the case where $G^-(\lambda) < \rho$. The proof proceeds in the same fashion as that of Theorem 1. Namely, we first provide closed-form expressions for the worst-case regret over $\mathcal{F}(\lambda; G)$, given $q \leq \widetilde{M}$.



(a) $\Delta - \Delta^{\mathrm{ws}}$ vs. γ (b) $q^{\Delta} - q^{\Delta^{\mathrm{ws}}}$ vs. γ

Figure 6 Dependence of $q^{\Delta} - q^{\Delta^{\mathrm{ws}}}$ and $\Delta - \Delta^{\mathrm{ws}}$ on γ, for $D \sim$ Unif$[0,1]$. Here, we assume $M = 1$, $b = 9$ and $h = 1$.

LEMMA 4.

$$\text{Regret}(q) = \begin{cases} (b+h)\Big(\rho(\widetilde{M}-q) + \mathbb{E}_G[(q-D)\mathbb{1}\{D \le q\}] - \mathbb{E}_G\big[(\widetilde{M}-D)\mathbb{1}\{D<\lambda\}\big] - \frac{\gamma}{2}(\widetilde{M}-\lambda)^2\Big) & \text{if } q < \lambda \\ (b+h)\max\big\{(1-\rho)(q-\lambda), (\widetilde{M}-q)\rho - G^-(\lambda) - \gamma(q-\lambda)) - \frac{\gamma}{2}(\widetilde{M}-q)^2\big\} & \text{if } q \ge \lambda. \end{cases}$$

We defer the proof of the lemma to Appendix B.1.2. We use these expressions to compute the minimax optimal ordering quantity $q_G^{\dagger}$ in this case. For all $q < \lambda$, we have:

$$\frac{d}{dq}\Big[\rho(\widetilde{M}-q) + \mathbb{E}_G[(q-D)\mathbb{1}\{D \le q\}] - \mathbb{E}_G\big[(\widetilde{M}-D)\mathbb{1}\{D<\lambda\}\big] - \frac{\gamma}{2}(\widetilde{M}-\lambda)^2\Big] = -\rho + G(q) < 0,$$

since $G^-(\lambda) < \rho$. Therefore:

$$\begin{aligned} \inf_{q<\lambda} \text{Regret}(q) &= (b+h)\Big(\rho(\widetilde{M}-\lambda) + \mathbb{E}_G[(\lambda-D)\mathbb{1}\{D \le \lambda\}] - \mathbb{E}_G\big[(\widetilde{M}-D)\mathbb{1}\{D<\lambda\}\big] - \frac{\gamma}{2}(\widetilde{M}-\lambda)^2\Big) \\ &= (b+h)\Big(\rho(\widetilde{M}-\lambda) - (\widetilde{M}-\lambda)G^-(\lambda) - \frac{\gamma}{2}(\widetilde{M}-\lambda)^2\Big) \\ &= (b+h)\Big((\widetilde{M}-\lambda)(\rho - G^-(\lambda)) - \frac{\gamma}{2}(\widetilde{M}-\lambda)^2\Big) = \text{Regret}(\lambda). \end{aligned}$$

Thus we have, $q_G^{\dagger} = \arg\min_{q \ge \lambda} \max\big\{(1-\rho)(q-\lambda), (\widetilde{M}-q)(\rho - G^-(\lambda) - \gamma(q-\lambda)) - \frac{\gamma}{2}(\widetilde{M}-q)^2\big\}$. Equivalently, $q_G^{\dagger}$ is the smallest solution to:

$$\begin{aligned} (1-\rho)(q-\lambda) &= (\widetilde{M}-q)(\rho - G^-(\lambda) - \gamma(q-\lambda)) - \frac{\gamma}{2}(\widetilde{M}-q)^2 \\ \iff q_G^{\dagger} &= \lambda + \frac{1 - G^-(\lambda) - \sqrt{(1-G^-(\lambda))^2 + (\rho - G^-(\lambda) - \gamma(\widetilde{M}-\lambda))^2 - (\rho - G^-(\lambda))^2}}{\gamma} \ge \lambda, \end{aligned}$$

where the final solution simply follows from algebra; we omit it as such. □

B.1.1. Proof of Lemma 3

Proof. To see that $f(x) = 0$ for all $x > \frac{1}{\gamma}$, note that $F(x) \ge \gamma x$ for all x in the support of F. Therefore, $F(x) = 1$ for some $x \le \frac{1}{\gamma}$, which proves the claim.

For the lower bound on $q_F^{\star}$, we have:

$$q_F^{\star} = \inf\{q > 0 : F(q) \ge \rho\} \le \inf\{q > 0 : G^-(\lambda) + \gamma(q-\lambda) \ge \rho\}.$$

Rearranging the above inequality yields the result. □

B.1.2. Proof of Lemma 4

Proof. As in the proof of Lemma 2 (see Eq. (31)), it suffices to analyze:

$$
\begin{aligned}
&\sup_{F\in\mathcal{F}(\lambda;G)} C_F(q)-C_F(q_F^\star)\\
&=\sup_{\tilde q\in[\lambda,\widetilde{M}]} b(\tilde q-q)+(b+h)\left(\sup_{\substack{F\in\mathcal{F}(\lambda;G):\\ q_F^\star=\tilde q}} \mathbb{E}_F[(q-D)\mathbb{1}\{D\le q\}-(\tilde q-D)\mathbb{1}\{D\le\tilde q\}]\right)\\
&=(b+h)\left(\sup_{\tilde q\in[\lambda,\widetilde{M}]} \rho(\tilde q-q)+\left(\sup_{\substack{F\in\mathcal{F}(\lambda;G):\\ q_F^\star=\tilde q}} \mathbb{E}_F[(q-D)\mathbb{1}\{D\le q\}-(\tilde q-D)\mathbb{1}\{D\le\tilde q\}]\right)\right)
\end{aligned}
\tag{20}
$$

Note that, for $\tilde q=q$, (20) $=0$. Therefore, we partition our analysis into two cases: $\tilde q<q$ and $\tilde q>q$.

(i) $q>\tilde q\ge\lambda$: In this case, by the proof of Lemma 2 (Eq. (32)), the worst-case distribution F that has $\tilde q$ as its newsvendor solution maximizes:

$$
\mathbb{E}_F[(q-\tilde q)\mathbb{1}\{D\le\tilde q\}+(q-D)\mathbb{1}\{\tilde q<D\le q\}]. \tag{21}
$$

Observe that $q-D<q-\tilde q$ for all $D\in(\tilde q,q]$. Hence, any worst-case distribution F should minimize $\mathbb{P}_F(D\in(\tilde q,q])$, which occurs by letting $\tilde q$ be the maximum of the support of the distribution. Therefore:

$$
\begin{aligned}
&(21)=(q-\tilde q)\\
\Longrightarrow\ &\text{Regret}(q)=(b+h)\sup_{\tilde q\in[\lambda,\widetilde{M}]}(1-\rho)(q-\tilde q)=(b+h)(1-\rho)(q-\lambda).
\end{aligned}
$$

(ii) $q<\lambda\le\tilde q$. As in the proof of Lemma 2 (Eq. (35)), for a fixed $\tilde q$, it suffices to find the distribution that maximizes:

$$
\sup_{\substack{F\in\mathcal{F}(\lambda;G):\\ q_F^\star=\tilde q}} \mathbb{E}_G[(q-D)\mathbb{1}\{D\le q\}]-\mathbb{E}_G[(\tilde q-D)\mathbb{1}\{D<\lambda\}]-\mathbb{E}_F[(\tilde q-D)\mathbb{1}\{\lambda\le D<\tilde q\}]. \tag{22}
$$

Suppose first that $\tilde q>\lambda$. Since $\tilde q-D>0$ for all $D\in[\lambda,\tilde q)$, a similar argument as above establishes that any worst-case distribution $F\in\mathcal{F}(\lambda;G)$ necessarily sets $\mathbb{P}_F(D\in[\lambda,\tilde q))=\gamma$. Using this fact above, we obtain:

$$
\begin{aligned}
(22)&=\mathbb{E}_G[(q-D)\mathbb{1}\{D\le q\}]-\mathbb{E}_G[(\tilde q-D)\mathbb{1}\{D<\lambda\}]-\gamma\int_\lambda^{\tilde q}(\tilde q-x)dx\\
&=\mathbb{E}_G[(q-D)\mathbb{1}\{D\le q\}]-\mathbb{E}_G[(\tilde q-D)\mathbb{1}\{D<\lambda\}]-\frac{\gamma}{2}(\tilde q-\lambda)^2.
\end{aligned}
$$

Plugging this back into (20), we have:

$$
(20)=(b+h)\left(\sup_{\tilde q\in[\lambda,\widetilde{M}]}\rho(\tilde q-q)+\mathbb{E}_G[(q-D)\mathbb{1}\{D\le q\}]-\mathbb{E}_G[(\tilde q-D)\mathbb{1}\{D<\lambda\}]-\frac{\gamma}{2}(\tilde q-\lambda)^2\right).
$$

By the first-order condition, the unconstrained supremum of the above is achieved at the smallest value of $\tilde q$ such that:

$$
\rho-G^-(\lambda)-\gamma(\tilde q-\lambda)\le 0 \iff \tilde q\ge\lambda+\frac{\rho-G^-(\lambda)}{\gamma}.
$$

Since $\tilde{q} \le \widetilde{M} \le \frac{\rho - G^-(\lambda)}{\gamma}$, the constrained supremum is attained at $\widetilde{M}$, and

$$\text{Regret}(q) = (b+h)\Big(\rho(\widetilde{M} - q) + \mathbb{E}_G[(q-D)\mathbb{1}\{D \le q\}] - \mathbb{E}_G\big[(\widetilde{M} - D)\mathbb{1}\{D < \lambda\}\big] - \frac{\gamma}{2}(\widetilde{M} - \lambda)^2\Big). \tag{23}$$

Now, if $\tilde{q} = \lambda$:

$$\text{Regret}(q) = (b+h)\Big(\rho(\lambda - q) + \mathbb{E}_G[(q-D)\mathbb{1}\{D \le q\}] - \mathbb{E}_G[(\lambda - D)\mathbb{1}\{D < \lambda\}]\Big). \tag{24}$$

Comparing (23) and (24), setting $\tilde{q} > \lambda$ is optimal if and only if:

$$(\rho - G^-(\lambda))(\widetilde{M} - \lambda) - \frac{\gamma}{2}(\widetilde{M} - \lambda)^2 \ge 0 \iff \rho - G^-(\lambda) - \frac{\gamma}{2}(\widetilde{M} - \lambda) \ge 0.$$

The above holds, since by definition $\widetilde{M} \le \lambda + \frac{\rho - G^-(\lambda)}{\gamma}$, and $\rho - G^-(\lambda) > 0$.

(iii) $\lambda \le q < \tilde{q}$: In this case, by the proof of Lemma 2 (Eq. (38)), it suffices to find a distribution F that maximizes:

$$-(\tilde{q} - q)G^-(\lambda) - (\tilde{q} - q)\mathbb{P}_F(\lambda \le D \le q) - \mathbb{E}_F\Big[(\tilde{q} - D)\mathbb{1}\{q < D \le \tilde{q}\}\Big]. \tag{25}$$

Since $0 < \tilde{q} - D < \tilde{q} - q$ for all $D \in (q, \tilde{q})$, the supremum of the above is achieved for F such that $f(x) = \gamma$ for all $x \in [\lambda, \tilde{q})$. In this case, we have:

$$\begin{aligned}(25) &= -(\tilde{q} - q)G^-(\lambda) - \gamma(\tilde{q} - q)(q - \lambda) - \gamma \int_q^{\tilde{q}} (\tilde{q} - x)dx \\ &= -(\tilde{q} - q)(G^-(\lambda) + \gamma(q - \lambda)) - \frac{\gamma}{2}(\tilde{q} - q)^2.\end{aligned}$$

Plugging this back into (20), we have:

$$(20) = (b+h)\left(\sup_{\tilde{q} \in [\lambda, \widetilde{M}]} \rho(\tilde{q} - q) - (\tilde{q} - q)(G^-(\lambda) + \gamma(q - \lambda)) - \frac{\gamma}{2}(\tilde{q} - q)^2\right).$$

By the first-order condition, the unconstrained supremum of the above is attained at the smallest $\tilde{q}$ such that:

$$\rho - (G^-(\lambda) + \gamma(q - \lambda)) - \gamma(\tilde{q} - q) \le 0 \iff \tilde{q} \ge q + \frac{\rho - (G^-(\lambda) + \gamma(q - \lambda))}{\gamma} = \lambda + \frac{\rho - G^-(\lambda)}{\gamma}.$$

Again, using the fact that $\tilde{q} \le \widetilde{M} \le \lambda + \frac{\rho - G^-(\lambda)}{\gamma}$, we obtain that the constrained supremum is attained at $\tilde{q} = \widetilde{M}$, and

$$\text{Regret}(q) = (b+h)\Big((\widetilde{M} - q)(\rho - (G^-(\lambda) + \gamma(q - \lambda)) - \frac{\gamma}{2}(\widetilde{M} - q)^2\Big).$$

Putting these three cases together, we obtain the result. □

B.2. Practical Performance Improvements in the Knife-Edge Regime

We next numerically investigate how the well-separated condition can generate sample complexity gains in the knife-edge regime, where testing for identifiability is the hardest. To do so, we introduce a new algorithm, which we denote $\mathsf{RCN}^{\mathsf{ws}}$. Similar to RCN, $\mathsf{RCN}^{\mathsf{ws}}$ proceeds hierarchically, by first testing the identifiability regime, and then estimating the appropriate $q^{\Delta^{\mathsf{ws}}}$. However, its identifiability test leverages the well-separatedness condition via the following observations:

ALGORITHM 3: Robust Censored Well-Separated Newsvendor (RCN$^{\text{ws}}$)

Input: Censored demand samples s^{off}, confidence terms $\zeta_k, k \in [K]$, well-separated parameter γ
Output: Ordering quantity q^{alg}
For all $k \in [K]$, compute censored SAA of $G^-(q_k^{\text{off}})$: $\widehat{G}_k^-(q_k^{\text{off}}) = \frac{1}{N_k}\sum_{i\in[N_k]} \mathbb{1}\{s_{ki}^{\text{off}} < q_k^{\text{off}}\}$.
Define $\widehat{G}^-(\lambda) = \widehat{G}_K^-(q_K^{\text{off}})$.
Compute censored SAA of $q_G^\star$: $q_{\widehat{G}}^\star = \inf\left\{x \mid \widehat{G}(x) \geq \rho\right\}$, where $\widehat{G}(x) = \frac{1}{N_K}\sum_{i\in[N_K]} \mathbb{1}\{s_{iK}^{\text{off}} \leq x\}$.
if $\gamma\lambda \geq \rho$ or $\widehat{G}^-(\lambda) \geq \rho + \zeta_K$ or $\widehat{G}_k^-(q_k^{\text{off}}) \geq \rho + \zeta_k - \gamma(\lambda - q_k^{\text{off}})$ for some $k \in [K]$ **then** `// Likely identifiable`
    Let $q^{\text{alg}} = q_{\widehat{G}}^\star$.
else if $\widehat{G}^-(\lambda) > \rho$ **then** `// Borderline: use separation test`
    if $\gamma(\lambda - q_{\widehat{G}}^\star) \geq \zeta_K$ **then** `// Separation overcomes noise`
        Let $q^{\text{alg}} = q_{\widehat{G}}^\star$. `// Likely identifiable`
    else
        Let $q^{\text{alg}} = \lambda$. `// Knife-edge`
else if $\widehat{G}^-(\lambda) \in (\rho - \zeta, \rho]$ **then** `// Knife-edge`
    Let $q^{\text{alg}} = \lambda$.
else `// Likely unidentifiable`
    Compute empirical estimate of $q_G^\dagger$:

$$q_{\widehat{G}}^\dagger = \frac{\widetilde{M}(\rho - \widehat{G}^-(\lambda) + \gamma\lambda) + \lambda(1-\rho) - \frac{\gamma}{2}(\widetilde{M}^2 + \lambda^2)}{1 - \widehat{G}^-(\lambda)} \quad \text{where } \widetilde{M} = \min\left\{M, \frac{1}{\gamma}, \lambda + \frac{\rho - \widehat{G}^-(\lambda)}{\gamma}\right\}. \tag{26}$$

    Let $q^{\text{alg}} = q_{\widehat{G}}^\dagger$.
return q^{alg}

1. If $\gamma\lambda \geq \rho$, by the well-separatedness condition, $G^-(\lambda) \geq \lambda\gamma \geq \rho$, and the problem is guaranteed to be identifiable. In this case, no statistical test is needed.
2. If $\widehat{G}^-(\lambda) > \rho$, $q_{\widehat{G}}^\star < \lambda$. Then, by the well-separatedness condition, with high probability:

$$G^-(\lambda) \geq G^-(q_{\widehat{G}}^\star) + \gamma(\lambda - q_{\widehat{G}}^\star) \geq \widehat{G}_K^-(q_{\widehat{G}}^\star) - \zeta_K + \gamma(\lambda - q_{\widehat{G}}^\star) \geq \rho - \zeta_K + \gamma(\lambda - q_{\widehat{G}}^\star).$$

   Thus, if $\gamma(\lambda - q_{\widehat{G}}^\star) \geq \zeta_K$, $G^-(\lambda) \geq \rho$, and the problem is identifiable.
3. If $\widehat{G}_k^-(q_k^{\text{off}}) \geq \rho + \zeta_k - \gamma(\lambda - q_k^{\text{off}})$ for some $k \in [K]$, we similarly have:

$$G^-(\lambda) \geq G^-(q_k^{\text{off}}) + \gamma(\lambda - q_k^{\text{off}}) \geq \widehat{G}_k^-(q_k^{\text{off}}) - \zeta_k + \gamma(\lambda - q_k^{\text{off}}) \geq \rho,$$

   which again guarantees identifiability.

