A data-driven distributionally robust optimization approach for risk-averse newsvendor problem with demand censoring

Xiaoli Su[a,b], Bo Shao[c], Chenghu Yang[a], Zhe Yuan[b*]

[a]School of Economics and Management, Fuzhou University, Fuzhou, China;

[b]Léonard de Vinci Pôle Universitaire, Research Center, 92916 Paris La Défense, France;

[c]Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, United States

*Correspondence: zhe.yuan@devinci.fr

Abstract: Limited historical censored sales data fails to accurately represent true demand, leading risk-averse decision makers (DMs) to adopt more conservative order decisions. To tackle this challenge, we develop data-driven distributionally robust optimization models for the risk-averse newsvendor problem with demand censoring (RA-DRN). In the RA-DRN models, we effectively capture the effect of demand censoring through the Kaplan-Meier (KM) estimator, and achieve an efficient collaboration between limited historical censored data and external demand information with different granularities. We transform the RA-DRN models into equivalent single-layer convex optimization problems using Lagrange multipliers, duality theory, and minimax inequality, thereby simplifying the solution process. Furthermore, through the Monte Carlo Simulation, we investigate the comprehensive effects of risk aversion and demand censoring on order decisions. Our findings reveal that the proposed models with the synergy of local-external demand information possess stronger robustness and effectively mitigate the adverse effect of demand censoring. When incorporating more granular external demand information, the order bias resulting from demand censoring might even be disregarded. In the majority of cases, risk-averse DMs tend to make more conservative order decisions and have a preference for collecting external demand information with greater granularity. In addition, the value of the synergy increases monotonically with respect to risk aversion, demand censoring, and profit structures.

Keywords: Demand censoring, Risk aversion, Distributionally robust optimization, Different information granularities

Acknowledgments: The authors would like to thank the session chair of the 33rd EURO Conference for encouraging us to submit this paper to the special issue of “Recent advances in stochastic and robust optimization”.

1. Introduction

The newsvendor problem serves as a cornerstone for various operations management issues, including inventory control, dynamic pricing, sourcing management, and platform operations. The classical formulation of the newsvendor problem assumes that complete demand information, including its distribution, is known. However, this is an unrealistic assumption in practical scenarios (Saghafian &Tomlin, 2016; Xue et al., 2022). In particular, it is difficult for decision makers (DMs) to gather sufficient historical demand information to accurately estimate the volatile demands of short life-cycle products or new products (Natarajan et al., 2017). Reliable demand data is often costly and challenging to obtain, and DMs may be unable to receive comprehensive insights about demand distribution but only limited historical sales data (Yang et al., 2019). Limited historical sales data may lead to significant errors in distribution fitting, resulting in suboptimal decisions. Furthermore, with underordering products, lost sales (excess demand) are unable to be observed for DMs. As such, the historical sales data, regarded as censored data, only captures only a fraction of the true demand, contributing to a more ambiguous understanding of market demand (Huh et al., 2011; Bu et al., 2022; Ding et al., 2024).

With the intrinsic uncertainty of demand and the scarcity of reliable historical data, distributionally robust optimization (DRO) approaches have been proposed in newsvendor problems (DRN) as a solution. These DRN models can effectively exploit limited demand data to enhance order decisions, offering resilience for DMs against the risk of wrong order decisions. Depending on the characteristics of the limited demand information derived from historical sales data, DRN problems can be classified into two types (Natarajan et al., 2017; Zhang et al., 2020). The first type is DRN involving a moment-based uncertainty set, where limited demand information is expressed through the range, mean, variance, and asymmetry of the demand distribution (Natarajan et al., 2017). The second type, known as the DRN problem with a distribution-based uncertainty set, utilizes empirical or prior nominal distributions to represent the limited demand information (Zhang et al., 2020). These DRN models can effectively exploit limited demand information to enhance order decisions. However, the majority of existing DRN models do not consider the risk of wrong order decisions for DMs.

Behavioral factors, such as cognitive abilities and biases, can systematically drive outcomes that diverge from model predictions, rendering conventional models less effective in representing actual behavior (Long et al., 2015; Surti et al., 2020). In practice, DMs often aim to mitigate the risk associated with inventory decisions, placing greater emphasis on integrating risk management strategies into the order decision-making process. As demonstrated by Eeckhoudt et al. (1995), risk-averse newsvendors tend to order less than risk-neutral ones. In recent studies, a multitude of models have been proposed to incorporate DMs' risk-averse preferences and to assess their impact on ordering decisions (Change et al., 2024; Liu and Zhu., 2024). These models diverge from the traditional Expected Utility Theory (EUT) framework, which is frequently employed in decision-making analysis(Sayın et al., 2014; Wang et al., 2009). Instead, Prospect Theory (PT), a widely recognized measure of risk, is adopted to elucidate the underlying causes of the 'asymmetry in ordering' phenomenon (Zhang and Siemsen, 2019; Shinde et al., 2021). PT explains that risk-averse DMs order more (less) than the order quantity with maximizing expected profit, called a pull-to-center pattern (Nagarajan and Shechter, 2014).

In addition, the majority of existing DRN models overlook the information loss caused by demand censoring, which can further degrade the performance of the models (Bu et al., 2022). Due to the intrinsic uncertainty of demand and the scarcity of reliable historical data, common issues such as sampling errors and model specification errors have the potential to misguide decision-making processes (Lin et al., 2022). Shinde et al. (2021) employed DRN models, incorporating moment-based uncertainty sets, to analyze DMs' ordering behavior in the context of risk and ambiguity. This research indicates that demand censoring can amplify distribution estimation errors, prompting risk-averse decision-makers to adopt more conservative strategies. Furthermore, they observe that with a small sample size, the precision of the estimated mean and standard deviation is compromised, which can further influence the accuracy of decision-making under uncertainty.

Fortunately, enhancing demand information can improve the performance of DRN issues with limited historical censored data. In practice, DMs exploit knowledge transferred from external multisource demand information to refine the precision and robustness of DRN problems (Zhang et al., 2020; Yang et al., 2023). With the development of advanced data technologies, DMs can observe different types of demand information at low cost from multiple external sources, such as Google, cloud platforms, social media marketing, and online retail websites. For example, by using Amazon's "Seller Central" (a third-party information services platform), online vendors can obtain relevant demand information, such as price, ratings, units sold prediction, sales trends, and more (Ha et al., 2022). Similarly, retailers on Tmall can collect demand insights through the Tmall Innovation Center (Li et al., 2021). DMs make an efficient synergy between historical sales data and multisource external demand information, to effectively utilize local-external demand information and obtain more powerful order decisions (Zhang et al., 2020; Yang et al., 2023).

Building on this analysis, we try to address the following research issues:

1. With demand censoring, how do risk-averse DMs utilize limited historical sale data and multisource external information to determine order decisions?

2. How do risk-averse DMs mitigate the superimposed asymmetry effect on ordering induced by demand censoring and risk aversion coefficients?