We use these three observations in our modified algorithm, presented in Algorithm 3.

B.2.1. Experimental setup. Across all experiments we let $h = 1$ and $b = 9$, so that $\rho = 0.9$. We consider the following demand distributions:

- *Continuous Uniform*: Demand is drawn from a continuous uniform distribution over $[0, 100]$. We let $M = 100$, and $\gamma = 1/100$.
- *Truncated Exponential*: Demand is drawn from a truncated exponential distribution with rate $1/80$ with support over $[0, 325]$. We let $M = 325$, and $\gamma = \frac{1}{80}e^{-325/80}$.

We moreover consider $K = 2$ historical ordering quantities and vary λ as in Section 6.2.1.

B.2.2. Results. We demonstrate the improvements in classifying the problem as identifiable by comparing the performance of RCN$^{\text{ws}}$ to that of RCN in the knife-edge regime where $\lambda \in [q_G^\star, 1.05q_G^\star]$. Additionally, in order to isolate the *sample complexity* gains of RCN$^{\text{ws}}$, we modify RCN to instead estimate the value of

λ	90.89	91.79	92.69	93.59	94.49
True SAA	0.1 (0.2%)	0.1 (0.2%)	0.1 (0.1%)	0.1 (0.2%)	0.1 (0.2%)
RCN	0.1 (0.2%)	0.2 (0.4%)	0.4 (0.9%)	0.7 (1.5%)	0.7 (1.5%)
RCN$^{\text{ws}}$	0.1 (0.3%)	0.1 (0.2%)	0.1 (0.2%)	0.1 (0.3%)	0.1 (0.3%)

(a) Uniform, $q^\star_G = 89$

λ	186.69	190.63	194.58	198.52	202.47
True SAA	0.2 (0.1%)	0.2 (0.1%)	0.3 (0.2%)	0.3 (0.2%)	0.2 (0.1%)
RCN	0.1 (0.0%)	0.2 (0.1%)	0.5 (0.3%)	1.1 (0.6%)	1.8 (1.1%)
RCN$^{\text{ws}}$	0.1 (0.0%)	0.2 (0.1%)	0.5 (0.3%)	1.1 (0.6%)	1.8 (1.1%)

(b) Truncated Exponential, $q^\star_G = 184.14$

Table 6 Vanilla-Regret(q^π) vs. λ. The relative vanilla regret $\mathcal{R}^{id}(q^\pi)\%$ is shown in parentheses.

$q^{\Delta^{\text{ws}}}$ associated with the well-separated setting. Our results can be found in Table 6. The performance of the two algorithms in the likely unidentifiable and likely identifiable regimes is similar, and omitted as such.

For the Continuous Uniform distribution (Table 6a), RCN$^{\text{ws}}$ yields modest but consistent performance gains over RCN across all tested values of λ. Specifically, for $\lambda \in \{90.89, 91.79, 92.69\}$, RCN$^{\text{ws}}$ achieves a 0% misclassification rate thanks to the "extra conditions" for identifiability, and therefore correctly outputs an estimate of $q^\star_G$. RCN, on the other hand, misclassifies the problem as *knife-edge* in 100% of replications, therefore outputting λ. For $\lambda = 93.59$ and $\lambda = 94.49$, RCN$^{\text{ws}}$ applies the identifiability condition in 90% and 60% of runs respectively (again with no misclassifications), while RCN continues to misclassify the problem at those same rates.

For the Truncated Exponential distribution (Table 6b), RCN and RCN$^{\text{ws}}$ achieve identical performance. This behavior arises because the chosen value of γ severely underestimates the local probability density around λ, rendering the additional identifiability conditions in RCN$^{\text{ws}}$ inactive. This highlights the importance of having tight bounds on γ in order to achieve sample complexity gains from this additional distributional information.

Appendix C: Useful Facts

C.1. Closed-form expression for difference in newsvendor costs

We repeatedly rely on the following closed-form expression for the difference between the newsvendor cost under two arbitrary ordering quantities q_1 and q_2.

PROPOSITION 3. *For any distribution $F \in \mathcal{G}$, $q_1, q_2 \in [0, M]$:*

$$C_F(q_1) - C_F(q_2) = b(q_2 - q_1) + (b+h)\mathbb{E}_F[(q_1 - D)\mathbb{1}\{D \le q_1\} - (q_2 - D)\mathbb{1}\{D \le q_2\}]. \tag{27}$$

Proof. By definition, for any q:

$$\begin{aligned}
C_F(q) &= b\mathbb{E}_F[(D-q)^+] + h\mathbb{E}_F[(q-D)^+] \\
&= b\mathbb{E}_F[(D-q)\mathbb{1}\{D > q\}] + h\mathbb{E}_F[(q-D)\mathbb{1}\{D \le q\}] \\
&= b\mathbb{E}_F[(D-q)(1-\mathbb{1}\{D \le q\})] + h\mathbb{E}_F[(q-D)\mathbb{1}\{D \le q\}] \\
&= b\mathbb{E}_F[D-q] - b\mathbb{E}_F[(D-q)\mathbb{1}\{D \le q\}] + h\mathbb{E}_F[(q-D)\mathbb{1}\{D \le q\}] \\
&= b\mathbb{E}_F[D-q] + (b+h)\mathbb{E}_F[(q-D)\mathbb{1}\{D \le q\}].
\end{aligned} \tag{28}$$

Applying (28) to q_1 and q_2, we obtain:

$$\begin{aligned}
C_F(q_1) - C_F(q_2) &= b\mathbb{E}_F[D-q_1] - b\mathbb{E}_F[D-q_2] + (b+h)\mathbb{E}_F[(q_1-D)\mathbb{1}\{D \le q_1\}] \\
&\quad - (b+h)\mathbb{E}_F[(q_2-D)\mathbb{1}\{D \le q_2\}] \\
&= b(q_2-q_1) + (b+h)\mathbb{E}_F[(q_1-D)\mathbb{1}\{D \le q_1\} - (q_2-D)\mathbb{1}\{D \le q_2\}].
\end{aligned}$$

□

C.2. On the relationship between λ and M

We implicitly rely on the following fact in all of our results.

PROPOSITION 4. *Suppose $G^-(\lambda) < \rho$. Then, $\lambda \le q_G^\star \le M$.*

Proof. We argue that $G^-(\lambda) < \rho$ implies $\lambda \le q_G^\star$ via the converse. Namely, suppose $\lambda > q_G^\star$. Recall, by definition of $q_G^\star$, $G(q_G^\star) \ge \rho$, which implies there exists $q < \lambda$ such that $G(q) \ge \rho$, i.e., $G^-(\lambda) \ge \rho$. Note that $q_G^\star \le M$ by assumption, and we obtain the claim. □

C.3. Worst-case regret in the identifiable regime

We use the following fact in the subsequent.

PROPOSITION 5. *Suppose $G^-(\lambda) \ge \rho$. Then, for any $q \in [0, M]$,*

$$\begin{aligned}
\mathrm{Regret}(q) &:= \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) \\
&= \begin{cases} C_G(q) - C_G(q_G^\star) & \text{if } q < \lambda \\ b(q_G^\star - q) + (b+h)\Big[(q-\lambda) + \mathbb{E}_G\big[(\lambda - D)\mathbb{1}\{D<\lambda\} - (q_G^\star - D)\mathbb{1}\{D \le q_G^\star\}\big]\Big] & \text{if } q \ge \lambda. \end{cases}
\end{aligned}$$

Proof. Recall that, in this regime, by Lemma 1, $q_F^\star = q_G^\star < \lambda$ for all $F \in \mathcal{F}(\lambda; G)$. Applying this fact to Proposition 3, for any $F \in \mathcal{F}(\lambda; G)$ we have:

$$C_F(q) - C_F(q_F^\star) = b(q_G^\star - q) + (b+h)\mathbb{E}_F\big[(q-D)\mathbb{1}\{D \le q\} - (q_G^\star - D)\mathbb{1}\{D \le q_G^\star\}\big]. \tag{29}$$

If $q < \lambda$, since $F \in \mathcal{F}(\lambda; G)$ and $q_F^\star < \lambda$, the above evaluates to:

$$\begin{aligned}
C_F(q) - C_F(q_F^\star) &= b(q_G^\star - q) + (b+h)\mathbb{E}_G\big[(q-D)\mathbb{1}\{D \le q\} - (q_G^\star - D)\mathbb{1}\{D \le q_G^\star\}\big] \\
&= C_G(q) - C_G(q_G^\star),
\end{aligned}$$

which proves the first case.

Suppose now that $q \ge \lambda$. Note that only the second term of (29) has dependence on F. Again, since $q_G^\star < \lambda$, for all $F \in \mathcal{F}(\lambda; G)$ we have:

$$\begin{aligned}
&\mathbb{E}_F\big[(q-D)\mathbb{1}\{D \le q\} - (q_G^\star - D)\mathbb{1}\{D \le q_G^\star\}\big] \\
&= \mathbb{E}_G[(q-D)\mathbb{1}\{D<\lambda\}] + \mathbb{E}_F[(q-D)\mathbb{1}\{\lambda \le D < q\}] - \mathbb{E}_G\big[(q_G^\star - D)\mathbb{1}\{D \le q_G^\star\}\big] \\
&\le \mathbb{E}_G[(q-D)\mathbb{1}\{D<\lambda\}] + (q-\lambda)\mathbb{P}_F(D=\lambda) - \mathbb{E}_G\big[(q_G^\star - D)\mathbb{1}\{D \le q_G^\star\}\big],
\end{aligned} \tag{30}$$

where the inequality is attained by the distribution F that places all remaining mass on λ. Re-arranging (30), we obtain:

$$\begin{aligned}
&\sup_{F \in \mathcal{F}(\lambda;G)} \mathbb{E}_F\big[(q-D)\mathbb{1}\{D \le q\} - (q_G^\star - D)\mathbb{1}\{D \le q_G^\star\}\big] \\
&= \mathbb{E}_G[(q-\lambda)\mathbb{1}\{D<\lambda\}] + \mathbb{E}_G[(\lambda - D)\mathbb{1}\{D<\lambda\}] \\
&\qquad + (q-\lambda)(1 - G^-(\lambda)) - \mathbb{E}_G\big[(q_G^\star - D)\mathbb{1}\{D \le q_G^\star\}\big] \\
&= (q-\lambda) + \mathbb{E}_G[(\lambda - D)\mathbb{1}\{D<\lambda\}] - \mathbb{E}_G\big[(q_G^\star - D)\mathbb{1}\{D \le q_G^\star\}\big].
\end{aligned}$$

Plugging this back into (29), the second case is shown. □

Appendix D: Omitted Proofs

D.1. Section 3 Omitted Proofs

D.1.1. Proof of Proposition 1

Proof. Suppose first that $\Delta > 0$. For any policy π, $\sup_{F\in\mathcal{F}(\lambda;G)}\{C_F(q^\pi) - C_F(q_F^\star)\} \geq \Delta > 0$, by definition of Δ. Therefore, for all $\epsilon \in (0,\Delta)$, $\sup_{F\in\mathcal{F}(\lambda;G)}\left\{C_F(q^\pi) - C_F(q_F^\star)\right\} > \epsilon$, which implies that the problem is unidentifiable.

The proof of the converse is a corollary of our main algorithmic result, in which we design a policy π such that $\sup_{F\in\mathcal{F}(\lambda;G)}\left\{C_F(q^\pi) - C_F(q_F^\star)\right\} = \widetilde{O}(1/\sqrt{N})$ with probability at least $1 - O(1/\sqrt{N})$ when $\Delta = 0$ (see Theorem 2). Taking $N \to \infty$ then implies that the problem is identifiable. □

D.1.2. Proof of Lemma 1

Proof. By definition, $q_G^\star = \inf\{q : G(q) \geq \rho\}$. If $G^-(\lambda) \geq \rho$, there must exist $q < \lambda$ such that $G(q) \geq \rho$, which implies $q_G^\star < \lambda$.

Consider now $q_F^\star$. Again, by definition,

$$\begin{aligned} q_F^\star &= \inf\{q : F(q) \geq \rho\} \\ &= \inf\{q : G(q)\mathbb{1}\{q<\lambda\} + (\mathbb{P}_G(D<\lambda) + \mathbb{P}_F(\lambda \leq D < q))\mathbb{1}\{q \geq \lambda\} \geq \rho\}, \end{aligned}$$

where the second equality uses the fact that $F \in \mathcal{F}(\lambda;G)$, hence $F(x) = G(x)\ \forall\, x < \lambda$. Again, using the fact that $G^-(\lambda) \geq \rho$, there must be $q < \lambda$ such that $G(q) \geq \rho$. Hence, the above infimum is necessarily achieved at $q_G^\star$. □

D.1.3. Proof of Lemma 2

Proof. By Proposition 3, we have:

$$\begin{aligned} &\sup_{F\in\mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) \\ &= \sup_{\tilde{q}\in[\lambda,M]} \sup_{\substack{F\in\mathcal{F}(\lambda;G):\\ q_F^\star=\tilde{q}}} b(\tilde{q}-q) + (b+h)\mathbb{E}_F[(q-D)\mathbb{1}\{D\leq q\} - (\tilde{q}-D)\mathbb{1}\{D\leq\tilde{q}\}] \\ &= \sup_{\tilde{q}\in[\lambda,M]} b(\tilde{q}-q) + (b+h)\left(\sup_{\substack{F\in\mathcal{F}(\lambda;G):\\ q_F^\star=\tilde{q}}} \mathbb{E}_F[(q-D)\mathbb{1}\{D\leq q\} - (\tilde{q}-D)\mathbb{1}\{D\leq\tilde{q}\}]\right). \end{aligned} \tag{31}$$

Here, the supremum is over all distributions $F \in \mathcal{F}(\lambda;G)$ such that $\tilde{q} = q_F^\star \geq \lambda$, since $q_F^\star < \lambda \implies F^-(\lambda) \geq \rho$. However, since $F \in \mathcal{F}(\lambda;G)$, $F^-(\lambda) = G^-(\lambda) < \rho$, a contradiction.

We analyze (31) depending on the following three cases: (i) $q > \tilde{q}$, (ii) $q < \tilde{q}$, and (iii) $q = \tilde{q}$.

(i) $q > \tilde{q}$: In this case, the worst-case distribution F that has $\tilde{q}$ as its newsvendor solution maximizes:

$$\mathbb{E}_F[(q-D)\mathbb{1}\{D\leq q\} - (\tilde{q}-D)\mathbb{1}\{D\leq\tilde{q}\}] = \mathbb{E}_F[(q-\tilde{q})\mathbb{1}\{D\leq\tilde{q}\} + (q-D)\mathbb{1}\{\tilde{q}<D\leq q\}]. \tag{32}$$

Observe that $q - D < q - \tilde{q}$ for all $D \in (\tilde{q}, q]$. Hence, for any distribution F such that $\mathbb{P}_F(D \in (\tilde{q}, q]) > 0$, there exists a strictly worse distribution that sets $\mathbb{P}_F(D \in (\tilde{q}, q]) = 0$, moving all of the remaining mass to $\tilde{q}$. (Note that such a distribution will still have $q_F^\star = \tilde{q}$.) For any such F, then, (32) = $(q-\tilde{q})\mathbb{P}_F(D \leq \tilde{q})$. Using this in (31):

$$\sup_{\tilde{q}\in[\lambda,q)} b(\tilde{q}-q) + (b+h)\left(\sup_{\substack{F\in\mathcal{F}(\lambda;G):\\ q_F^\star=\tilde{q}}} \mathbb{E}_F[(q-D)\mathbb{1}\{D\leq q\} - (\tilde{q}-D)\mathbb{1}\{D\leq\tilde{q}\}]\right)$$

$$
\begin{aligned}
&= \sup_{\tilde q \in [\lambda, q)} b(\tilde q - q) + (b+h)(q - \tilde q)\left(\sup_{\substack{F \in \mathcal{F}(\lambda; G): \\ q_F^\star = \tilde q}} \mathbb{P}_F(D \le \tilde q) \right) \\
&= \sup_{\tilde q \in [\lambda, q)} b(\tilde q - q) + (b+h)(q - \tilde q), \qquad (33)
\end{aligned}
$$

achieved by letting F be such that $\mathbb{P}_F(D \le \tilde q) = 1$, since $q > \tilde q$ by assumption. Further simplifying, we obtain:

$$
(33) = \sup_{\tilde q \in [\lambda, q)} h(q - \tilde q) = h(q - \lambda). \qquad (34)
$$

(ii) $q < \tilde q$: Suppose first that $q < \lambda$. In this case, for a fixed $\tilde q$:

$$
\begin{aligned}
&\sup_{\substack{F \in \mathcal{F}(\lambda; G): \\ q_F^\star = \tilde q}} \mathbb{E}_F[(q - D)\mathbb{1}\{D \le q\} - (\tilde q - D)\mathbb{1}\{D \le \tilde q\}] \\
&= \sup_{\substack{F \in \mathcal{F}(\lambda; G): \\ q_F^\star = \tilde q}} \mathbb{E}_G[(q - D)\mathbb{1}\{D \le q\}] - \mathbb{E}_G[(\tilde q - D)\mathbb{1}\{D < \lambda\}] - \mathbb{E}_F[(\tilde q - D)\mathbb{1}\{\lambda \le D < \tilde q\}]. \qquad (35)
\end{aligned}
$$

Since $\tilde q - D > 0$ for all $D \in [\lambda, \tilde q)$, a similar argument as above establishes that any worst-case distribution $F \in \mathcal{F}(\lambda; G)$ necessarily sets $\mathbb{P}_F(D \in [\lambda, \tilde q)) = 0$. Using this fact above, we obtain:

$$
\begin{aligned}
&\sup_{\substack{F \in \mathcal{F}(\lambda; G): \\ q_F^\star = \tilde q}} \mathbb{E}_F[(q - D)\mathbb{1}\{D \le q\} - (\tilde q - D)\mathbb{1}\{D \le \tilde q\}] \\
&= \mathbb{E}_G[(q - D)\mathbb{1}\{D \le q\}] - \mathbb{E}_G[(\tilde q - D)\mathbb{1}\{D < \lambda\}].
\end{aligned}
$$

Further using this in (31), we then have:

$$
\begin{aligned}
&\sup_{\tilde q \in [\lambda, M]} b(\tilde q - q) + (b+h)\left(\sup_{\substack{F \in \mathcal{F}(\lambda; G): \\ q_F^\star = \tilde q}} \mathbb{E}_F[(q - D)\mathbb{1}\{D \le q\} - (\tilde q - D)\mathbb{1}\{D \le \tilde q\}] \right) \\
&= \sup_{\tilde q \in [\lambda, M]} b(\tilde q - q) + (b+h)(\mathbb{E}_G[(q - D)\mathbb{1}\{D \le q\}] - \mathbb{E}_G[(\tilde q - D)\mathbb{1}\{D < \lambda\}]) \\
&= -bq + (b+h)(\mathbb{E}_G[(q - D)\mathbb{1}\{D \le q\}] + \mathbb{E}_G[D\mathbb{1}\{D < \lambda\}]) \\
&\quad + \sup_{\tilde q \in [\lambda, M]} \left\{ \tilde q (b - (b+h)G^-(\lambda)) \right\}, \qquad (36)
\end{aligned}
$$

where the supremum is attained at $\tilde q = M$, since $G^-(\lambda) < \rho \implies b - (b+h)G^-(\lambda) > 0$. Therefore:

$$
(36) = b(M - q) + (b+h)(\mathbb{E}_G[(q - D)\mathbb{1}\{D \le q\}] - \mathbb{E}_G[(M - D)\mathbb{1}\{D < \lambda\}]). \qquad (37)
$$

Now, if $q \ge \lambda$, for a fixed $\tilde q$ we have:

$$
\begin{aligned}
&\mathbb{E}_F[(q - D)\mathbb{1}\{D \le q\} - (\tilde q - D)\mathbb{1}\{D \le \tilde q\}] \\
&= \mathbb{E}_F[(q - \tilde q)\mathbb{1}\{D \le q\} - (\tilde q - D)\mathbb{1}\{q < D \le \tilde q\}] \\
&= \mathbb{E}_F[-(\tilde q - q)\mathbb{1}\{D \le q\} - (\tilde q - D)\mathbb{1}\{q < D \le \tilde q\}] \\
&= \mathbb{E}_F[-(\tilde q - q)\mathbb{1}\{D < \lambda\} - (\tilde q - q)\mathbb{1}\{\lambda \le D \le q\} - (\tilde q - D)\mathbb{1}\{q < D \le \tilde q\}] \\
&= -(\tilde q - q)G^-(\lambda) - (\tilde q - q)\mathbb{P}_F(\lambda \le D \le q) - \mathbb{E}_F\Big[(\tilde q - D)\mathbb{1}\{q < D \le \tilde q\}\Big]. \qquad (38)
\end{aligned}
$$

Since $0 < \tilde{q} - D < \tilde{q} - q$ for all $D \in (q, \tilde{q})$, the supremum of the above is achieved for F such that $\mathbb{P}_F(\lambda \leq D < \tilde{q}) = 0$. Then,

$$\begin{aligned}
&\sup_{\tilde{q} \in [q,M]} b(\tilde{q} - q) + (b+h)\left(\sup_{\substack{F \in \mathcal{F}(\lambda;G):\\ q_F^\star = \tilde{q}}} \mathbb{E}_F[(q-D)\mathbb{1}\{D \leq q\} - (\tilde{q} - D)\mathbb{1}\{D \leq \tilde{q}\}] \right) \\
&= \sup_{\tilde{q} \in [q,M]} b(\tilde{q} - q) - (b+h)(\tilde{q} - q)G^-(\lambda) \\
&= \sup_{\tilde{q} \in [q,M]} (\tilde{q} - q)(b - (b+h)G^-(\lambda)) \\
&= (M - q)(b - (b+h)G^-(\lambda)),
\end{aligned} \tag{39}$$

again, since $G^-(\lambda) < \rho$ by assumption.

(iii) $q = \tilde{q}$: In this case, $C_F(q) - C_F(\tilde{q}) = 0$ for all $F \in \mathcal{F}(\lambda; G)$.

Putting these three cases together and applying them to (31), we have:

$$\begin{aligned}
&\sup_{F \in \mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) \\
&= \begin{cases} b(M-q) + (b+h)(\mathbb{E}_G[(q-D)\mathbb{1}\{D \leq q\}] - \mathbb{E}_G[(M-D)\mathbb{1}\{D < \lambda\}]) & \text{if } q < \lambda \\ \max\{h(q-\lambda), (b-(b+h)G^-(\lambda))(M-q)\} & \text{if } q \geq \lambda. \end{cases}
\end{aligned}$$

Noting that $h(q-\lambda) > (b-(b+h)G^-(\lambda))(M-q) \iff q > q_G^\dagger$, with equality at $q_G^\dagger$, we obtain the result. $\square$

D.1.4. Proof of Theorem 1

Proof. We proceed separately for each regime.

Case I: $G^-(\lambda) \geq \rho$. By Lemma 1, $q_F^\star = q_G^\star < \lambda$ for all $F \in \mathcal{F}(\lambda; G)$. Therefore, for all $F \in \mathcal{F}(\lambda; G)$, $q \in [0, M]$:

$$C_F(q) - C_F(q_F^\star) = C_F(q) - C_F(q_G^\star).$$

Letting $q = q_G^\star$, this implies that $C_F(q_G^\star) - C_F(q_F^\star) = 0$ for all $F \in \mathcal{F}(\lambda; G)$. Using the fact that $\sup_{F \in \mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) \geq 0$ for all q by definition of $q_F^\star$, we then have $\Delta := \inf_q \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) = 0$, precisely achieved at $q^\Delta = q_G^\star$.

Case II: $G^-(\lambda) < \rho$. Recall, $q_G^\dagger := \frac{bM + h\lambda - (b+h)G^-(\lambda)M}{(b+h)(1-G^-(\lambda))}$ by definition. We analyze $\inf_{q \in [0,M]} \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star)$ by partitioning the proof into three cases: (i) $q < \lambda$, (ii) $q \in [\lambda, q_G^\dagger]$, and (iii) $q > q_G^\dagger$.

By Lemma 2, for $q < \lambda$:

$$\sup_{F \in \mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) = b(M-q) + (b+h)\Big[\mathbb{E}_G\Big[(q-D)\mathbb{1}\{D \leq q\}] - (M-D)\mathbb{1}\{D < \lambda\}\Big]\Big].$$

Hence, we have:

$$\begin{aligned}
&\inf_{q<\lambda} \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) \\
&= \inf_{q<\lambda} bM - (b+h)\mathbb{E}_G[(M-D)\mathbb{1}\{D < \lambda\}] - bq + (b+h)\mathbb{E}_G[(q-D)\mathbb{1}\{D \leq q\}],
\end{aligned} \tag{40}$$

where (40) uses the fact that $F \in \mathcal{F}(\lambda; G)$ for the first expectation. Now, the newsvendor cost can equivalently be written as

$$C_G(q) = b(\mathbb{E}_G[D] - q) + (b+h)\mathbb{E}_G[(q-D)\mathbb{1}\{D \le q\}]. \tag{41}$$

Plugging this formulation into (40), we obtain:

$$\inf_{q<\lambda} \sup_{F\in\mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) = bM - (b+h)\mathbb{E}_G[(M-D)\mathbb{1}\{D<\lambda\}] - b\mathbb{E}_G[D] + \inf_{q<\lambda} C_G(q). \tag{42}$$

Recall, $G^-(\lambda) < \rho$ implies $q_G^\star \ge \lambda$. Hence, by convexity of $C_G(q)$, $C_G(\lambda) \le C_G(q)$ for all $q < \lambda$. Then:

$$\begin{aligned}
&\inf_{q<\lambda} \sup_{F\in\mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) \\
&= bM - (b+h)\mathbb{E}_G[(M-D)\mathbb{1}\{D<\lambda\}] - b\mathbb{E}_G[D] + C_G(\lambda) \\
&= bM - (b+h)\mathbb{E}_G[(M-D)\mathbb{1}\{D<\lambda\}] - b\lambda + (b+h)\mathbb{E}_G[(\lambda-D)\mathbb{1}\{D\le\lambda\}],
\end{aligned}$$

where the second equality follows from (41). Simplifying, we finally obtain:

$$\begin{aligned}
\inf_{q<\lambda} \sup_{F\in\mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) &= b(M-\lambda) - (b+h)\mathbb{E}_G\Big[(M-\lambda)\mathbb{1}\{D<\lambda\}\Big] \\
&= (M-\lambda)(b-(b+h)G^-(\lambda)). \qquad (43)
\end{aligned}$$

Suppose now $q \in \left[\lambda, q_G^\dagger\right]$. Again, by Lemma 2:

$$\begin{aligned}
\min_{q\in[\lambda,q_G^\dagger]} \sup_{F\in\mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) &= \min_{q\in[\lambda,q_G^\dagger]} (M-q)(b-(b+h)G^-(\lambda)) \\
&= (M-q_G^\dagger)(b-(b+h)G^-(\lambda)), \qquad (44)
\end{aligned}$$

where the final equality follows from the fact that the right-hand side of (44) is decreasing in q, since $G^-(\lambda) < \rho$. It is easy to derive that $q_G^\dagger \ge \lambda$ for $G^-(\lambda) < \rho$. Comparing (43) and (44), this implies:

$$\min_{q\in[\lambda,q_G^\dagger]} \sup_{F\in\mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) \le \inf_{q<\lambda} \sup_{F\in\mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star).$$

Hence, setting $q = q_G^\dagger$ (weakly) dominates setting $q < \lambda$.

We conclude by considering $q \in (q_G^\dagger, M]$. By Lemma 2:

$$\inf_{q\in(q_G^\dagger,M]} \sup_{F\in\mathcal{F}(\lambda;G)} C_F(q) - C_F(q_F^\star) = \inf_{q\in(q_G^\dagger,M]} h(q-\lambda) = h(q_G^\dagger - \lambda).$$

By definition of $q_G^\dagger$,

$$h(q_G^\dagger - \lambda) = (M-q_G^\dagger)(b-(b+h)G^-(\lambda)) = \frac{h(b-(b+h)G^-(\lambda)(M-\lambda)}{(b+h)(1-G^-(\lambda))}.$$

Putting this together with (44), we conclude that $q^\Delta = q_G^\dagger$, with

$$\Delta = \frac{h(b-(b+h)G^-(\lambda)(M-\lambda)}{(b+h)(1-G^-(\lambda))}.$$

$\square$

D.2. Worst-Case Nature of Two-Point Mass Distributions

D.2.1. Proof of Proposition 2

Proof. This fact is a corollary of the derivation of Regret(q) in the proof of Lemma 2. To show this, recall our three cases: (i) $q > \tilde{q} \geq \lambda$, (ii) $q < \lambda \leq \tilde{q}$, and (iii) $\lambda \leq q < \tilde{q}$, where $\tilde{q} := q_F^\star$.

(i) $q > \tilde{q} \geq \lambda$: In this case, we showed that any worst-case distribution F is such that $\tilde{q} = \lambda$, and $\mathbb{P}_F(D \leq \lambda) = 1$. Therefore, the worst-case distribution places the entirety of the remaining mass at λ.