To explore these issues, we propose a data-driven distributionally robust optimization approach to risk-averse newsvendor models by introducing the Kaplan-Meier (KM) estimator. Multisource external demand information is integrated to reduce the negative effects of limited historical censored data and substantially improve the precision of order decisions.

Our first contribution is to formulate a data-driven distributionally robust newsvendor problem with demand censoring faced by risk-averse DMs and fully exploit limited historical censored data and external multisource demand information. Considering the characteristics of local-external demand information, coarse and finer-grained demand information is integrated into the proposed models based on the phi-divergence and Wasserstein measures, respectively. Our second contribution is to transform the two RA-DRN models into equivalent single-layer convex optimization problems based on Lagrange multipliers, duality theory, and minimax inequality, which are then amenable to algorithmic implementation. Our third contribution is to validate the performances of the RA-DRN models under different conditions. We show the performance of the RA-DRN models monotonically increases with respect to the loss aversion coefficients of DMs, the tail of distributions, demand censoring, and profit structures. Incorporating external demand information can eliminate the adverse effect of demand censoring, and even ignore demand censoring problems with finer-grained external demand information.

The remainder of our work is organized as follows. In Section 2, the literature review and research gaps are analyzed. In Section 3, we propose the RA-DRN models with different granularities of multisource demand information. Furthermore, we transform the RA-DRN models into a single-level optimization problem. Section 4 presents numerical examples to validate the RA-DRN models with different conditions. We also provide conclusions and implications in Section 5.

2. Literature Review

This study is related to two streams of literature, namely, (1) the distributionally robust newsvendor problem, and (2) the newsvendor problem with censored demand. In the following, we briefly review each stream of literature.

2.1 Distributionally Robust Newsvendor Problem

The distributionally robust optimization approach is a primary method for solving the newsvendor problem with partial demand information. This review addressed the problem from two perspectives: Distributionally Robust Newsvendor with Moment Information (DRN-M) and Distributionally Robust Newsvendor with Distance-based Ambiguity Sets (DRN-D).

For the DRN-M problem, DMs exploit the common statistical properties of demand distribution (i.e., moment, unimodality, and asymmetry) to make robust order decisions (Delage and Ye, 2010; Natarajan et al., 2017). The maximin and minimax regret criteria are commonly adopted to handle partial information (Perakis and Roels, 2008; Natarajan et al., 2021). The maximin (minimax) criterion aims to maximize (minimize) the worst-case expected profit (cost) over all possible demand distributions. For example, Scarf (1958) proposes an order policy is robust to any distribution with a given mean and variance. Delage and Ye (2010) develop a new form of distributional set incorporating support, mean, and second-moment matrix derived from historical data of random variables. Natarajan et al. (2017) utilize second-order partitioned statistics, such as semivariance, to measure distribution asymmetry and derive a closed-form robust order quantity using mean, variance, and semivariance information. Das et al. (2021) address the distributionally robust newsvendor (DRN) problem with heavy-tailed demand, incorporating the first and the α-*th* moment as demand information. Under volatile demand and limited censored demand data, the order decisions with the maximin (minimax) criterion may be conservative. To address this issue, researchers usually used the minimax regret criterion to reduce decision conservation (Perakis and Roels, 2008). Zhu et al. (2013) adopt the relative minimax regret criterion in two information scenarios (i.e., the mean and the support, the mean and the variance). Yang et al. (2019) derive the robust closed-form order quantity and emissions control decisions only considering the support and the mean. The literature mentioned assumes that the moments of the underlying demand distributions can be accurately estimated before making decisions. However, accurately estimating the moments based on limited historical censored data presents a significant challenge.

Conversely, an alternative approach, known as distribution-based DRN, utilizes statistical distance measures to construct an uncertainty set through historical data. This includes methods such as ϕ-divergences (Ben-Tal et al., 2013; Rahimian et al., 2019) and Wasserstein distance (Lee et al., 2021; Chen and Xie, 2021; Laan et al., 2022). Rahimian et al. (2019) explore the effectiveness of the minimax distributionally robust optimization (DRO) model using the total variation distance. Lee et al. (2021) derived a closed-form robust order quantity with the Wasserstein metric. Considering the randomness of both demand and yield. Chen and Xie (2021) propose a minimax regret decision criterion to determine the optimal order quantity, extending their analysis to the Hurwicz criterion model for broader applicability. Gao and Kleywegt (2022) explore distributionally robust stochastic optimization (DRSO) using the Wasserstein distance, which was transformed into an infinite dimensional space of the inner maximization problem. Cheramin et al. (2022) propose computationally efficient inner and outer approximations for DRO using moment-based and Wasserstein distance-based ambiguity sets, demonstrating their effectiveness in production-transportation and multiproduct newsvendor problems. Gao (2022) provided the first finite-sample guarantees for Wasserstein distributionally robust optimization (DRO), overcoming the curse of dimensionality by balancing empirical mean and loss variation. Wang and Delage (2024) present a column generation-based decomposition scheme for a distributionally robust multi-item newsvendor problem, demonstrating superiority in out-of-sample performance with Wasserstein and mean absolute deviation ambiguity sets. However, if demand censoring is not considered, the poor quality of small-scale censored historical data may degrade the performance of DRN models.

2.2 Newsvendor Problem with Censored Demand

Due to inventory constraints, demand censoring is a common phenomenon but a key challenge for DMs in making order decisions. To tackle the inherent challenges posed by demand uncertainty and the loss of demand information, researchers develop different approaches based on the Bayesian update theory and the nonparametric approach (Bisi et al., 2011; Ban and Rudin, 2019; Chen et al., 2021).

Godfrey and Powell (2001) pioneer a distribution-free algorithm to tackle the newsvendor problem under conditions of censored demand. Their approach revolves around refining order decisions through continuous gradient estimates, thereby adeptly handling the complexities of demand uncertainty and censored data. Building upon this seminal work, Ding et al. (2002) elucidate that in the context of newsvendor models, dealing with censored demand necessitates higher optimal inventory levels than what would be suggested by a Bayesian myopic policy. Bensoussan et al. (2009) advance these insights by demonstrating that the myopic order quantity is always less than or equal to the optimal order quantity. They achieve this by employing unnormalized probability, which not only simplified the proof but also allowed for its extension to an infinite-horizon framework. Huh and Rusmevichientong (2009) propose nonparametric adaptive policies for inventory planning with censored demand, where only sales quantities are observed, and lost sales are unobservable. Besbes and Muharremoglu (2013) propose an exploration-exploitation framework to handle historical sales data with demand censoring. Jain et al. (2014) and Mersereau (2015) used Bayesian approaches to improve demand estimation and update demand distributions under censored observations in inventory management. However, these methods depend on specific assumptions regarding the demand distribution and the relationships among multisource data information. The assumptions may not be realistic under the conditions of limited historical censored data.

In the nonparametric domain, Huh et al. (2011) introduce nonparametric adaptive inventory control policies with demand censoring, using the Kaplan-Meier estimator. Bisi et al. (2015) present a non-parametric adaptive algorithm for the censored newsvendor problem, addressing both non-perishable and perishable inventory issues. Amjad and Shah (2017) propose the Universal Singular Value Thresholding (USVT) algorithm, to estimate true product demand at a specific store location and period based on noisy and potentially censored observations. Ban (2020) develops advanced machine-learning algorithms and nonparametric estimation procedures to improve inventory management, demonstrating significant cost reductions and computational efficiency while addressing demand censoring and feature incorporation in the newsvendor problem. Chen et al. (2020) and Chen et al. (2023) introduce nonparametric data-driven algorithms for dynamic pricing and inventory control with censored demand. Bu et al. (2023) establish a novel algorithm for single-product pricing with censored demand in an offline data-driven setting, maximizing expected revenue by hedging against distributional uncertainty from historical censored data. Lyu et al. (2024) build an upper confidence bound (UCB)-type learning framework for optimizing periodic-review lost-sales inventory systems with lead times using censored demand. Ding et al. (2024) develop the feature-based adaptive inventory algorithm and the dynamic shrinkage algorithm for stochastic periodic-review inventory systems with censored demand. However, in the aforementioned literature, DMs are assumed to be risk-neutral. According to Long and Nasiry (2015) and Shinde et al. (2021), the asymmetry effect on ordering can be observed under demand censoring and risk-averse DMs. It is worth exploring whether the risk-averse DMs with censored demand enhance the asymmetry effect, and how to reduce the additional effects.

Departing from the extant literature, we formulate a data-driven distributionally robust newsvendor problem that confronts risk-averse decision-makers with the challenge of demand censoring. We comprehensively harness limited historical censored data and external multi-source demand information to address this issue. We introduced the KM estimator into the proposed models, to capture the phenomenon of lost sales. Moreover, we adopt Phi-divergence and Wasserstein measures to achieve the local-external synergies with coarse-grained and finer-grained external demand information. Furthermore, we employ Prospect Theory to quantify the risk aversion of DMs. Interestingly, we create the innovative relationship of risk-averse coefficients, demand censoring, and profit structures of products, to eliminate risks associated with demand censoring. Moreover, we explained the asymmetry in ordering effects with censored demand and risk aversion.

3. The model

3.1 Problem Description

Considering a data-driven newsvendor problem, in which risk-averse DMs can exploit multisource demand information to determine order quantity q before a selling season. It is assumed that the demand x is stochastic, and the probability distribution function $F(\cdot)$ is unknown.

Instead, DMs can observe a limited historical censored sales dataset $\{(D_n, \delta_n)\}_{n=1}^{N}$. Constrained by available inventory Y_n^C, the sales data D_n may not fully recover actual demand Y_n^T ($n = 1, 2, \ldots, N$). The binary indicator δ_n describes whether actual demand exceeds inventory. When $Y_n^C < Y_n^T$, D_n is censored and the actual demand Y_n^T is not observed ($\delta_n = 0$). Conversely, the historical sales data $D_n = Y_n^T$ and $\delta_n = 1$. We have $D_n = \min(Y_n^C, Y_n^T)$ and $\delta_n = \mathbf{I}\left[Y_n^C \geq Y_n^T\right]$.

Facing locally limited historical censored data, DMs typically gather K sources of external demand information to enhance decision accuracy. In practices, the external information can be divided into two types: coarse-grained and finer-grained information. Coarse-grained information provides a broad, general representation of demand information with less detail, such as market trends. It is collected through various channels, including online retail websites (i.e., Amazon, Walmart, and Alibaba), Social Media Marketing (i.e., Facebook, TikTok, and X), and so on. While a discrete empirical distribution cannot fully capture the true demand distribution, it can still be obtained easily and inexpensively from observed data, expert opinions, simulations, and similar sources (Bayraksan and Love, 2015). Therefore, coarse-grained information is often represented by discrete empirical distributions. Let $\hat{P}_k^c$ denote the k-th source of the external information $\hat{\mathbf{P}}_{\mathbf{c}} = \left\{ \hat{P}_1^c, \ldots, \hat{P}_k^c, \ldots, \hat{P}_K^c \right\}$. $\hat{P}_k^c = \left\{ \hat{p}_{k,1}^c, \ldots, \hat{p}_{k,m}^c, \ldots, \hat{p}_{k,M}^c \right\}$, where the demand market has M scenarios and $\hat{p}_{k,m}^c$ is the corresponding probability of the scenario $\bar{d}_m$. Let $\bar{\mathbf{d}} = \left\{ \bar{d}_1, \ldots, \bar{d}_m, \ldots, \bar{d}_M \right\}$. Finer-grained information, on the other hand, assumes that the distribution type of external demand information and the true demand distribution belong to the same family and are known, but the corresponding parameters are not precisely estimated. Let $\hat{\mathbf{F}}_c = \left\{ \hat{F}_1^c, \ldots, \hat{F}_k^c, \ldots, \hat{F}_K^c \right\}$ represent the finer-grained distributional parameters. The distributions of $\hat{F}_{\mathrm{c}}$ are assumed to belong to the same family as the specific known distribution.

According to the granularity of the external demand information, we develop two distributionally robust newsvendor models to effectively fuse and utilize different multisource demand information, aiming to assist risk-averse DMs make more informed order decisions with limited demand information. For coarse-grained external information, we use the KL-divergence measure to construct the uncertainty set of the DRN model, which can effectively fuse isomorphic multisource information, denoted by the RA-DRN-KL model. On the contrary, we use the Wasserstein measure to build the uncertainty set, achieving the fusion of heterogeneous multisource information, denoted by the RA-DRN-WM model. The process of decision-making integrating multisource information is shown in Figure 1.

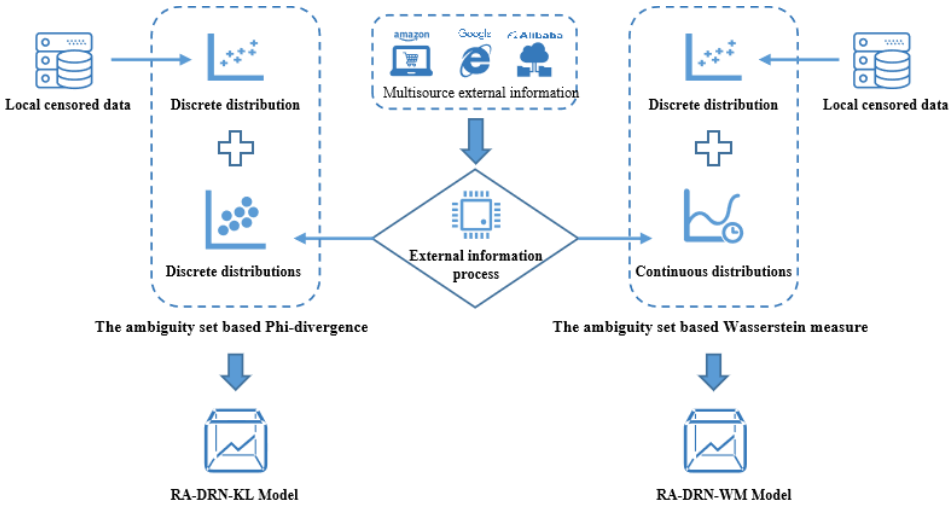


Figure 1: The process of decision-making integrating multisource information

The key notation is listed in Table 1.