(ii) $q < \lambda \leq \tilde{q}$: In this case, we argued that any worst-case distribution F is such that $\mathbb{P}_F(\lambda \leq D < \tilde{q}) = 0$ and $\tilde{q} = M$. This is satisfied by a distribution F that places the entirety of the remaining mass at M.

(iii) $\lambda \leq q < \tilde{q}$: Here we similarly argued that any worst-case distribution F is such that $\mathbb{P}_F(\lambda \leq D < \tilde{q}) = 0$ and $\tilde{q} = M$. Again, this is achieved by a distribution F that places the entirety of the remaining mass on M.

□

Interested readers may refer to `https://anonymous.4open.science/r/rcn_msom_sig` for the proofs of all remaining results.

D.3. Proof of Theorem 2

Proof. Proposition 6 below upper bounds our algorithm's regret by (i) the minimax risk Δ, and (ii) the cost difference between q^{alg} and q^{Δ}, for any distribution $F \in \mathcal{F}(\lambda; G)$. We defer its proof to Appendix D.3.1.

PROPOSITION 6. *The regret of Algorithm 1 is upper bounded by:*

$$\text{Regret}(q^{\text{alg}}) \leq \Delta + \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^{\text{alg}}) - C_F(q^{\Delta}). \tag{45}$$

Hence, the remainder of the proof focuses on bounding the second term. Before doing so, we show a series of useful facts. First off, Lemma 5 establishes Lipschitzness of the newsvendor cost function. Its proof can be found in Appendix D.3.2.

LEMMA 5. *For any distribution F, and ordering quantities q, q':*

$$|C_F(q) - C_F(q')| \leq \max\{b, h\}|q - q'|.$$

Moreover, by Fact 1, $\mathbb{1}\left\{s_i^{\text{off},\lambda} < \lambda\right\} = \mathbb{1}\left\{d_i^{\text{off}} < \lambda\right\}$ for all $i \in [N]$. We use this to prove Lemma 6, which establishes that $\widehat{G}^-(\lambda)$ and $G^-(\lambda)$ are close with constant probability, below. We defer its proof to Appendix D.3.3.

LEMMA 6. *Let $\mathcal{E} = \left\{|\widehat{G}^-(\lambda) - G^-(\lambda)| < \zeta\right\}$. Then, $\mathbb{P}_G(\mathcal{E}) \geq 1 - \delta$.*

We moreover use Fact 1 to show that $q^{\star}_{\widehat{G}}$ is precisely equal to the actual SAA of $q^{\star}_G$. To formalize this, we introduce some additional notation. Let $\widehat{G}^{\text{uc}}(x) = \frac{1}{N}\sum_{i \in [N]} \mathbb{1}\left\{d_i^{\text{off}} \leq x\right\}$ be the empirical cdf of the (true) uncensored demand, and $q^{\star}_{\widehat{G}^{\text{uc}}} = \inf\{x : \widehat{G}^{\text{uc}}(x) \geq \rho\}$ be the corresponding newsvendor quantile. We repeatedly rely on the following lemma in the remainder of the proof. Its proof can be found in Appendix D.3.4.

LEMMA 7. *Suppose $\widehat{G}^-(\lambda) \geq \rho$. Then, $q^{\star}_{\widehat{G}} = q^{\star}_{\widehat{G}^{\text{uc}}} < \lambda$.*

Having established these useful facts, we may begin with the proof of our upper bound. We prove each of the four regret bounds separately, beginning with the two "extreme" cases of $G^-(\lambda) \geq \rho + 2\zeta$ and $G^-(\lambda) < \rho - 2\zeta$. Moreover, in the remainder of the proof we condition on the good event $\mathcal{E}$ defined in Lemma 6.

Case I: $G^-(\lambda) \geq \rho + 2\zeta$. Under event $\mathcal{E}$, $\widehat{G}^-(\lambda) > \rho + \zeta$, which implies $q^{\text{alg}} = q^{\star}_{\widehat{G}}$ by construction. Since $\widehat{G}^-(\lambda) \geq \rho$, by Lemma 7, $q^{\star}_{\widehat{G}} = q^{\star}_{\widehat{G}^{\text{uc}}} < \lambda$. By Proposition 5, then,

$$\text{Regret}(q^{\text{alg}}) \leq C_G(q^{\star}_{\widehat{G}^{\text{uc}}}) - C_G(q^{\star}_G).$$

Hence, bounding the regret of q^{alg} in this region reduces to the question of bounding the regret of the SAA of $q^{\star}_G$.

The following lemma, whose proof is adapted from Chen and Ma (2024) (see Appendix D.3.5), allows us to bound this SAA error.

LEMMA 8. *Suppose $G^-(\lambda) \geq \rho$, and $\widehat{G}^-(\lambda) \geq \rho$. With probability at least $1 - \delta$,*

$$C_G(q^{\star}_{\widehat{G}^{\text{uc}}}) - C_G(q^{\star}_G) \leq \lambda(b + h)\sqrt{\frac{\log(2/\delta)}{2N}}.$$

Taking a union bound over the event $\mathcal{E}$ and this latter good event, we obtain the first case in the theorem statement.

Case II: $G^-(\lambda) < \rho - 2\zeta$. By Theorem 1, $q^\Delta = q^\dagger_G$ in this case. Moreover, under event $\mathcal{E}$, $\widehat{G}^-(\lambda) < \rho - \zeta$, and our algorithm outputs $q^{\text{alg}} = q^\dagger_{\widehat{G}}$. Then, by Proposition 6:

$$\text{Regret}(q^{\text{alg}}) \le \Delta + \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^\dagger_{\widehat{G}}) - C_F(q^\dagger_G) \le \Delta + \max\{b,h\}|q^\dagger_{\widehat{G}} - q^\dagger_G|, \tag{46}$$

where the second inequality follows from the fact that C_F is $\max\{b,h\}$-Lipschitz (Lemma 5). It suffices then to bound the error in our algorithm's estimate of $q^\dagger_G$. Lemma 9, whose proof can be found in Appendix D.3.6, leverages the fact that $|G^-(\lambda) - \widehat{G}^-(\lambda)| < \zeta$ under event $\mathcal{E}$ to bound this error.

LEMMA 9. *Under event $\mathcal{E}$, $|q^\dagger_G - q^\dagger_{\widehat{G}}| \le \frac{M-\lambda}{1-\rho}\zeta$.*

Plugging this back into (46), we obtain:

$$\text{Regret}(q^{\text{alg}}) \le \Delta + \frac{\max\{b,h\}(M-\lambda)}{1-\rho}\zeta, \tag{47}$$

completing the proof of this case in the theorem statement.

We now turn our attention to the remaining case, where $G^-(\lambda) \in [\rho - 2\zeta, \rho + 2\zeta)$, further partitioning the analysis based on whether $G^-(\lambda) \ge \rho$.

Case III: $G^-(\lambda) \in [\rho, \rho + 2\zeta)$. By Theorem 1, $q^\Delta = q^\star_G$, and $\Delta = 0$. Moreover, given $\mathcal{E}$, $\widehat{G}^-(\lambda) \in (\rho - \zeta, \rho + 3\zeta)$.

Suppose first that $\widehat{G}^-(\lambda) \in [\rho + \zeta, \rho + 3\zeta)$. Then, $q^{\text{alg}} = q^\star_{\widehat{G}} < \lambda$. Identical arguments as those used for Case I yield:

$$\text{Regret}(q^{\text{alg}}) \le C_G(q^\star_{\widehat{G}^{\text{uc}}}) - C_G(q^\star_G) \le \lambda(b+h)\zeta. \tag{48}$$

Suppose now that $\widehat{G}^-(\lambda) \in (\rho - \zeta, \rho + \zeta)$. In this case, $q^{\text{alg}} = \lambda$. Using the fact that $q^\Delta = q^\star_G < \lambda$ and $\Delta = 0$, by Proposition 6, we have:

$$\text{Regret}(q^{\text{alg}}) \le \Delta + \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^{\text{alg}}) - C_F(q^\Delta) = \sup_{F \in \mathcal{F}(\lambda;G)} C_F(\lambda) - C_F(q^\star_G).$$

Proposition 5 and Lemma 1 together imply that:

$$\begin{aligned}
&\sup_{F \in \mathcal{F}(\lambda;G)} C_F(\lambda) - C_F(q^\star_G) \\
&= b(q^\star_G - \lambda) + (b+h)\mathbb{E}_G\left[(\lambda - D)\mathbb{1}\{D < \lambda\} - (q^\star_G - D)\mathbb{1}\left\{D \le q^\star_G\right\}\right] \\
&= b(q^\star_G - \lambda) + (b+h)\mathbb{E}_G\left[(\lambda - q^\star_G)\mathbb{1}\{D < \lambda\} + (q^\star_G - D)\mathbb{1}\{D < \lambda\} - (q^\star_G - D)\mathbb{1}\left\{D \le q^\star_G\right\}\right] \\
&= b(q^\star_G - \lambda) + (b+h)\mathbb{E}_G\left[(\lambda - q^\star_G)\mathbb{1}\{D < \lambda\} + (q^\star_G - D)\mathbb{1}\left\{q^\star_G < D < \lambda\right\}\right] \\
&\le (\lambda - q^\star_G)\Big((b+h)G^-(\lambda) - b\Big) \\
&= (b+h)(\lambda - q^\star_G)(G^-(\lambda) - \rho).
\end{aligned}$$

Using the fact that $q^\star_G \in [0, \lambda)$ for $G^-(\lambda) \ge \rho$, and $G^-(\lambda) \le \rho + 2\zeta$ by assumption, we obtain the final bound of:

$$\text{Regret}(q^{\text{alg}}) \le 2\lambda(b+h)\zeta.$$

We proceed to our final case.

Case IV: $G^-(\lambda) \in [\rho - 2\zeta, \rho)$. By Theorem 1, $q^\Delta = q^\dagger_G$. Moreover, under $\mathcal{E}$, $\widehat{G}^-(\lambda) \in (\rho - 3\zeta, \rho + \zeta)$.

Suppose first that $\widehat{G}^-(\lambda) \in (\rho - 3\zeta, \rho - \zeta)$. In this case, the algorithm outputs $q^{\mathrm{alg}} = q^\dagger_{\widehat{G}}$. By the same arguments as those used in Case II, we obtain:

$$\mathrm{Regret}(q^{\mathrm{alg}}) \le \Delta + \frac{\max\{b,h\}(M-\lambda)}{(1-\rho)}\zeta.$$

Suppose now that $\widehat{G}^-(\lambda) \in [\rho - \zeta, \rho + \zeta)$. In this case, $q^{\mathrm{alg}} = \lambda$. Again, leveraging Proposition 6, we have:

$$\mathrm{Regret}(q^{\mathrm{alg}}) \le \Delta + \sup_{F \in \mathcal{F}(\lambda;G)} C_F(\lambda) - C_F(q^\dagger_G) \le \Delta + \max\{b,h\}|\lambda - q^\dagger_G|, \tag{49}$$

where the final inequality follows from Lipschitzness of the newsvendor cost function (Lemma 5). Plugging in the definition of $q^\dagger_G$, algebra gives us that:

$$\begin{aligned}
|\lambda - q^\dagger_G| &= \frac{(M-\lambda)(b-(b+h)G^-(\lambda))}{(b+h)(1-G^-(\lambda))} \\
&\le \frac{(M-\lambda)(b-(b+h)(\rho-2\zeta))}{(b+h)(1-\rho)} \\
&= \frac{2(M-\lambda)(b+h)\zeta}{h} \\
&= \frac{2(M-\lambda)}{1-\rho}\zeta,
\end{aligned}$$

where the first inequality uses the fact that $G^-(\lambda) \in [\rho - 2\zeta, \rho)$, and the subsequent equalities plug in the definition of $\rho = b/(b+h)$. Plugging this bound back into (49), we obtain:

$$\mathrm{Regret}(q^{\mathrm{alg}}) \le \Delta + \max\{b,h\}\frac{2(M-\lambda)}{1-\rho}\zeta,$$

concluding the proof of the theorem. □

D.3.1. Proof of Proposition 6

Proof. Adding and subtracting $C_F(q^\Delta)$ from the definition of regret, we have:

$$\begin{aligned}
\mathrm{Regret}(q^{\mathrm{alg}}) &= \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^{\mathrm{alg}}) - C_F(q^\Delta) + C_F(q^\Delta) - C_F(q^\star_F) \\
&\le \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^{\mathrm{alg}}) - C_F(q^\Delta) + \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^\Delta) - C_F(q^\star_F) \\
&= \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^{\mathrm{alg}}) - C_F(q^\Delta) + \Delta,
\end{aligned}$$

where the final equality uses the definition of $\Delta = \sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^\Delta) - C_F(q^\star_F)$. □

D.3.2. Proof of Lemma 5

Proof. By definition:

$$|C_F(q) - C_F(q')| = \left| \mathbb{E}_F\left[b\left((D-q)^+ - (D-q')^+\right) + h\left((q-D)^+ - (q'-D)^+\right)\right] \right|.$$

Without loss of generality, suppose $q < q'$. Consider first the term $b((D-q)^+ - (D-q')^+)$. We have:

$$(D-q)^+ - (D-q')^+ = \begin{cases} q' - q & \text{if } D \ge q' \\ 0 & \text{if } D \le q \\ D - q & \text{if } D \in (q, q'). \end{cases}$$

Similarly:

$$(q-D)^+ - (q'-D)^+ = \begin{cases} 0 & \text{if } D \geq q' \\ q - q' & \text{if } D \leq q \\ D - q' & \text{if } D \in (q, q'). \end{cases}$$

Putting these two together, we have:

$$b\big((D-q)^+ - (D-q')^+\big) + h\big((q-D)^+ - (q'-D)^+\big) = \begin{cases} b(q'-q) & \text{if } D \geq q' \\ h(q-q') & \text{if } D \leq q \\ b(D-q) + h(D-q') & \text{if } D \in (q, q') \end{cases}$$

$$\Longrightarrow \left| \mathbb{E}_F\left[b\big((D-q)^+ - (D-q')^+\big) + h\big((q-D)^+ - (q'-D)^+\big)\right] \right| \leq \max\{b, h\}|q - q'|.$$

□

D.3.3. Proof of Lemma 6

Proof. Since $\mathbb{1}\left\{s_i^{\mathrm{off},\lambda} < \lambda\right\} = \mathbb{1}\left\{d_i^{\mathrm{off}} < \lambda\right\}$ for all i, we have:

$$\widehat{G}^-(\lambda) = \frac{1}{N} \sum_{i \in [N]} \mathbb{1}\left\{d_i^{\mathrm{off}} < \lambda\right\} \Longrightarrow \mathbb{E}_G[\widehat{G}^-(\lambda)] = G^-(\lambda).$$

Hence, by Hoeffding's inequality (Boucheron et al. 2013):

$$\mathbb{P}_G(|\widehat{G}^-(\lambda) - G^-(\lambda)| \geq \zeta) \leq 2\exp(-2N\zeta^2).$$

Letting $\zeta = \sqrt{\frac{\log(2/\delta)}{2N}}$, we obtain the result. □

D.3.4. Proof of Lemma 7

Proof. We first show that $\widehat{G}^-(\lambda) \geq \rho$ implies $q^\star_{\widehat{G}^{\mathrm{uc}}} < \lambda$. Again, since $\mathbb{1}\left\{s_i^{\mathrm{off},\lambda} < \lambda\right\} = \mathbb{1}\left\{d_i^{\mathrm{off}} < \lambda\right\}$,

$$\frac{1}{N} \sum_{i=1}^N \mathbb{1}\left\{d_i^{\mathrm{off}} < \lambda\right\} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}\left\{s_i^{\mathrm{off},\lambda} < \lambda\right\} = \widehat{G}^-(\lambda) \geq \rho,$$

by assumption. Then, it must be that $\widehat{G}^{\mathrm{uc}}(x) \geq \rho$ for some $x < \lambda$, which implies that $q^\star_{\widehat{G}^{\mathrm{uc}}} < \lambda$.