Table 1: Summary of key notation

Indices	Definitions

m	Index of demand scenario ($m \in \{1,2,\ldots, M\}$)
n	Index of local historical sales period ($n \in \{1,2,\ldots, N\}$)
k	Index of external demand information source ($k \in \{1,2,\ldots, K\}$)
Parameters	Definitions
x	Random demand
r	Unit revenue
c	Unit procurement cost
l	Unit shortage cost
s	Unit salvage value
Y_n^C	The available inventory of the n-th historical period
Y_n^T	The realized demand of the n-th historical period
δ_n	A binary censoring indicator $\delta_n = \mathbf{I}\left[Y_n^C \geq Y_n^T\right]$
$\hat{P}_e$	Discrete empirical demand distribution from historical sales dataset
$\hat{P}_c$	Discrete reference distribution by learning external information
$\hat{F}_c$	Continuous reference distribution by learning external information
Decisions variable	
q	Order quantity of a product in the regular sales period

3.2 Benchmark

In the benchmark, we consider that risk-averse DMs determine the optimal order quantity based on the true demand distribution. It is assumed that unit procurement cost is c, and unit revenue is r. If the demand cannot be satisfied, a shortage cost l is incurred. If the retailer orders too many products, the leftover products are sold at salvage value s, and $s < r$. Thus, the profit function of DMs is developed as follows:

$$\pi(q) = r\min(x,q) + s(q-x)^+ - l(x-q)^+ - cq \tag{1}$$

where $(\cdot)^+ = \max(\cdot,0)$.

To characterize the risk aversion properties of DMs, the following segmented linear utility function is used:

$$U(W)=\begin{cases} W-W_0 & (W\ge W_0) \\ \lambda(W-W_0) & (W<W_0) \end{cases} \tag{2}$$

where W_0 is the reference point. Without the loss of generality, we set $W_0=0$.

In Eq. (2), λ is the risk aversion coefficient, which indicates the sensitivity of DMs to losses over gains. We assume that DMs are equally opposed to expected surplus loss and expected shortage loss. Based on PT, we can formulate the expected utility function with risk aversion as follows:

$$\begin{aligned} EU_F(q)=&(\lambda-1)(r-s)\left[E_F\min\left\{x,\frac{c-s}{r-s}q\right\}-\frac{c-s}{r-s}q\right]+(r+l-s)E_F\min\{x,q\} \\ &+(\lambda-1)lE_F\min\left\{x,\frac{r+l-c}{l}q\right\}-(c-s)q-\lambda lE_F[x] \end{aligned} \tag{3}$$

Eq. (3) reflects the expected surplus losses and gains with overordering, and the expected shortage losses and gains with underordering. Without the loss of generality, Eq. (3) is normalized. Let $r+l-s=1$, $\alpha=r-s$, $\beta=c-s$, where α and β mean unit shortage loss and unit residual loss, respectively. Eq. (3) can be transformed into:

$$\begin{aligned} EU_F(q)=&E_F\left[(\lambda-1)\min\{\alpha x,\beta q\}+(\lambda-1)\min\{(1-\alpha)x,(1-\beta)q\}\right] \\ &+E_F\left[\min\{x,q\}-\lambda(\beta q+(1-\alpha)x)\right] \end{aligned} \tag{4}$$

Given the known demand distribution function $F(\cdot)$, DMs maximize the expected utility function in Eq. (4) as follows:

$$\max_{q\ge 0} EU_F(q) \tag{5}$$

Let $q^*=\arg\max_{q\ge 0} EU_F(q)$ and $R^*=EU_F(q^*)$ be the optimal order quantity and the corresponding expected utility, respectively.

In practice, the full knowledge of the demand cannot be perfectly observed. The risk-averse DMs only capture the limited historical censored data, which does not ensure the precision of the order quantity. External multisource demand information can be exploited to improve the order decision. Next, we develop risk-averse DRN models by leveraging multisource demand information. According to the granularity of multisource external demand information, the proposed RA-DRN models fall into two categories, which are detailed in the following sections.

3.3 The construction of the uncertainty sets

The key step of RA-DRN models is the construction of the uncertainty sets. In this section, we construct the uncertainty sets based on the characteristics of the multisource demand information.

The multisource demand information exploited by DMs includes limited historical sales data with demand censoring and multisource external demand information.

For their partial information, DMs can use local historical sales data D to learn discrete empirical distribution. We assume that the interval of each scenario is denoted as $\left(d_m, d_{m+1}\right]$ $\left(m=1,2,\ldots,M\right)$. The midpoints of the intervals are taken as the demand level of each scenario, $\left\{\bar{d}_1, \bar{d}_2, \ldots, \bar{d}_M\right\}$. Referring to Kaplan and Meier (1958), DMs can adopt the Kaplan-Meier (KM) estimator to learn empirical distribution from censored observations $\{(D_n, \delta_n)\}_{n=1}^{N}$, which can consider the censored demand data and censoring indicator. We define this setting as a partially censored (PC) case. Here, $\hat{p}_m^{e,PC}$ is represented as follows:

$$\hat{p}_m^{e,PC} = \prod_{n_1: D_n \le d_m} \left(\frac{N - n_1}{N - n_1 + 1} \right)^{\delta_{n_1}} - \prod_{n_2: D_n \le d_{m+1}} \left(\frac{N - n_2}{N - n_2 + 1} \right)^{\delta_{n2}}, m = \left\{1, 2, \ldots, M\right\} \tag{6}$$

Let $\hat{P}_e^{PC} = (\hat{p}_1^{e,PC}, \hat{p}_2^{e,PC}, \cdots \hat{p}_M^{e,PC})$; n_1 and n_2 are a sequence of satisfying the condition $D_n \le d_m$ and $D_n \le d_{m+1}$, respectively.

For multisource external information, the demand information shared from multiple external sources may be incomplete, inaccurate or fuzzy. If the external demand information is coarse-grained, we select the phi-divergence measure to formulate the uncertainty set. Firstly, we fuse multisource external information into a barycenter $\hat{P}_c = \sum_{k=1}^{K} w_k \hat{P}_k^c$. This uncertainty set is constructed as follows:

$$F\ \left(\hat{P}_e^{PC}, \hat{P}_c\right) = \left\{ P \in \Re^M \ \middle|\ P \ge 0, \sum_{m=1}^{M} p_m = 1, I_\phi\left(P \middle\| \hat{P}_e^{PC}\right) \le \rho_e, I_\phi\left(P \middle\| \hat{P}_c\right) \le \rho_c \right\} \tag{7}$$

where $I_\phi\left(P \middle\| \hat{P}_e^{PC}\right) = \sum_{m=1}^{M} \phi\left(\frac{p_m}{\hat{p}_m^{e,PC}}\right) \hat{p}_m^{e,PC}$ and $I_\phi\left(P \middle\| \hat{P}_c\right) = \sum_{m=1}^{M} \phi\left(\frac{p_m}{\hat{p}_m^{c}}\right) \hat{p}_m^{c}$.