We use this to show that $q^\star_{\widehat{G}} = q^\star_{\widehat{G}^{\mathrm{uc}}}$. By definition, $q^\star_{\widehat{G}^{\mathrm{uc}}}$ satisfies:

$$\begin{cases} \frac{1}{N} \sum_{i \in [N]} \mathbb{1}\left\{d_i^{\mathrm{off}} \leq x\right\} < \rho & \forall\, x < q^\star_{\widehat{G}^{\mathrm{uc}}} \\ \frac{1}{N} \sum_{i \in [N]} \mathbb{1}\left\{d_i^{\mathrm{off}} \leq x\right\} \geq \rho & \forall\, x \geq q^\star_{\widehat{G}^{\mathrm{uc}}}. \end{cases}$$

Fix $x \geq q^\star_{\widehat{G}^{\mathrm{uc}}}$. Since $s_i^{\mathrm{off},\lambda} \leq d_i^{\mathrm{off}}$ for all i,

$$\frac{1}{N} \sum_{i \in [N]} \mathbb{1}\left\{s_i^{\mathrm{off},\lambda} \leq x\right\} \geq \frac{1}{N} \sum_{i \in [N]} \mathbb{1}\left\{d_i^{\mathrm{off}} \leq x\right\} \geq \frac{1}{N} \sum_{i \in [N]} \mathbb{1}\left\{d_i^{\mathrm{off}} \leq q^\star_{\widehat{G}^{\mathrm{uc}}}\right\} \geq \rho,$$

by definition of $q^\star_{\widehat{G}^{\mathrm{uc}}}$.

Now, fix $x < q^\star_{\widehat{G}^{\mathrm{uc}}}$. Since $x < q^\star_{\widehat{G}^{\mathrm{uc}}} < \lambda$, if $d_i^{\mathrm{off}} \geq \lambda$, $s_i^{\mathrm{off},\lambda} = \lambda$, which then implies that $\mathbb{1}\left\{s_i^{\mathrm{off},\lambda} \leq x\right\} = 0$. We moreover have in this case that $\mathbb{1}\left\{d_i^{\mathrm{off}} \leq x\right\} = 0$. Putting these two together, we have that $\mathbb{1}\left\{s_i^{\mathrm{off},\lambda} \leq x\right\} = \mathbb{1}\left\{d_i^{\mathrm{off}} \leq x\right\}$ if $d_i^{\mathrm{off}} \geq \lambda$.

If $d_i^{\text{off}} < \lambda$, on the other hand, $s_i^{\text{off},\lambda} = d_i^{\text{off}}$ by definition. Therefore, $\mathbb{1}\left\{s_i^{\text{off},\lambda} \le x\right\} = \mathbb{1}\left\{d_i^{\text{off}} \le x\right\}$ whenever $d_i^{\text{off}} < \lambda$.

Since $\mathbb{1}\left\{s_i^{\text{off},\lambda} \le x\right\} = \mathbb{1}\left\{d_i^{\text{off}} \le x\right\}$ whenever $x < q^\star_{\widehat{G}^{\text{uc}}}$, we obtain:

$$\frac{1}{N}\sum_{i\in[N]} \mathbb{1}\left\{s_i^{\text{off},\lambda} \le x\right\} = \frac{1}{N}\sum_{i\in[N]} \mathbb{1}\left\{d_i^{\text{off}} \le x\right\} < \rho \quad \forall\, x < q^\star_{\widehat{G}^{\text{uc}}}.$$

Putting these two facts together, we have:

$$\begin{cases} \frac{1}{N}\sum_{i\in[N]} \mathbb{1}\left\{s_i^{\text{off},\lambda} \le x\right\} < \rho & \forall\, x < q^\star_{\widehat{G}^{\text{uc}}} \\ \frac{1}{N}\sum_{i\in[N]} \mathbb{1}\left\{s_i^{\text{off},\lambda} \le x\right\} \ge \rho & \forall\, x \ge q^\star_{\widehat{G}^{\text{uc}}}, \end{cases}$$

which implies that $q^\star_{\widehat{G}} = q^\star_{\widehat{G}^{\text{uc}}}$. □

D.3.5. Proof of Lemma 8

Proof. The proof is modified from Chen and Ma (2024).

By the Dvoretzky–Kiefer–Wolfowitz–Massart inequality (DKW) inequality (Massart 1990):

$$\mathbb{P}\left[\sup_{a\ge 0}|\widehat{G}^{\text{uc}}(a) - G(a)| \le \sqrt{\frac{\log(2/\delta)}{2N}}\right] \ge 1 - 2\exp\left(-2N\left(\sqrt{\frac{\log(2/\delta)}{2N}}\right)^2\right) = 1-\delta.$$

Therefore with probability at least $1-\delta$:

$$\sup_{a\ge 0}|\widehat{G}^{\text{uc}}(a) - G(a)| \le \sqrt{\frac{\log(2/\delta)}{2N}}. \tag{50}$$

We condition our analysis on the above event, partitioning the proof based on (i) $q^\star_{\widehat{G}^{\text{uc}}} \le q^\star_G$, and (ii) $q^\star_{\widehat{G}^{\text{uc}}} > q^\star_G$.

Case I: $q^\star_{\widehat{G}^{\text{uc}}} \le q^\star_G$. By Proposition 3:

$$\begin{aligned} &C_G(q^\star_{\widehat{G}^{\text{uc}}}) - C_G(q^\star_G) \\ &= b(q^\star_G - q^\star_{\widehat{G}^{\text{uc}}}) + (b+h)\mathbb{E}_G\left[(q^\star_{\widehat{G}^{\text{uc}}} - D)\mathbb{1}\left\{D \le q^\star_{\widehat{G}^{\text{uc}}}\right\} - (q^\star_G - D)\mathbb{1}\left\{D \le q^\star_G\right\}\right] \\ &= b(q^\star_G - q^\star_{\widehat{G}^{\text{uc}}}) + (b+h)\mathbb{E}_G\left[(q^\star_{\widehat{G}^{\text{uc}}} - q^\star_G)\mathbb{1}\left\{D \le q^\star_{\widehat{G}^{\text{uc}}}\right\} - (q^\star_G - D)\mathbb{1}\left\{q^\star_{\widehat{G}^{\text{uc}}} < D \le q^\star_G\right\}\right] \\ &\le (b+h)(q^\star_G - q^\star_{\widehat{G}^{\text{uc}}})(\rho - G(q^\star_{\widehat{G}^{\text{uc}}})), \end{aligned} \tag{51}$$

where the upper bound follows from the fact that $(q^\star_G - D)\mathbb{1}\left\{q^\star_{\widehat{G}^{\text{uc}}} < D \le q^\star_G\right\} \ge 0$, and $b = (b+h)\rho$ by definition. Moreover:

$$\rho - G(q^\star_{\widehat{G}^{\text{uc}}}) = \rho - \widehat{G}^{\text{uc}}(q^\star_{\widehat{G}^{\text{uc}}}) + \widehat{G}^{\text{uc}}(q^\star_{\widehat{G}^{\text{uc}}}) - G(q^\star_{\widehat{G}^{\text{uc}}}) \le \sup_{a\ge 0}|\widehat{G}^{\text{uc}}(a) - G(a)|, \tag{52}$$

where the inequality follows from $\widehat{G}^{\text{uc}}(q^\star_{\widehat{G}^{\text{uc}}}) \ge \rho$, by definition of $q^\star_{\widehat{G}^{\text{uc}}}$.

Using this fact in (51), we have:

$$\begin{aligned} C_G(q^\star_{\widehat{G}^{\text{uc}}}) - C_G(q^\star_G) &\le (b+h)(q^\star_G - q^\star_{\widehat{G}^{\text{uc}}})(\rho - G(q^\star_{\widehat{G}^{\text{uc}}})) \\ &\le (b+h)\cdot|q^\star_G - q^\star_{\widehat{G}^{\text{uc}}}|\cdot\sup_{a\ge 0}|\widehat{G}^{\text{uc}}(a) - G(a)|. \end{aligned}$$

Case II: $q^\star_{\widehat{G}^{\mathrm{uc}}} > q^\star_G$. Again, by Proposition 3:

$$\begin{aligned}
&C_G(q^\star_{\widehat{G}^{\mathrm{uc}}}) - C_G(q^\star_G) \\
&= b(q^\star_G - q^\star_{\widehat{G}^{\mathrm{uc}}}) + (b+h)\mathbb{E}_G\left[(q^\star_{\widehat{G}^{\mathrm{uc}}} - D)\mathbb{1}\left\{D \le q^\star_{\widehat{G}^{\mathrm{uc}}}\right\} - (q^\star_G - D)\mathbb{1}\left\{D \le q^\star_G\right\}\right] \\
&= -\rho(b+h)(q^\star_{\widehat{G}^{\mathrm{uc}}} - q^\star_G) + (b+h)\mathbb{E}_G\left[(q^\star_{\widehat{G}^{\mathrm{uc}}} - q^\star_G)\mathbb{1}\left\{D \le q^\star_G\right\} + (q^\star_{\widehat{G}^{\mathrm{uc}}} - D)\mathbb{1}\left\{q^\star_G < D < q^\star_{\widehat{G}^{\mathrm{uc}}}\right\}\right] \\
&\le (b+h)(q^\star_{\widehat{G}^{\mathrm{uc}}} - q^\star_G)\left(-\rho + G^-\left(q^\star_{\widehat{G}^{\mathrm{uc}}}\right)\right),
\end{aligned} \tag{53}$$

where the inequality follows from $q^\star_{\widehat{G}^{\mathrm{uc}}} - D < q^\star_{\widehat{G}^{\mathrm{uc}}} - q^\star_G$ for all $D > q^\star_G$.

For all $q < q^\star_{\widehat{G}^{\mathrm{uc}}}$:

$$G(q) - \rho = G(q) - \widehat{G}^{\mathrm{uc}}(q) + \widehat{G}^{\mathrm{uc}}(q) - \rho \le \sup_{a\ge 0}|\widehat{G}^{\mathrm{uc}}(a) - G(a)|,$$

where the inequality follows from the fact that $\widehat{G}^{\mathrm{uc}}(q) < \rho$ for all $q < q^\star_{\widehat{G}^{\mathrm{uc}}}$, by definition of $q^\star_{\widehat{G}^{\mathrm{uc}}}$. We similarly use this to obtain the final upper bound:

$$C_G(q^\star_{\widehat{G}^{\mathrm{uc}}}) - C_G(q^\star_G) \le (b+h)\cdot|q^\star_{\widehat{G}^{\mathrm{uc}}} - q^\star_G|\cdot\sup_{a\ge 0}|\widehat{G}^{\mathrm{uc}}(a) - G(a)|.$$

Applying (50) to both cases, and using the fact that both $q^\star_G \le \lambda$ and $q^\star_{\widehat{G}^{\mathrm{uc}}} \le \lambda$ under the assumption that $G^-(\lambda) \ge \rho$ and $\widehat{G}^-(\lambda) \ge \rho$, we obtain the claim. □

D.3.6. Proof of Lemma 9

Proof. Consider the function $Q^\dagger(x) = \frac{bM + h\lambda - (b+h)Mx}{(b+h)(1-x)}$. Observe that $q^\dagger_G = Q^\dagger(G^-(\lambda))$, and $q^\dagger_{\widehat{G}} = Q^\dagger(\widehat{G}^-(\lambda))$ by definition. For any $x \in [0, \rho - \zeta]$:

$$\frac{dQ^\dagger}{dx} = \frac{h(\lambda - M)}{(b+h)(1-x)^2} \implies \left|\frac{dQ^\dagger}{dx}\right| \le \frac{h(M-\lambda)}{(b+h)(1-\rho+\zeta)^2}.$$

Hence $Q^\dagger$ is $\frac{h(M-\lambda)}{(b+h)(1-\rho+\zeta)^2}$-Lipschitz, which implies

$$|q^\dagger_G - q^\dagger_{\widehat{G}}| \le \frac{h(M-\lambda)}{(b+h)(1-\rho+\zeta)^2}|G^-(\lambda) - \widehat{G}^-(\lambda)| \le \frac{h(M-\lambda)}{(b+h)(1-\rho+\zeta)^2}\zeta = \frac{(1-\rho)(M-\lambda)}{(1-\rho+\zeta)^2}\zeta$$

under event $\mathcal{E}$. Using the fact that since $\rho < 1$ and $\zeta > 0$, $(1-\rho)^2 \le (1-\rho+\zeta)^2$, which completes the proof of the claim. □

D.4. Proof of Corollary 2

Proof. For ease of notation, we let $R^{\mathrm{id}} = 2\lambda(b+h)\zeta$ be the upper bound on the regret of RCN derived in Theorem 2; we also let $R^{\mathrm{id}}_{\max} = \max_{q\in[0,M]}\sup_{F\in\mathcal{F}(\lambda;G)}\mathrm{Regret}(q)$ when $G^-(\lambda) \ge \rho$. Similarly, let $R^{\mathrm{ui}} = \Delta + 2\max\left\{\frac{b}{h}, 1\right\}(M-\lambda)(b+h)\zeta$ and $R^{\mathrm{ui}}_{\max} = \max_{q\in[0,M]}\sup_{F\in\mathcal{F}(\lambda;G)}\mathrm{Regret}(q)$ when $G^-(\lambda) < \rho$. Applying the bounds from Theorem 2, we have:

$$\mathbb{E}_G[\mathrm{Regret}(q^{\mathrm{alg}})] \le \begin{cases} R^{\mathrm{id}} + R^{\mathrm{id}}_{\max}\cdot 2\delta & \text{if } G^-(\lambda) \ge \rho \\ R^{\mathrm{ui}} + R^{\mathrm{ui}}_{\max}\cdot 2\delta & \text{if } G^-(\lambda) < \rho. \end{cases} \tag{54}$$

We first bound $R^{\text{id}}_{\max}$. By Proposition 5, if $q < \lambda$, $\text{Regret}(q) = C_G(q) - C_G(q^\star_G)$. By Lemma 5:

$$C_G(q) - C_G(q^\star_G) \le \max\{b, h\}|q - q^\star_G| \le (b+h)\max\{\rho, 1-\rho\}\lambda.$$

If $q \ge \lambda$, by Proposition 5,

$$\begin{aligned}
\text{Regret}(q) &= b(q^\star_G - q) + (b+h)\left[(q-\lambda) + \mathbb{E}_G\left[(\lambda - D)\mathbb{1}\{D < \lambda\} - (q^\star_G - D)\mathbb{1}\left\{D \le q^\star_G\right\}\right]\right] \\
&= bq^\star_G + hq - (b+h)\lambda + (b+h)\mathbb{E}_G\left[(\lambda - D)\mathbb{1}\{D < \lambda\} - (q^\star_G - D)\mathbb{1}\left\{D \le q^\star_G\right\}\right] \\
&\le bq^\star_G + hq + (b+h)\left(-\lambda + \lambda G^-(\lambda) - q^\star_G\rho - \mathbb{E}_G\left[D\mathbb{1}\left\{D \in (q^\star_G, \lambda)\right\}\right]\right).
\end{aligned}$$

Using the fact that $(b+h)\rho q^\star_G = bq^\star_G$ and $\mathbb{E}_G\left[D\mathbb{1}\left\{D \in (q^\star_G, \lambda)\right\}\right] \ge 0$, we obtain:

$$\text{Regret}(q) \le hM - (b+h)\lambda(1 - G^-(\lambda)) \le hM.$$

Putting these two bounds together, we obtain:

$$R^{\text{id}}_{\max} \le (b+h)\max\{\max\{\rho, (1-\rho)\}\lambda, (1-\rho)M\} = (b+h)\max\{\rho\lambda, (1-\rho)M\}.$$

We now bound $R^{\text{ui}}_{\max}$. By Lemma 2, for $q < \lambda$,

$$\text{Regret}(q) = b(M - q) + (b+h)\left[\mathbb{E}_G\left[(q - D)\mathbb{1}\{D \le q\} - (M - D)\mathbb{1}\{D < \lambda\}\right]\right] \le bM,$$

where the inequality follows from the fact that $-bq + (b+h)\mathbb{E}_G[(q-D)\mathbb{1}\{D \le q\}]$ is decreasing in q for $q < \lambda < q^\star_G$.