In Eq.(7), $\phi(\cdot)$ defined as a phi-divergence function, is a convex function on $\Re^+$; such that $\phi(1) := 0$, $0 \cdot \phi\left(\frac{0}{0}\right) := 0$, and $\phi\left(\frac{\varpi}{0}\right) := \varpi \lim_{\varsigma \to \infty} \frac{\phi(\varsigma)}{\varsigma}, (\forall \varpi > 0)$. According to Pardo (2006), ρ_e and ρ_c^k are set as $\frac{\phi^{''}(1)}{2N} \chi^2_{M-1,1-\tau}$, where $\chi^2_{M-1,1-\tau}$ is the $1-\tau$ quantile of the chi-squared distribution χ^2_{M-1} with *M*-1 degrees of freedom.

For the finer-grained information, the construction between local reference demand distribution

$\hat{P}_e^{PC}$ and external demand distributions $\hat{\mathbf{F}}_c$ is heterogeneous. Here, we use the Wasserstein measure to incorporate heterogeneous demand information into the uncertainty set. Similarly, we fuse multisource external information into a barycenter $\hat{F}_c = \sum_{k=1}^{K} w_k \hat{F}_k^c$. The uncertainty set with 1-Wasserstein distance can be defined as follows:

$$W_z(F,\hat{P}_e^{PC}) = \min_{\kappa\in\Gamma\left(F,\hat{P}_e\right)} \int_{\Xi\times\Xi} d\left(x,\hat{x}_e\right) d\kappa\left(x,\hat{x}_e\right) \tag{8}$$

$$W_z(F,\hat{F}_c) = \min_{\kappa\in\Gamma\left(F,\hat{F}_c\right)} \int_{\Xi\times\Xi} d\left(x,\hat{x}_c\right) d\kappa\left(x,\hat{x}_c\right) \tag{9}$$

where $W_z(\cdot)$ describes the 1-Wasserstein distance between true distribution F and reference distribution ($\hat{P}_e, \hat{F}_c$). $\Gamma(\cdot)$ denotes the set of all probability measures on $\Xi\times\Xi$ with marginal probabilities F and $\hat{P}_e\left(\hat{F}_c\right)$ on x and $\hat{x}_e\left(\hat{x}_c\right)$. And $d\left(x,\hat{x}\right)$ is an arbitrary norm function, denoted as $d\left(x,\hat{x}\right) := \left|x-\hat{x}\right|$.

Thus, this uncertainty set $F\left(\hat{P}_e^C, \hat{\mathbf{P}}_\mathbf{c}\right)$ is reformulated based on the heterogeneous demand information:

$$F_W\left(\hat{P}_e^{PC}, \hat{F}_c\right) = \left\{F \in \Re^+ \middle| W_z\left(F, \hat{P}_e^{PC}\right) \le \rho_W^e, W_z\left(F, \hat{F}_c\right) \le \rho_W^c \right\} \tag{10}$$

where ρ_W^c and ρ_W^e are the radius of the Wasserstein distance metric, respectively.

Next, we develop RA-DRN models based on the aforementioned uncertainty sets.

3.4 The formulation of the RA-DRN models

According to the uncertainty set $F\left(\hat{P}_e^{PC}, \hat{\mathbf{P}}_\mathbf{c}\right)$, we formulate a data-driven RA-DRN model (RA-DRN-KL) as follows:

$$\begin{aligned}
&\max_q \min_P EU_P(q) \\
&s.t\ I_\phi\left(P \middle\| \hat{P}_e^{PC}\right) \le \rho_e \\
&\quad I_\phi\left(P \middle\| \hat{P}_c\right) \le \rho_c \\
&\quad \sum_{m=1}^{M} p_m = 1 \\
&\quad q \ge 0
\end{aligned} \tag{11}$$

When DMs only use local censored data, the constraint $I_\phi\left(P \middle\| \hat{P}_c\right) \le \rho_c$ from Eq. (11) are removed, denoted by the RA-DRN-KL-S model.

In addition, Eq. (11) is a two-layer convex optimization problem, which can be solved by applying the Lagrange multiplier theory and duality theory. Taking the KL divergence (i.e., $\phi(t) = t\log t - t + 1$) as an example, the RA-DRN-KL model can be transformed into the simpler form, as shown in Proposition 1.

Proposition 1. With the KL divergence, the RA-DRN-KL model is equivalent to a single-level convex optimization problem:

$$\begin{aligned} &\max_{q,\lambda_e,\lambda_c} -(\lambda_e\rho_e + \lambda_c\rho_c) - (\lambda_e + \lambda_c)\log \mathrm{E}_{\hat{P}_e^{PC}}\left(e^{\mathbf{Z}}\right) \\ &\quad s.t. \quad q, \lambda_e, \lambda_c \geq 0 \end{aligned}$$

where $\mathbf{Z} = \{Z_1, Z_2, \ldots, Z_M\}$, $Z_m = \dfrac{-\pi_m + \lambda_c \log \vartheta_m}{\lambda_e + \lambda_c}$, $\vartheta_m = \dfrac{\hat{p}_m^c}{\hat{p}_m^{e,PC}}$ and

$$\pi_m = \left((\lambda - 1)\alpha \min\left\{ \bar{d}_m, \frac{\beta}{\alpha} q \right\} + (\lambda - 1)(1-\alpha)\min\left\{ \bar{d}_m, \frac{1-\beta}{1-\alpha} q \right\} + \min\{\bar{d}_m, q\} - \lambda\beta q - \lambda(1-\alpha)d_m \right).$$

Proposition 1 transforms the RA-DRN-KL model into a single optimization problem. This transformation not only facilitates the integration of coarse-grained external demand information but also mitigates the complexity of solving the model.

When the uncertainty set is $F_W\left(\hat{P}_e^{PC}, \hat{F}_c\right)$, we formulate a data-driven RA-DRN model (RA-DRN-WM) as follows:

$$\begin{aligned} &\max_q \min_F EU_F(q) \\ &s.t. W_z\left(F, \hat{P}_e^{PC}\right) \leq \rho_W^e \\ &\quad W_z\left(F, \hat{F}_c\right) \leq \rho_W^c \end{aligned} \tag{12}$$

If DMs can not collect finer-grained external demand information, the constraint $W_z\left(F, \hat{F}_c\right) \leq \rho_W^c$ removes from Eq. (12), which is denoted by the RA-DRN-WM-S model.

Similarly, the RA-DRN-WM model belongs to a two-layer convex optimization, which can be solved by Lagrange multiplier theory and maximin inequality theory. The RA-DRN-WM model can be transformed into a simpler form, as shown in Proposition 2.

Proposition 2: With the 1-Wasserstein distance, the RA-DRN-WM model can be equivalent to the following single-layer convex optimization problem:

$$\max_{q\geq 0} -\frac{\nu}{2}\rho_W^c - \frac{\nu}{2}\rho_W^e + \frac{1}{2}EU_{\hat{P}_e^{PC}}(q) + \frac{1}{2}EU_{\hat{F}_c}(q) \tag{13}$$

where $\nu = \max\{\lambda\alpha, \lambda(1-\alpha)\}$.

From Proposition 2, DMs can effectively collaborate local historical sales data and finer-grained external demand information, without increasing the complexity of the optimization model. Interestingly, the solution of order quantities q and dual variables (λ_e and λ_c) are independent and separated. The optimal dual variables are $\lambda_e^* = \lambda_c^* = \frac{1}{2}\max\left\{\lambda\alpha, \lambda(1-\alpha)\right\}$.

With solely local censored data, Eq. (13) can be rewritten as (RA-DRN-WM-S):

$$\max_{q\geq 0} -\nu\rho_W^e + EU_{\hat{P}_e^{PC}}(q) \tag{14}$$

4. Numerical experiment

In this section, we conduct numerical experiments to validate the effectiveness of the proposed models by exploiting multisource demand information. Accordingly, our analysis focuses on the following two aspects: (1) the performances of the RA-DRN models with different conditions; and (2) the impacts of the demand granularity, information process, demand censoring, and risk aversion level on the order decision.

4.1 Data generating

In our numerical experiments, data samples are generated based on the following data-generating procedure.

Parameter setting: According to Natarajan et al. (2017), the demand is usually assumed to be light-tailed demand (i.e., Normal distribution) and heavy-tailed distribution (i.e., Lognormal distribution). In addition, the non-negative of the demand should not be ignored. Thus, we discuss the performance of the proposed RA-DRN models on two underlying unknown distributions: the truncated Normal distribution and the truncated Lognormal distribution. The parameters of the distributions are set as shown in Table 2.

Table 2: The parameters setting of the demand distributions

Distribution	Parameters setting	Truncate interval
Truncated normal $N(\mu,\sigma^2)$	1. The mean μ is normalized to 100. 2. The normal variance is generated randomly by $\sigma \in (50,110)$.	$(0, \mu+2\sigma]$
Truncated lognormal $LN(\mu,\sigma^2)$	1. The mean μ is normalized to 100. 2. The normal variance is generated randomly by $\sigma \in (50,110)$.	$(0, \mu+2\sigma]$

Generation of historical dataset: For local censored information, we generate the censored observations $\{(D_n,\delta_n)\}_{n=1}^{N}$ in two steps. We first generate 30 samples ($N=30$) based on the underlying unknown distributions, $Y^T=\left\{Y_1^T,Y_2^T,\ldots,Y_{30}^T\right\}$. Then, we randomly select N_c samples from Y^T to replace as censored data $Y_n^C=\left(1-\nu_c\right)Y_n^T$, where ν_c depicts the proportion of stockout inventory. The corresponding censoring indicator is denoted by $\delta_n=\mathbf{I}\left[Y_n^C\geq Y_n^T\right]=0$. The remaining unselected samples are uncensored data $D_n=Y_n^T$, which the available inventory can fully satisfy actual demand ($Y_n^C\geq Y_n^T,\delta_n=1$). To obtain empirical distribution, we divide the market demand into 3 scenarios ($M=3$) based on the truncated demand interval, and the midpoint of each demand interval is taken as the demand level for each scenario.

Generation of external information: DMs can collect two external demand information from platforms/Apps ($K=2$). When the external demand information is coarse-grained, we independently draw 30 samples from the underlying unknown distribution twice and estimate discrete empirical distributions $\hat{\mathbf{P}}_{\mathbf{c}}=\left\{\hat{P}_1^c,\hat{P}_2^c\right\}$ by using the maximum likelihood estimation (MLE) method. If the external demand information is finer-grained, the fused continuous distribution $\hat{F}_c$ follows the same type of underlying unknown distribution but has a different standard deviation. We denote the standard deviation of true unknown demand distribution, and the collected demand distribution are σ and $\sigma+20$, respectively.

The other parameters in the experiments are shown as follows: $\tau=0.05$, $N=30$, $\rho_e=\rho_c=\frac{\phi^{''}(1)}{2N}\chi^2_{M-1,1,1-\tau}=0.12$, $\rho_c^W=\rho_c^W=1$, $M=3$, $K=2$, $\alpha=0.9$.

4.2 Experimental Setting

To evaluate the performance of different RA-DRN models, we populate numerical experiments based on the following procedure.

Step 1. Benchmark Setup. In the benchmark approach Eq. (5), DMs obtain order quantity q^* and corresponding expected utility $R^*=EU_F(q^*)$ with the true distribution F .

Step 2. Reference Models Setup. To explore the effect of information utilization and demand censoring on order decisions, we develop reference models with different information sources and censored demand learning process, as summarized in Table 2. In Table 3, we consider other two settings differentiated by the data available and learning to explore the effect of demand censoring and the value of external demand information. In the complete observable (CO) case, DMs can observe actual demand $Y^T=\left\{Y_1^T,Y_2^T,\ldots,Y_N^T\right\}$, which can use the traditional MLE method to obtain a discrete empirical reference distribution $\hat{P}_e^{CO}=(\hat{p}_1^{e,CO},\hat{p}_2^{e,CO},\mathrm{L}\ ,\hat{p}_M^{e,CO})$. In the second setting, DMs collect censored observations $\{(D_n,\delta_n)\}_{n=1}^N$, but directly use the MLE method to obtain the discrete empirical distribution $\hat{P}_e^{CC}=(\hat{p}_1^{e,CC},\hat{p}_2^{e,CC},\cdots,\hat{p}_M^{e,CC})$. We define this setting as the complete censored (CC) case, which does not utilize the censoring indicator (δ_n) information. Compared to the performance of RA-DRN models with CO and CC cases, we can discuss the effect of demand censoring on decision accuracy. Meanwhile, we can investigate the value of the censoring indicator by analyzing the differences between RA-DRN models with CO and PC cases. For simplicity, we use KL-S instead of RA-DRN models, and the rest are the same.