For $q \in \left[\lambda, q^\dagger_G\right]$,

$$\text{Regret}(q) = (b - (b+h)G^-(\lambda))(M - q) \le (b - (b+h)G^-(\lambda))(M - \lambda).$$

Finally, for $q > q^\dagger_G$,

$$\text{Regret}(q) = h(q - \lambda) \le h(M - \lambda).$$

Putting these three bounds together, we obtain:

$$R^{\text{ui}}_{\max} \le (b+h)\max\{\rho M, (\rho - G^-(\lambda))(M-\lambda), (1-\rho)(M-\lambda)\} = (b+h)\max\{\rho M, (1-\rho)(M-\lambda)\}.$$

We conclude by plugging these bounds back into (54). Since $\zeta = \sqrt{\log(2/\delta)/2N} = c'\sqrt{\log N/N}$ for some $c' > 0$ and $\delta = c/\sqrt{N}$, if $G^-(\lambda) \ge \rho$, we have:

$$\mathbb{E}_G[\text{Regret}(q^{\text{alg}})] \le 2c'(b+h)(\lambda + \max\{\rho\lambda, (1-\rho)M\})\sqrt{\log N/N}.$$

If $G^-(\lambda) < \rho$:

$$\mathbb{E}_G[\text{Regret}(q^{\text{alg}})] \le \Delta + 2c'(b+h)\left(\max\left\{\frac{\rho}{1-\rho}, 1\right\}(M - \lambda) + \max\{\rho M, (1-\rho)(M-\lambda)\}\right)\sqrt{\log N/N}.$$

Renaming $c' = 2c'$, we obtain the result. □

D.5. Proof of Theorem 3

Proof. We consider an instance for which $\lambda \in (0,1)$, and present the set of "hard" distributions associated with each regime:

(i) strictly unidentifiable regime:

$$G_0^{\text{ui}} = \text{Ber}(1-\rho+\delta_0^{\text{ui}}), \quad G_1^{\text{ui}} = \text{Ber}(1-\rho+\delta_0^{\text{ui}}+\delta_1^{\text{ui}}),$$

where $\delta_0^{\text{ui}} = \min\left\{\frac{\rho}{2}, \frac{1-\rho}{2}\right\}$ and $\delta_1^{\text{ui}} = \min\left\{\frac{\rho}{4}, \frac{3\rho-1}{4}, \frac{1}{2}\sqrt{\frac{1-\rho}{N}}\right\}$. For these distributions, $G_0^{\text{ui},-}(\lambda) = \rho - \delta_0^{\text{ui}}$, and $G_1^{\text{ui},-}(\lambda) = \rho - (\delta_0^{\text{ui}} + \delta_1^{\text{ui}})$.

(ii) knife-edge regime:

$$G_0^{\text{ke}} = \text{Ber}(1-\rho+\delta^{\text{ke}}), \quad G_1^{\text{ke}} = \text{Ber}(1-\rho-\delta^{\text{ke}}),$$

where $\delta^{\text{ke}} = \min\left\{\frac{\rho}{2}, \frac{1-\rho}{2}, \frac{1}{4}\sqrt{\frac{1-\rho}{N}}\right\}$. Here, $G_0^{\text{ke},-}(\lambda) = \rho - \delta^{\text{ke}}$, $G_1^{\text{ke},-}(\lambda) = \rho + \delta^{\text{ke}}$.

(iii) strictly identifiable regime: $G_0^{\text{id}}, G_1^{\text{id}}$ are respectively defined by cdf's:

$$G_0^{\text{id}}(x) = \begin{cases} 0 & \forall\, x < 0 \\ \rho - \delta^{\text{id}} & \forall\, x \in [0,H) \\ 1 & \forall\, x \geq H, \end{cases} \qquad G_1^{\text{id}}(x) = \begin{cases} 0 & \forall\, x < 0 \\ \rho + \delta^{\text{id}} & \forall\, x \in [0,H) \\ 1 & \forall\, x \geq H, \end{cases}$$

where $\delta^{\text{id}} = \min\left\{\frac{\rho}{2}, \frac{1-\rho}{2}, \frac{1}{4}\sqrt{\frac{1-\rho}{N}}\right\}$, and $H = \lambda/2$. In this case, $G_0^{\text{id},-}(\lambda) = G_1^{\text{id},-}(\lambda) = 1$.

Note that all instances are effectively uncensored. For $k \in \{\text{ui},\text{ke}\}$, this holds because $s^{\text{off},\lambda} = \lambda \implies d_i^{\text{off}} = 1$. For $k = \text{id}$, this holds because $d_i^{\text{off}} < \lambda$ for all i, which then implies that $d_i^{\text{off}} = s_i^{\text{off},\lambda}$ for all i.

For all $k \in \{\text{id},\text{ui},\text{ke}\}$:

$$\sup_{G \in \mathcal{G}^k} \mathbb{E}_G\left[\sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q_F^\star) - \Delta_G\right] \geq \sup_{G \in \{G_0^k, G_1^k\}} \mathbb{E}_G\left[\sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q_F^\star) - \Delta_G\right]. \quad (55)$$

For clarity, in the remainder of the proof we make clear the dependence of Δ and q^Δ on the underlying demand distribution G.

Recall, by Theorem 1, for any distribution G, the minimax optimal ordering quantity and minimax risk are respectively given by:

$$q_G^\Delta = \begin{cases} q_G^\star & \text{if } G^-(\lambda) \geq \rho \\ q_G^\dagger & \text{if } G^-(\lambda) < \rho \end{cases} \qquad \Delta_G = \begin{cases} 0 & \text{if } G^-(\lambda) \geq \rho \\ \frac{h(b-(b+h)G^-(\lambda))(M-\lambda)}{(b+h)(1-G^-(\lambda)} & \text{if } G^-(\lambda) < \rho, \end{cases} \quad (56)$$

with $\frac{h(b-(b+h)G^-(\lambda))(M-\lambda)}{(b+h)(1-G^-(\lambda)} = (b-(b+h)G^-(\lambda))(M-q_G^\dagger) = h(q_G^\dagger - \lambda)$ for G such that $G^-(\lambda) < \rho$ (see proof of Theorem 1).

Lemma 10 provides the minimax optimal ordering quantity for the three sets of distributions described above. We defer its proof to Appendix D.5.1.

LEMMA 10. *The following holds, in each regime:*

(i) *strictly unidentifiable regime:*

$$q_G^\Delta = q_G^\dagger \quad \forall\, G \in \{G_0^{\text{ui}}, G_1^{\text{ui}}\}$$

(ii) *knife-edge regime:*

$$q_{G_0^{\text{ke}}}^\Delta = q_{G_0^{\text{ke}}}^\dagger, \quad q_{G_1^{\text{ke}}}^\Delta = q_{G_1^{\text{ke}}}^\star = 0$$

(iii) *strictly identifiable regime:*

$$q^{\Delta}_{G_0^{\rm id}} = q^{\star}_{G_0^{\rm id}} = H, \quad q^{\Delta}_{G_1^{\rm id}} = q^{\star}_{G_1^{\rm id}} = 0.$$

Lemma 11 next formalizes the idea that minimizing the worst-case regret in excess of Δ_G reduces to the problem of estimating q^{Δ}_G. We defer its proof to Appendix D.5.2.

LEMMA 11. *For* $k \in \{\text{id}, \text{ui}, \text{ke}\}$, $G \in \{G_0^k, G_1^k\}$*:*

$$\sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^{\pi}) - C_F(q^{\star}_F) - \Delta_G \geq \begin{cases} (b+h)\delta_0^k \; |q^{\pi} - q^{\Delta}_G| & \textit{if } k = \text{ui} \\ (b+h)\delta^k \; |q^{\pi} - q^{\Delta}_G| & \textit{if } k \in \{\text{ke}, \text{id}\}. \end{cases}$$

Applying Lemma 11 to (55), it suffices to lower bound the worst-case absolute difference between q^{π} and q^{Δ}_G. Namely:

$$\sup_{G \in \mathcal{G}^k} \mathbb{E}_G \left[\text{Regret}(q^{\pi}) - \Delta_G \right] \geq (b+h)\delta_0^k \sup_{G \in \{G_0^k, G_1^k\}} \mathbb{E}_G \left[|q^{\pi} - q^{\Delta}_G| \right] \quad \text{if } k = \text{ui} \tag{57}$$

$$\sup_{G \in \mathcal{G}^k} \mathbb{E}_G \left[\text{Regret}(q^{\pi}) - \Delta_G \right] \geq (b+h)\delta^k \sup_{G \in \{G_0^k, G_1^k\}} \mathbb{E}_G \left[|q^{\pi} - q^{\Delta}_G| \right] \quad \text{if } k \in \{\text{ke}, \text{id}\}. \tag{58}$$

We next lower bound $\sup_{G \in \{G_0^k, G_1^k\}} \mathbb{E}_G \left[|q^{\pi} - q^{\Delta}_G| \right]$ via a reduction to hypothesis testing. To formalize this, we introduce some additional notation. Let $\mathcal{D}^N$ be the set of all possible demand samples. For $k \in \{\text{id}, \text{ui}, \text{ke}\}$, let $\Psi^k : \mathcal{D}^N \mapsto \{G_0^k, G_1^k\}$ denote a mapping from the observed demand samples to a prediction of the underlying demand distribution. We moreover let $G_0^{k,N}$ and $G_1^{k,N}$ be the joint distributions of the N independent demand samples generated under G_0^k and G_1^k, respectively.

By Proposition 9.2.1. in Duchi (2024):

$$\begin{aligned} &\sup_{G \in \{G_0^k, G_1^k\}} \mathbb{E}_G \left[|q^{\pi} - q^{\Delta}_G| \right] \\ &\quad \geq \frac{1}{2} \left| q^{\Delta}_{G_0^k} - q^{\Delta}_{G_1^k} \right| \inf_{\Psi^k} \left\{ \mathbb{P}_{G_1^k}\left(\Psi^k(d^{\text{off}}) = G_0^k \right) + \mathbb{P}_{G_0^k}\left(\Psi^k(d^{\text{off}}) = G_1^k \right) \right\} \\ &\quad \geq \frac{1}{4} \underbrace{\left| q^{\Delta}_{G_0^k} - q^{\Delta}_{G_1^k} \right|}_{(I)} \underbrace{\exp\Big(-d_{KL}(G_1^{k,N} \,||\, G_0^{k,N}) \Big)}_{(II)}, \end{aligned} \tag{59}$$

where $d_{KL}(G_1^{k,N} \,||\, G_0^{k,N})$ denotes the Kullback-Leibler (KL) divergence, and the second inequality follows from the Bretagnolle-Huber inequality (Lattimore and Szepesvári 2020).

Equation (59) demonstrates the main drivers of regret in our setting. In particular, any algorithm must incur higher regret if (i) the minimax optimal ordering quantities under G_0^k and G_1^k are far apart (term (I)), or (ii) G_0^k and G_1^k are distributionally "close enough" that any hypothesis test that aims to distinguish between the two incurs high error rate (term (II)). Lemmas 12 and 13 respectively lower bound these two component terms for all regimes. We defer their proofs to Appendix D.5.3 and D.5.4, respectively.

LEMMA 12. *For* $k \in \{\text{ui}, \text{ke}, \text{id}\}$*:*

$$\left| q^{\Delta}_{G_0^k} - q^{\Delta}_{G_1^k} \right| \geq \begin{cases} \delta_1^{\text{ui}} \cdot \frac{h(M-\lambda)}{b+h} & \textit{if } k = \text{ui} \\ \lambda & \textit{if } k = \text{ke} \\ H & \textit{if } k = \text{id}. \end{cases}$$

LEMMA 13. *For* $k \in \{\text{ui}, \text{ke}, \text{id}\}$*:*

$$\exp\Big(-d_{KL}(G_1^{k,N} \,||\, G_0^{k,N})\Big) \geq \begin{cases} \exp\Big(-2N \cdot \frac{(\delta_1^k)^2}{1-\rho}\Big) & \text{if } k = \text{ui} \\ \exp\Big(-8N \cdot \frac{(\delta^k)^2}{1-\rho}\Big) & \text{if } k \in \{\text{ke}, \text{id}\}. \end{cases}$$

Applying these two lemmas to (59), and plugging back into (57) and (58), we obtain:

$$\sup_{G \in \mathcal{G}^k} \mathbb{E}_G\Big[\text{Regret}(q^\pi) - \Delta_G\Big] \geq \begin{cases} \frac{1}{4} h(M-\lambda)\delta_0^{\text{ui}}\delta_1^{\text{ui}} \exp\Big(-2N \cdot \frac{(\delta_1^{\text{ui}})^2}{1-\rho}\Big) & \text{if } k = \text{ui} \\ \frac{1}{4}\lambda(b+h)\delta^{\text{ke}} \exp\Big(-8N \cdot \frac{(\delta^{\text{ke}})^2}{1-\rho}\Big) & \text{if } k = \text{ke} \\ \frac{1}{4}H(b+h)\delta^{\text{id}} \exp\Big(-8N \cdot \frac{(\delta^{\text{id}})^2}{1-\rho}\Big) & \text{if } k = \text{id}. \end{cases}$$

For $\delta_0^{\text{ui}} = \min\Big\{\frac{\rho}{2}, \frac{1-\rho}{2}\Big\}$ and $\delta_1^{\text{ui}} = \min\Big\{\frac{\rho}{4}, \frac{3\rho-1}{4}, \frac{1}{\sqrt{2 \cdot 2N/(1-\rho)}}\Big\}$, we obtain, for $k = \text{ui}$:

$$\begin{aligned} &\sup_{G \in \mathcal{G}^k} \mathbb{E}_G\Big[\text{Regret}(q^\pi) - \Delta_G\Big] \\ &\geq \frac{h(M-\lambda)}{8} \min\{\rho, 1-\rho\} \cdot \min\Big\{\frac{\rho}{4}, \frac{3\rho-1}{4}, \frac{\sqrt{1-\rho}}{2\sqrt{N}}\Big\} \cdot e^{-1/2} \\ &\geq \frac{h(M-\lambda)\sqrt{1-\rho}\min\{\rho, 1-\rho\} \cdot \min\{\rho, 3\rho-1\} \cdot e^{-1/2}}{64\sqrt{N}}, \end{aligned}$$

where the second inequality follows from the fact that $\min\{x, y\} \geq xy$ for $x, y \in [0, 1]$.