Table 3: The summary of each Model with different local informational levels

Local informational	Information source	Uncertain set	Order quantity	Expected utility
CO	single	KL divergence	q_{KL-S}^{CO}	R_{KL-S}^{CO}
		Wasserstein distance	q_{WM-S}^{CO}	R_{WM-S}^{CO}
	Multiple	KL divergence	q_{KL}^{CO}	R_{KL}^{CO}
		Wasserstein distance	q_{WM}^{CO}	R_{WM}^{CO}
CC	single	KL divergence	q_{KL-S}^{CC}	R_{KL-S}^{CC}
		Wasserstein distance	q_{WM-S}^{CC}	R_{WM-S}^{CC}
	Multiple	KL divergence	q_{KL}^{CC}	R_{KL}^{CC}
		Wasserstein distance	q_{WM}^{CC}	R_{WM}^{CC}
PC	single	KL divergence	q_{KL-S}^{PC}	R_{KL-S}^{PC}
		Wasserstein distance	q_{WM-S}^{PC}	R_{WM-S}^{PC}
	Multiple	KL divergence	q_{KL}^{PC}	R_{KL}^{PC}
		Wasserstein distance	q_{WM}^{PC}	R_{WM}^{PC}

Note: The expected utility under different models is calculated by substituting the obtained order quantity into the expected utility function Eq. (4) with the true distribution. We use subscripts $i\in\left\{CO,CC,PC\right\}$ and $j\in\left\{KL-S,WM-S,KL,WM\right\}$ to denote each model.

Step 3. Performance comparison. We compare the candidate approaches with the benchmark approach based on two indicators. Two indicators include the ratio of optimal order quantity Δq_j^i and the disappointment level of expected utility ΔR_j^i, where $\Delta q_j^i=\left|q_j^i-q^*\right|/q^*$, $\Delta R_j^i=\left|R_j^i-R^*\right|/R^*$,

$i \in \{CO, CC, PC\}$, and $j \in \{KL-S, WM-S, KL, WM\}$. The smaller ΔR_j^i represents the closer the ideal expected utility and the better performance of the proposed RA-DRN models.

Step 4. Robustness Validation. We repeat the above three steps 500 times and provide the box plots, the mean values of two indicators, to identify the effectiveness of the RA-DRN models with demand censoring and validate the values of multisource demand information, investigate the marginal value of proper censored demand information learning. Furthermore, we discuss how the demand censoring level affects the order decisions and the values of multisource demand information. Furthermore, the statistical significance is also validated through the Diebold-Mariano (DM) test.

4.3 The performance of RA-DRN models

Given demand censoring level ($N_c = 5$ and $\nu_c = 0.5$), we present the order decisions of risk-averse DMs ($\lambda = 1.5$) by using RA-DRN models under different demand distributions and residual loss β.

First, Table 4 displays the average difference of the RA-DRN models to the optimal order quantity and the corresponding expected utility with normal distributions. From Table 3, we can see that (1) under the same informational cases and information utilization methods, incorporating external information helps improve the accuracy of order decisions. For example, under residual loss $\beta = 0.5$ and the informational case is completely observable (CO case), the KL model with the local-external information reduces the relative errors of order decisions ΔR from 14.47% to 10.01%, compared with that of the KL-S model with single local demand information. (2) compared with the CO case, the complete censored data in the CC case contains little demand information, which amplifies the magnitude of decision errors. (3) the WM model significantly outperforms the KL model. This implies that the decision accuracy increases as the granularity of external demand information increases. Specifically, integrating finer-grained external demand information can reduce censoring-induced errors, because the performance of the WM model with CC and CO case is close. (4) the value of the censored indicator on decision accuracy is not always positive. For finer-grained external demand information, the ambiguity and uncertainty of the censored indicator (the WM model with the PC case) may induce DMs to make suboptimal order decisions.

Table 4: The performance of RA-DRN models under normal distributions

Informational cases	Model	$\beta = 0.4$		$\beta = 0.5$		$\beta = 0.6$	
		Δq_j^i	ΔR_j^i	Δq_j^i	ΔR_j^i	Δq_j^i	ΔR_j^i
CO	KL-S	26.31%	14.47%	34.28%	23.63%	28.80%	25.96%
	KL	20.73%	10.01%	33.26%	22.86%	28.96%	26.35%

	WM-S	10.44%	2.32%	28.25%	18.71%	35.86%	51.74%
	WM	4.24%	0.36%	3.33%	0.33%	3.47%	0.60%
CC	KL-S	39.22%	26.16%	40.74%	31.36%	32.08%	31.88%
	KL	25.25%	13.69%	35.28%	25.08%	30.37%	29.04%
	WM-S	13.34%	5.33%	32.43%	24.99%	37.74%	49.34%
	WM	4.24%	0.36%	3.33%	0.33%	3.47%	0.60%
PC	KL-S	25.85%	14.15%	35.16%	24.56%	29.37%	27.47%
	KL	21.66%	10.98%	34.06%	24.03%	29.10%	27.02%
	WM-S	10.59%	2.47%	28.19%	18.75%	36.66%	54.54%
	WM	9.02%	1.60%	14.44%	5.60%	13.84%	8.59%

To show the impact of information learning and utilization more clearly, we present boxplots in Figure 2. The boxplot subfigures describe the statistical results of four RA-DRN models using a five-number summary, including the minimum, first quartile, mean, third quartile, and maximum. We compare four RA-DRN models with different scenarios to understand the efficient conditions of each information utilization strategy and the value of each demand information.

It is obvious that integrating external information always improves the performance of models, compared to the models with solely local historical data. We observe that the interquartile range (IQR) of data-driven RA-DRN models integrating external demand information is lower than that of local demand information across different scenarios. With single local historical data, the IQR and mean of the KL-S/WM-S model with the PC case are close to that of the KL-S/WM-S model with the CO case. This implied that without external demand information, using the KM estimator can reduce the negative effects of demand censoring with two information utilization strategies (the KL-S and WM-S models). Additionally, the above conclusions are also validated in Figure 2.

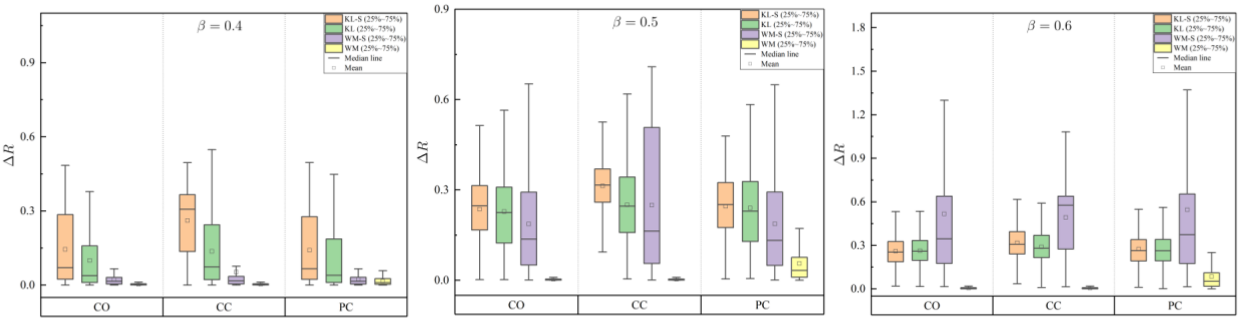


Figure 2 The comparison of the RA-DRN models under normal distribution

Moreover, we conduct the Diebold-Mariano (DM) test to further verify the significant differences between the RA-DRN models incorporating multisource demand information with the RA-DRN models using local historical data. The results are shown in Table 5. We can observe that for most conditions, incorporating external demand information is beneficial at the 1% significance level.