For $\delta^{\text{ke}} = \min\Big\{\frac{\rho}{2}, \frac{1-\rho}{2}, \frac{1}{\sqrt{2 \cdot 8N/(1-\rho)}}\Big\}$, we obtain, for $k = \text{ke}$:

$$\begin{aligned} &\sup_{G \in \mathcal{G}^k} \mathbb{E}_G\Big[\sup_{F \in \mathcal{F}(\lambda; G)} C_F(q^\pi) - C_F(q_F^\star) - \Delta_G\Big] \\ &\geq \frac{\lambda(b+h)}{4} \min\Big\{\frac{\rho}{2}, \frac{1-\rho}{2}, \frac{1}{4}\sqrt{\frac{1-\rho}{N}}\Big\} \cdot e^{-1/2} \\ &\geq \frac{\lambda(b+h)\sqrt{1-\rho}\min\{\rho, 1-\rho\}e^{-1/2}}{32\sqrt{N}}. \end{aligned}$$

Finally, using the same instantiation of δ^{id} as δ^{ke}, and setting $H = \lambda/2$, we obtain:

$$\begin{aligned} &\sup_{G \in \mathcal{G}^k} \mathbb{E}_G\Big[\sup_{F \in \mathcal{F}(\lambda; G)} C_F(q^\pi) - C_F(q_F^\star) - \Delta_G\Big] \\ &\geq \frac{H(b+h)\sqrt{1-\rho}\min\{\rho, 1-\rho\}e^{-1/2}}{32\sqrt{N}} \\ &= \frac{\lambda(b+h)\sqrt{1-\rho}\min\{\rho, 1-\rho\}e^{-1/2}}{64\sqrt{N}}. \end{aligned}$$

□

D.5.1. Proof of Lemma 10

Proof. We proceed regime-by-regime.

Case I: Strictly unidentifiable regime. This follows from the fact that $G^-(\lambda) < \rho$ for all $G \in \{G_0^{\text{ui}}, G_1^{\text{ui}}\}$.

Case II: Knife-edge regime. This follows from the fact that $G_0^{\text{ke},-}(\lambda) < \rho$, and $G_1^{\text{ke},-}(\lambda) \geq \rho$. Moreover, since G_1^{ke} is Bernoulli, it must then be that $q^\star_{G_1^{\text{ke}}} = 0$.

Case III: Strictly identifiable regime. This follows from the fact that $G^-(\lambda) \geq \rho$ for all $G \in \{G_0^{\text{id}}, G_1^{\text{id}}\}$. For $G = G_0^{\text{id}}$, $\mathbb{P}_G(D = 0) < \rho$, which implies that $q^\star_G = H$. For $G = G_1^{\text{id}}$, $\mathbb{P}_G(D = 0) \geq \rho$, which implies $q^\star_G = 0$. □

D.5.2. Proof of Lemma 11

Proof. We prove the lemma separately for each regime. In all cases, we rely repeatedly on the closed-form characterizations of the minimax risk established in Lemma 2 and Proposition 5, bounding the loss based on the location of q^π relative to λ.

Case I: Strictly unidentifiable regime. By Lemma 10, for all $G \in \{G_0^{\text{ui}}, G_1^{\text{ui}}\}$, $q^\Delta_G = q^\dagger_G$.

Suppose first that $q^\pi < \lambda$. By Lemma 2, we have:

$$\begin{aligned}
\sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q^\star_F) &= b(M - q^\pi) + (b+h)\Big(q^\pi \mathbb{P}_G(D=0) - M\mathbb{P}_G(D=0)\Big) \\
&= (M - q^\pi)(b - (b+h)G^-(\lambda)),
\end{aligned}$$

since $\lambda \in (0,1)$ and G is Bernoulli. Subtracting $\Delta_G = (b - (b+h)G^-(\lambda))(M - q^\dagger_G)$ on both sides, we have:

$$\begin{aligned}
\sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q^\star_F) - \Delta_G &= (M - q^\pi)(b - (b+h)G^-(\lambda)) - (b - (b+h)G^-(\lambda))(M - q^\dagger_G) \\
&= (q^\dagger_G - q^\pi)(b - (b+h)G^-(\lambda)) \\
&\geq (q^\dagger_G - q^\pi)(b+h)\delta_0^{\text{ui}},
\end{aligned}$$

where the final inequality follows from the fact that $G^-(\lambda) \leq \rho - \delta_0^{\text{ui}}$ for $G \in \{G_0^{\text{ui}}, G_1^{\text{ui}}\}$, and $q^\dagger_G \geq \lambda > q^\pi$.

Consider now the case where $q^\pi \in [\lambda, q^\dagger_G]$. By Lemma 2, for $G^-(\lambda) < \rho$, we have:

$$\begin{aligned}
\sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q^\star_F) - \Delta_G &= (b - (b+h)G^-(\lambda))(M - q^\pi) - (b - (b+h)G^-(\lambda))(M - q^\dagger_G) \\
&= (b - (b+h)G^-(\lambda))(q^\dagger_G - q^\pi) \\
&\geq (b+h)\delta_0^{\text{ui}}(q^\dagger_G - q^\pi),
\end{aligned}$$

where again we used the fact that $G^-(\lambda) \leq \rho - \delta_0^{\text{ui}}$ for $G \in \{G_0^{\text{ui}}, G_1^{\text{ui}}\}$.

Finally, if $q^\pi \in (q^\dagger_G, M]$, using the alternative representation of $\Delta_G = h(q^\dagger_G - \lambda)$, by Lemma 2:

$$\sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q^\star_F) - \Delta_G = h(q^\pi - \lambda) - h(q^\dagger_G - \lambda) = h(q^\pi - q^\dagger_G).$$

Combining the three cases and using the fact that $\delta_0^{\text{ui}} < 1 - \rho \implies (b+h)\delta_0^{\text{ui}} < h$, we obtain the lower bound for the strictly unidentifiable regime.

Case II: Knife-edge regime. If $G = G_0^{\text{ke}}$, since $G_0^{\text{ke},-}(\lambda) < \rho$, by the same arguments as above:

$$\sup_{F \in \mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q^\star_F) - \Delta_G \geq (b+h)\delta^{\text{ke}}|q^\pi - q^\dagger_G|. \tag{60}$$

If $G = G_1^{\mathrm{ke}}$, by (56), $\Delta_G = 0$, since $G_1^{\mathrm{ke},-}(\lambda) \geq \rho$ by construction. Moreover, by Lemma 10, $q_G^\Delta = q_G^\star = 0$. Suppose first that $q^\pi < \lambda$. Applying Proposition 5 to $G = G_1^{\mathrm{ke}}$, we have:

$$\begin{aligned}\sup_{F\in\mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q_F^\star) &= -bq^\pi + (b+h)q^\pi\mathbb{P}_G(D=0)\\ &= -bq^\pi + (b+h)q^\pi(\rho+\delta^{\mathrm{ke}})\\ &= q^\pi(b+h)\delta^{\mathrm{ke}},\end{aligned}$$

where the first equality simplified the expression in Proposition 5 using the fact that $\lambda \in (0,1)$, and $q_G^\star = 0$. Suppose now that $q^\pi \geq \lambda$. Again, by Proposition 5:

$$\begin{aligned}\sup_{F\in\mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q_F^\star) &= -bq^\pi + (b+h)(q^\pi - \lambda + \lambda\mathbb{P}_G(D=0))\\ &= hq^\pi - \lambda(b+h)\mathbb{P}_G(D=1)\\ &\geq q^\pi\Big(h - (b+h)(1-\rho-\delta^{\mathrm{ke}})\Big)\\ &= q^\pi(b+h)\delta^{\mathrm{ke}}.\end{aligned}$$

Hence, in both cases:

$$\sup_{F\in\mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q_F^\star) \geq (q^\pi - q_G^\star)(b+h)\delta^{\mathrm{ke}}. \tag{61}$$

Since $q_G^\Delta = q_G^\dagger$ for $G = G_0^{\mathrm{ke}}$, and $q_G^\Delta = q_G^\star$ for $G = G_1^{\mathrm{ke}}$, putting (60) and (61) together we obtain:

$$\sup_{F\in\mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q_F^\star) - \Delta_G \geq (b+h)\delta^{\mathrm{ke}}|q^\pi - q^\Delta| \quad \forall\, G\in\{G_0, G_1\}.$$

Case III: Strictly identifiable regime. By Lemma 10, $q^\Delta_{G_0^{\mathrm{id}}} = q^\star_{G_0^{\mathrm{id}}} = H$, and $q^\Delta_{G_1^{\mathrm{id}}} = q^\star_{G_1^{\mathrm{id}}} = 0$. Moreover, for $G \in \{G_0^{\mathrm{id}}, G_1^{\mathrm{id}}\}$, $\Delta_G = 0$, by (56).

Suppose first that $G = G_0^{\mathrm{id}}$. By Proposition 5, if $q^\pi < \lambda$:

$$\begin{aligned}&\sup_{F\in\mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q_F^\star)\\ &= b(H-q^\pi) + (b+h)\Big(q^\pi\mathbb{P}_G(D=0) + (q^\pi - H)\mathbb{1}\{H\leq q^\pi\}\mathbb{P}_G(D=H) - H\mathbb{P}_G(D=0)\Big)\\ &= (H-q^\pi)(b - (b+h)\mathbb{P}_G(D=0) - (b+h)\mathbb{P}_G(D=H)\mathbb{1}\{H\leq q^\pi\})\\ &= (b+h)(H-q^\pi)\Big(\delta^{\mathrm{id}} - (1-\rho+\delta^{\mathrm{id}})\mathbb{1}\{H\leq q^\pi\}\Big)\\ &= (b+h)(1-\rho)(q^\pi - H)\mathbb{1}\{H\leq q^\pi\} + (b+h)(H-q^\pi)\delta^{\mathrm{id}}\mathbb{1}\{H > q^\pi\}\\ &\geq (b+h)|H-q^\pi|\delta^{\mathrm{id}} = (b+h)|q_G^\Delta - q^\pi|\delta^{\mathrm{id}},\end{aligned} \tag{62}$$

where the inequality follows from the fact that $\delta^{\mathrm{id}} < 1-\rho$.
If $q^\pi \geq \lambda$, on the other hand, by Proposition 5:

$$\begin{aligned}&\sup_{F\in\mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q_F^\star)\\ &= b(H-q^\pi) + (b+h)\Big(q^\pi - \lambda + \lambda\mathbb{P}_G(D=0) + (\lambda - H)\mathbb{P}_G(D=H) - H\mathbb{P}_G(D=0)\Big)\\ &= b(H-q^\pi) + (b+h)(q^\pi - H)\\ &= h(q^\pi - H)\\ &\geq (b+h)|H-q^\pi|\delta^{\mathrm{id}} = (b+h)|q_G^\Delta - q^\pi|\delta^{\mathrm{id}},\end{aligned} \tag{63}$$

where the inequality follows from the fact that $q^\pi \geq \lambda > H$, and $\delta^{\text{id}} < 1-\rho$.

Putting (62) and (63) together, we obtain the lower bound for $G = G_0^{\text{id}}$.

Suppose now that $G = G_1^{\text{id}}$. By Proposition 5, if $q^\pi < \lambda$:

$$\begin{aligned}
\sup_{F\in\mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q_F^\star) &= -bq^\pi + (b+h)\Big(q^\pi \mathbb{P}_G(D=0) + (q^\pi - H)\mathbb{P}_G(D=H)\mathbb{1}\{H \leq q^\pi\}\Big) \\
&= -bq^\pi + (b+h)\Big(q^\pi(\rho+\delta^{\text{id}}) + (q^\pi - H)\mathbb{1}\{H\leq q^\pi\}(1-\rho-\delta^{\text{id}})\Big) \\
&= (b+h)\Big(q^\pi\delta^{\text{id}} + (q^\pi - H)(1-\rho-\delta^{\text{id}})\Big)\mathbb{1}\{H\leq q^\pi\} \\
&\quad + (b+h)\delta^{\text{id}} q^\pi \mathbb{1}\{H > q^\pi\} \\
&\geq (b+h)\delta^{\text{id}} q^\pi = (b+h)\delta^{\text{id}}|q^\pi - q_G^\Delta|,
\end{aligned} \tag{64}$$

where the inequality follows from the fact that $\delta^{\text{id}} < 1-\rho$.

If $q^\pi \geq \lambda$, on the other hand, by Proposition 5:

$$\begin{aligned}
\sup_{F\in\mathcal{F}(\lambda;G)} C_F(q^\pi) - C_F(q_F^\star) &= -bq^\pi + (b+h)\Big[(q^\pi - \lambda) + \lambda\mathbb{P}_G(D=0) + (\lambda - H)\mathbb{P}_G(D=H)\Big] \\
&= hq^\pi - H(b+h)(1-\rho-\delta^{\text{id}}) \\
&\geq \Big(h - (b+h)(1-\rho-\delta^{\text{id}})\Big)q^\pi \\
&= (b+h)\delta^{\text{id}} q^\pi = (b+h)\delta^{\text{id}}|q^\pi - q_G^\Delta|,
\end{aligned} \tag{65}$$

where the inequality follows from the fact that $q^\pi \geq \lambda > H$.

Putting (64) and (65) together, we obtain the lower bound for $G = G_1^{\text{id}}$. □

D.5.3. Proof of Lemma 12

Proof. We proceed case-by-case.

Case I: Strictly unidentifiable regime. By Lemma 10, in this setting, for $G \in \{G_0^{\text{ui}}, G_1^{\text{ui}}\}$,

$$q_G^\Delta = q_G^\dagger = \frac{h\lambda + (b-(b+h)G^-(\lambda))M}{(b+h)(1-G^-(\lambda))}.$$

As in the proof of Theorem 2, define $Q_G^\dagger(\cdot)$ as:

$$Q_G^\dagger(x) = \frac{h\lambda + (b-(b+h)x)M}{(b+h)(1-x)}.$$

Recall, $G_0^{\text{ui},-}(\lambda) = \rho - \delta_0^{\text{ui}}$, and $G_1^{\text{ui},-}(\lambda) = \rho - \delta_0^{\text{ui}} - \delta_1^{\text{ui}}$. Hence, by the mean value theorem, for some $c \in (\rho - \delta_0^{\text{ui}} - \delta_1^{\text{ui}}, \rho - \delta_0^{\text{ui}})$:

$$\begin{aligned}
&Q_G^\dagger(\rho - \delta_0^{\text{ui}}) - Q_G^\dagger(\rho - (\delta_0^{\text{ui}} + \delta_1^{\text{ui}})) = \delta_1^{\text{ui}} \cdot \frac{dq_G^\dagger}{dx}\bigg|_{x=c} = \delta_1^{\text{ui}} \cdot \frac{h(\lambda - M)}{(b+h)(1-c)^2} \\
\Longrightarrow\ &\big|Q_G^\dagger(\rho - \delta_0^{\text{ui}}) - Q_G^\dagger(\rho - (\delta_0^{\text{ui}} + \delta_1^{\text{ui}}))\big| \geq \delta_1^{\text{ui}} \cdot \frac{h(M-\lambda)}{b+h},
\end{aligned} \tag{66}$$

thus providing the bound for the unidentifiable regime.

Case II: Knife-edge regime. By Lemma 10, in this setting $q^{\Delta}_{G_0^{\text{ke}}} = q^{\dagger}_{G_0^{\text{ke}}}$, and $q^{\Delta}_{G_1^{\text{ke}}} = q^{\star}_{G_1^{\text{ke}}} = 0$. Hence:

$$|q^{\Delta}_{G_0^{\text{ke}}} - q^{\Delta}_{G_1^{\text{ke}}}| = \frac{h\lambda + (b-(b+h)(\rho-\delta^{\text{ke}}))M}{(b+h)(1-\rho+\delta^{\text{ke}})} \geq \frac{h\lambda + (b-(b+h)\rho)M}{(b+h)(1-\rho)} = \lambda,$$

where the inequality follows from the fact that $Q^{\dagger}_G(x)$ is decreasing in x, and $\rho - \delta^{\text{ke}} \leq \rho$ for all $\delta \geq 0$.

Case III: Strictly identifiable regime. This follows from the fact that, by Lemma 10, $q^{\Delta}_{G_0^{\text{id}}} = H$ and $q^{\Delta}_{G_1^{\text{id}}} = 0$. □

D.5.4. Proof of Lemma 13

Proof. Since we are in the uncensored setting and demand samples are drawn i.i.d., we have that $d_{KL}(G_1^{k,N} \,||\, G_0^{k,N}) = N d_{KL}(G_1^k \,||\, G_0^k)$. Moreover, across all regimes, G_0^k and G_1^k are distributions with support over two point masses. Hence, we can use reverse Pinsker's inequality to upper bound the KL divergence from $G_1^{k,N}$ to $G_0^{k,N}$ (Sason (2015), Eqn. 10). Namely, for any two distributions G_0^k, G_1^k with support on the same two point masses a and b, with $a > b$, we have:

$$d_{KL}(G_1^k \,||\, G_0^k) = 2 \cdot \frac{(p-q)^2}{q},$$

where p and q respectively denote the mass G_1^k and G_0^k place on a. We apply this to each regime, to obtain:

Case I: Strictly unidentifiable regime.

$$\exp\left(-d_{KL}(G_1^{\text{ui},N} \,||\, G_0^{\text{ui},N})\right) \geq \exp\left(-2N \cdot \frac{(\delta_1^{\text{ui}})^2}{1-\rho+\delta_0^{\text{ui}}}\right) \geq \exp\left(-2N \cdot \frac{(\delta_1^{\text{ui}})^2}{1-\rho}\right).$$

Case II: Knife-edge regime.

$$\exp\left(-d_{KL}(G_1^{\text{ke},N} \,||\, G_0^{\text{ke},N})\right) \geq \exp\left(-2N \cdot \frac{(2\delta^{\text{ke}})^2}{1-\rho+\delta^{\text{ke}}}\right) \geq \exp\left(-8N \cdot \frac{(\delta^{\text{ke}})^2}{1-\rho}\right).$$

Case III: Strictly identifiable regime.

$$\exp\left(-d_{KL}(G_1^{\text{id},N} \,||\, G_0^{\text{id},N})\right) \geq \exp\left(-2N \cdot \frac{(2\delta^{\text{id}})^2}{1-\rho+\delta^{\text{id}}}\right) \geq \exp\left(-8N \cdot \frac{(\delta^{\text{id}})^2}{1-\rho}\right).$$

□

Appendix E: Analysis of RCN⁺

THEOREM 5. *Fix $\delta \in (0,1)$, and let $\zeta_k = \sqrt{\frac{\log(2K/\delta)}{2N_k}}$. With probability at least $1-2\delta$, Algorithm 1 outputs an ordering quantity q^{alg} such that:*

(i) if $G^-(q_k^{\text{off}}) \geq \rho + 2\zeta_k$ for some $k \in [K]$:

$$\text{Regret}(q^{\text{alg}}) \leq \lambda(b+h)\sqrt{\frac{\log(2/\delta)}{2\sum_{k\in\mathcal{U}_1} N_k}},$$

where $\mathcal{U}_1 = \{k : G^-(q_k^{\text{off}}) \geq \rho + 2\zeta_k\}$.

(ii) if $G^-(\lambda) \geq \rho$ and $G^-(q_k^{\text{off}}) < \rho + 2\zeta_k$ for all $k \in [K]$:

$$\text{Regret}(q^{\text{alg}}) \leq \lambda(b+h)\max\left\{2\zeta_K, \sqrt{\frac{\log(2/\delta)}{2\min_{k\in\mathcal{U}_2} N_k}}\right\},$$

where $\mathcal{U}_2 = \{k : G^-(q_k^{\text{off}}) \geq \rho\}$.

(iii) if $G^-(\lambda) < \rho$ and $G^-(q_k^{\text{off}}) \geq \rho - 2\zeta_k$ for some $k \in [K]$:

$$\text{Regret}(q^{\text{alg}}) \leq \Delta + 2\max\left\{\frac{b}{h}, 1\right\}(M-\lambda)(b+h)\zeta_K$$

(iv) if $G^-(q_k^{\text{off}}) < \rho - 2\zeta_k$ for all $k \in [K]$:

$$\text{Regret}(q^{\text{alg}}) \leq \Delta + \max\left\{\frac{b}{h}, 1\right\}(M-\lambda)(b+h)\zeta_K.$$

Proof. Abusing notation, we define $q^\dagger_{\widehat{G}} = \frac{bM+h\lambda-(b+h)\widehat{G}_K^-(\lambda)M}{(b+h)(1-\widehat{G}_K^-(\lambda))}$. We also let $N_{\text{est}} = \sum_{k\in\mathcal{U}_{\text{est}}} N_k$, $\widehat{G}(x) = \frac{1}{N_{\text{est}}}\sum_{k\in\mathcal{U}_{\text{est}}}\sum_{i\in[N_k]} \mathbb{1}\left\{s_{ki}^{\text{off}} \leq x\right\}$, and $\widehat{G}^{\text{uc}}(x) = \frac{1}{N_{\text{est}}}\sum_{k\in\mathcal{U}_{\text{est}}}\sum_{i\in[N_k]} \mathbb{1}\left\{d_{ki}^{\text{off}} \leq x\right\}$.

Recall, by Proposition 6,

$$\text{Regret}(q^{\text{alg}}) \leq \Delta + \sup_{F\in\mathcal{F}(\lambda;G)} C_F(q^{\text{alg}}) - C_F(q^\Delta). \tag{67}$$

Therefore, as in the proof of Theorem 2, we focus on bounding $\sup_{F\in\mathcal{F}(\lambda;G)} C_F(q^{\text{alg}}) - C_F(q^\Delta)$. We will rely heavily on the following facts in the remainder of our proof.

LEMMA 14. *Let $\mathcal{E} = \left\{|\widehat{G}_k^-(q_k^{\text{off}}) - G^-(q_k^{\text{off}})| < \zeta_k \text{ for all } k \in [K]\right\}$. Then, $\mathbb{P}_G(\mathcal{E}) \geq 1-\delta$.*

Lemma 14 is an analog of Lemma 6, where we additionally union bound over all selling seasons $k \in [K]$. It similarly follows from the fact that $\mathbb{1}\left\{s_{ki}^{\text{off}} < q_k^{\text{off}}\right\} = \mathbb{1}\left\{d_{ki}^{\text{off}} < q_k^{\text{off}}\right\}$ for all $k \in [K]$, $i \in [N_k]$. We omit the proof of this fact as such.

With these facts in hand, we prove each of the regret bounds separately, beginning with the two "extreme" cases of $G^-(q_k^{\text{off}}) \geq \rho + 2\zeta_k$ for some $k \in [K]$ and $G^-(q_k^{\text{off}}) < \rho - 2\zeta_k$ for all $k \in [K]$. Moreover, in the remainder of the proof we condition on the good event $\mathcal{E}$ defined in Lemma 14.

Case I: $G^-(q_k^{\text{off}}) \geq \rho + 2\zeta_k$ for some $k \in [K]$. Under event $\mathcal{E}$, for all $k \in \mathcal{U}_1$, $\widehat{G}_k^-(q_k^{\text{off}}) \geq \rho + \zeta_k$. Therefore, $\mathcal{U}_{\text{est}} \neq \emptyset$. Moreover, for all $k \in \mathcal{U}_{\text{est}}$, under event $\mathcal{E}$ we have $G^-(q_k^{\text{off}}) \geq \rho$; therefore, $q_G^\star \leq \min_{k\in\mathcal{U}_{\text{est}}} q_k^{\text{off}}$.

Note that, in this case, $G^-(\lambda) \geq \rho$. We therefore apply Proposition 5, obtaining:

$$\text{Regret}(q^{\text{alg}}) = C_G(q^{\text{alg}}) - C_G(q_G^\star). \tag{68}$$

Suppose first that $q^{\text{alg}} \leq q_G^\star < \min_{k\in\mathcal{U}_{\text{est}}} q_k^{\text{off}}$. By the same arguments as those used to establish (51), we have:

$$\begin{aligned} C_G(q^{\text{alg}}) - C_G(q_G^\star) &\leq (b+h)(q_G^\star - q^{\text{alg}})(\rho - G(q^{\text{alg}})) \\ &\leq (b+h)\lambda(\widehat{G}(q^{\text{alg}}) - G(q^{\text{alg}})), \end{aligned}$$

where we use the fact that $\widehat{G}(q^{\text{alg}}) \geq \rho$ by construction. Now, since $q^{\text{alg}} < \min_{k\in\mathcal{U}_{\text{est}}} q_k^{\text{off}}$, we have that $\mathbb{1}\left\{s_{ki}^{\text{off}} \leq q^{\text{alg}}\right\} = \mathbb{1}\left\{d_{ki}^{\text{off}} \leq q^{\text{alg}}\right\}$ for all $k \in \mathcal{U}_{\text{est}}$, $i \in [N_k]$. Therefore, $\widehat{G}(q^{\text{alg}}) = \widehat{G}^{\text{uc}}(q^{\text{alg}})$, and

$$\begin{aligned} C_G(q^{\text{alg}}) - C_G(q_G^\star) &\leq (b+h)\lambda(\widehat{G}(q^{\text{alg}}) - G(q^{\text{alg}})) \\ &\leq (b+h)\lambda(\widehat{G}^{\text{uc}}(q^{\text{alg}}) - G(q^{\text{alg}})). \end{aligned}$$

Applying the DKW inequality, we have that with probability at least $1-\delta$,

$$C_G(q^{\text{alg}}) - C_G(q_G^\star) \leq (b+h)\lambda\sqrt{\frac{\log(2/\delta)}{2N_{\text{est}}}}.$$

Suppose now that $q^{\text{alg}} > q^\star_G$. In this case, by the same arguments as those used to establish (53), we have:

$$\begin{aligned}
C_G(q^{\text{alg}}) - C_G(q^\star_G) &\le (b+h)(q^{\text{alg}} - q^\star_G)(G^-(q^{\text{alg}}) - \rho) \\
&\le (b+h)(q^{\text{alg}} - q^\star_G)\left(\lim_{x\to q^{\text{alg}}} \widehat{G}^{\text{uc}}(x) + \sqrt{\frac{\log(2/\delta)}{2N_{\text{est}}}} - \rho\right) \\
&\le (b+h)(q^{\text{alg}} - q^\star_G)\left(\lim_{x\to q^{\text{alg}}} \widehat{G}(x) + \sqrt{\frac{\log(2/\delta)}{2N_{\text{est}}}} - \rho\right) \\
&\le (b+h)\lambda\sqrt{\frac{\log(2/\delta)}{2N_{\text{est}}}},
\end{aligned}$$

where the second inequality applies the DKW inequality, the third inequality uses the fact that $\mathbb{1}\left\{s^{\text{off}}_{ki} \le x\right\} \ge \mathbb{1}\left\{d^{\text{off}}_{ki} \le x\right\}$ for all x, the fourth inequality uses the fact that $\lim_{x\to q^{\text{alg}}} \widehat{G}(x) < \rho$ by definition of q^{alg}, and the final inequality uses the fact that $q^{\text{alg}} \le \lambda$.

Using the fact that $k \in \mathcal{U}_1 \implies k \in \mathcal{U}_{\text{est}}$, we have that $N_{\text{est}} \ge \sum_{k\in\mathcal{U}_1} N_k$. Applying this bound to the above concludes the proof of the first case.

Case II: $G^-(q^{\text{off}}_k) < \rho - 2\zeta_k$ for all $k \in [K]$. In this case, $G^-(\lambda) < \rho$, and by Theorem 1, $q^\Delta = q^\dagger_G$. Moreover, under event $\mathcal{E}$, $\widehat{G}^-_k(q^{\text{off}}_k) < \rho - \zeta_k$ for all $k \in [K]$, and our algorithm outputs $q^{\text{alg}} = q^\dagger_{\widehat{G}}$. In this case, the same analysis as that of Case II in the proof of Theorem 2 applies, and we obtain:

$$\text{Regret}(q^{\text{alg}}) \le \Delta + \frac{\max\{b,h\}(M-\lambda)}{1-\rho}\zeta_K, \tag{69}$$

completing the proof of the second case.

We now turn our attention to the remaining case.

Case III: $G^-(q^{\text{off}}_k) < \rho + 2\zeta_k$ for all $k \in [K]$, and $G^-(q^{\text{off}}_k) \ge \rho - 2\zeta_k$ for some $k \in [K]$. We further partition the analysis based on whether $G^-(\lambda) \ge \rho$.

(i) $G^-(\lambda) \ge \rho$. By Theorem 1, $q^\Delta = q^\star_G$, and $\Delta = 0$. Moreover, given $\mathcal{E}$, $\widehat{G}^-_k(q^{\text{off}}_k) \in (\rho - \zeta_k, \rho + 3\zeta_k)$ for all $k \in \mathcal{U}_2$.

a. Suppose first that there exists $k \in \mathcal{U}_2$ such that $\widehat{G}^-_k(q^{\text{off}}_k) \in [\rho + \zeta_k, \rho + 3\zeta_k)$. Then, $\mathcal{U}_{\text{est}} \ne \emptyset$, and identical arguments as those used for Case I yield:

$$\text{Regret}(q^{\text{alg}}) \le (b+h)\lambda\sqrt{\frac{\log(2/\delta)}{2N_{\text{est}}}} \tag{70}$$

with probability at least $1-2\delta$. Using the fact that $N_{\text{est}} \ge \min_{k\in\mathcal{U}_2} N_k$ since $\mathcal{U}_{\text{est}} \subseteq \mathcal{U}_2$ under event $\mathcal{E}$, we obtain:

$$\text{Regret}(q^{\text{alg}}) \le (b+h)\lambda\sqrt{\frac{\log(2/\delta)}{2\min_{k\in\mathcal{U}_2} N_k}}. \tag{71}$$

b. Suppose now that $\widehat{G}^-_k(q^{\text{off}}_k) \in (\rho - \zeta_k, \rho + \zeta_k)$ for all $k \in \mathcal{U}_2$. In this case, $q^{\text{alg}} = \lambda$. Using the same arguments as those used in Case III of the proof of Theorem 2, we have:

$$\text{Regret}(q^{\text{alg}}) \le (b+h)(\lambda - q^\star_G)(G^-(\lambda) - \rho).$$

Using the fact that $q^\star_G \in [0,\lambda)$ for $G^-(\lambda) \ge \rho$, and $G^-(\lambda) \le \rho + 2\zeta_K$ by assumption, we obtain the final bound of:

$$\text{Regret}(q^{\text{alg}}) \le 2\lambda(b+h)\zeta_K.$$

(ii) $G^-(\lambda) < \rho$. By Theorem 1, $q^\Delta = q^\dagger_G$. Moreover, under $\mathcal{E}$, $\widehat{G}^-_k(q^{\text{off}}_k) \in (\rho - 3\zeta_k, \rho + \zeta_k)$ for all $k \in [K]$.

a. Suppose first that $\widehat{G}^-_k(q^{\text{off}}_k) \in (\rho - 3\zeta_k, \rho - \zeta_k)$ for all $k \in [K]$. In this case, the algorithm outputs $q^{\text{alg}} = q^\dagger_{\widehat{G}}$. By the same arguments as those used in Case II, we obtain:

$$\text{Regret}(q^{\text{alg}}) \leq \Delta + \frac{\max\{b,h\}(M-\lambda)}{(1-\rho)}\zeta_K.$$

b. Suppose now that there exists k such that $\widehat{G}^-_k(q^{\text{off}}_k) \in [\rho - \zeta_k, \rho + \zeta_k)$. In this case, $q^{\text{alg}} = \lambda$. By the same arguments as those used in Case IV of the proof of Theorem 2, we have:

$$\text{Regret}(q^{\text{alg}}) \leq \Delta + \max\{b,h\}\frac{2(M-\lambda)}{1-\rho}\zeta_K,$$

concluding the proof of the theorem.

□