Table 5: the significance of information learning under normal distribution

Information level	Model	$\beta=0.4$	$\beta=0.5$	$\beta=0.6$
		DM value$^{\text{Significance}}$		
CO	KL-S vs KL	5.03***	-0.32	-1.12
	WM-S vs WM	2.52***	15.46***	9.81***
CC	KL-S vs KL	13.44***	7.41***	7.03***
	WM-S vs WM	5.83***	18.21***	12.78***
PC	KL-S vs KL	-1	-1.23	-1
	WM-S vs WM	2.35***	13.46***	10.04***

Note: (1) DM value describes the values of the DM statistic. (2) “***” indicates significance at the level of 1%, i.e., p-value<0.01; “**” indicates significance at the level of 5%, i.e., p-value<0.05; “*” indicates significance at the level of 10%, i.e., p-value<0.2.

Further, we discuss the performance of the RA-DRN models under lognormal distribution in Figure 3 and Table 6. From Figure 3 and Table 6, we can find the result is similar to that of the normal distribution: (1) the external information can effectively eliminate the demand uncertainty (2) the processing of censored data can improve the performance of the RA-DRN models in most scenarios.

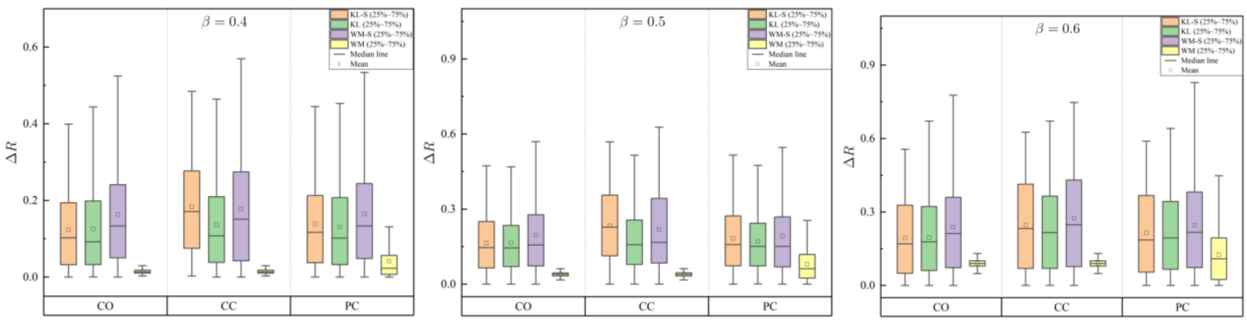
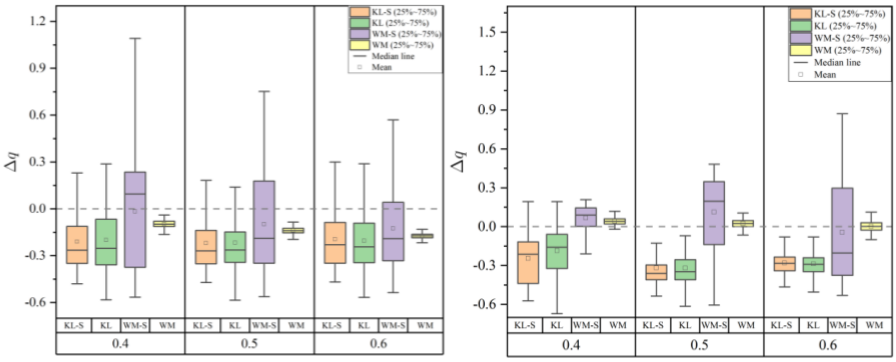


Figure 3: the comparison results of the RA-DRN models under lognormal distribution

Table 6: the significance of information learning under lognormal distribution

Information level	Model	$\beta=0.4$	$\beta=0.5$	$\beta=0.6$
		DM value$^{\text{Significance}}$		
CO	KL-S vs KL	-1.52	0.12	0.05
	WM-S vs WM	13.02***	14.12***	14.02***
CC	KL-S vs KL	6.16***	-1	-1
	WM-S vs WM	14.72***	17.22***	17.39***
PC	KL-S vs KL	0.61	2.61***	-1
	WM-S vs WM	10.49***	3.74***	7.75***

We can observe that the KL model with coarse-grained external demand information may be inferior to the WM-S model with solely local historical data in some cases. The result is because the order quantities deviation of the other three data-driven models have higher robustness and conservation, compared with the WM-S model. As we can see from Figure 4 with the CO case, for heavy-tailed distribution (lognormal distribution), the order quantities deviation of the WM-S model decreases with residual loss β increases. Conversely, for light-tailed distribution (normal distribution), the order quantities deviation of the WM-S model increases with residual loss β increases. This phenomenon is because the WM-S model with 1-Wasserstein distance is equivalent to the optimal solution of sample average approximation (SAA), in which the decisions depend on the characteristics of profit structures and market demand. Thus, with coarse-grained external information, the KL model sometimes is inferior to the WM-S model.



(a) Lognormal distribution (b) Normal distribution

Figure 4: The ratio differences of optimal order quantity under the CO case

To sum up, facing limited historical censored data, DMs prefer to collect external demand information. This is because the risk of wrong order decisions induced by limited historical sales data misguide risk-averse DMs in making more conservative order decisions. DMs with limited ability and budget, can both utilize the coarse-grained external demand information and the KM estimator, which can mitigate the negative effect of demand censoring. The finer-grained external demand information provides more distinct description of market demand, increasing risk-averse DMs' trust in demand information. Specifically, as residual loss increases, the available value of finer-grained external demand information and the aspiration of collecting finer-grained external demand information increases.

4.4 The effect of demand censoring

To further evaluate the effect of demand censoring on the performance of the proposed models, we explore the differences with three censored levels (N_c, ν_c), including ($5, 0.3$), ($10, 0.3$), and ($10, 0.5$). The experimental setup is the same as for Section 5.3. The statistical results of the disappointment level of the expected utility ΔR_j^i are shown in Figure 5.

Figure 5 depicts the relative loss of the RA-DRN models with complete censored (CC) and partial censored (PC) cases. The left and right panels of boxplots correspond to the normal distribution and lognormal distribution, respectively. The conclusions in Section 5.3 can also be obtained under the other censored levels. Meanwhile, we can find that as censored levels aggravate, the value of censored indicators in improving order accuracy increases. Novelty, its value is close to that of coarse-grained external information, because of the similar performance between the KL-S and KL models with PC strategies. These results imply that the value of coarse-grained demand information is limited. In addition, the statistical results of RA-DRN models with external demand information are similar with different censored levels. This suggests that integrating external demand information can mitigate the adverse impacts of limited data and demand censoring, maintain the robustness of order decisions, and enable DMs to effectively treat sales data as if it were uncensored. Moreover, the higher demand granularity can be collected for DMs, the more value external demand information has, which allows DMs to obtain more accurate order decisions.

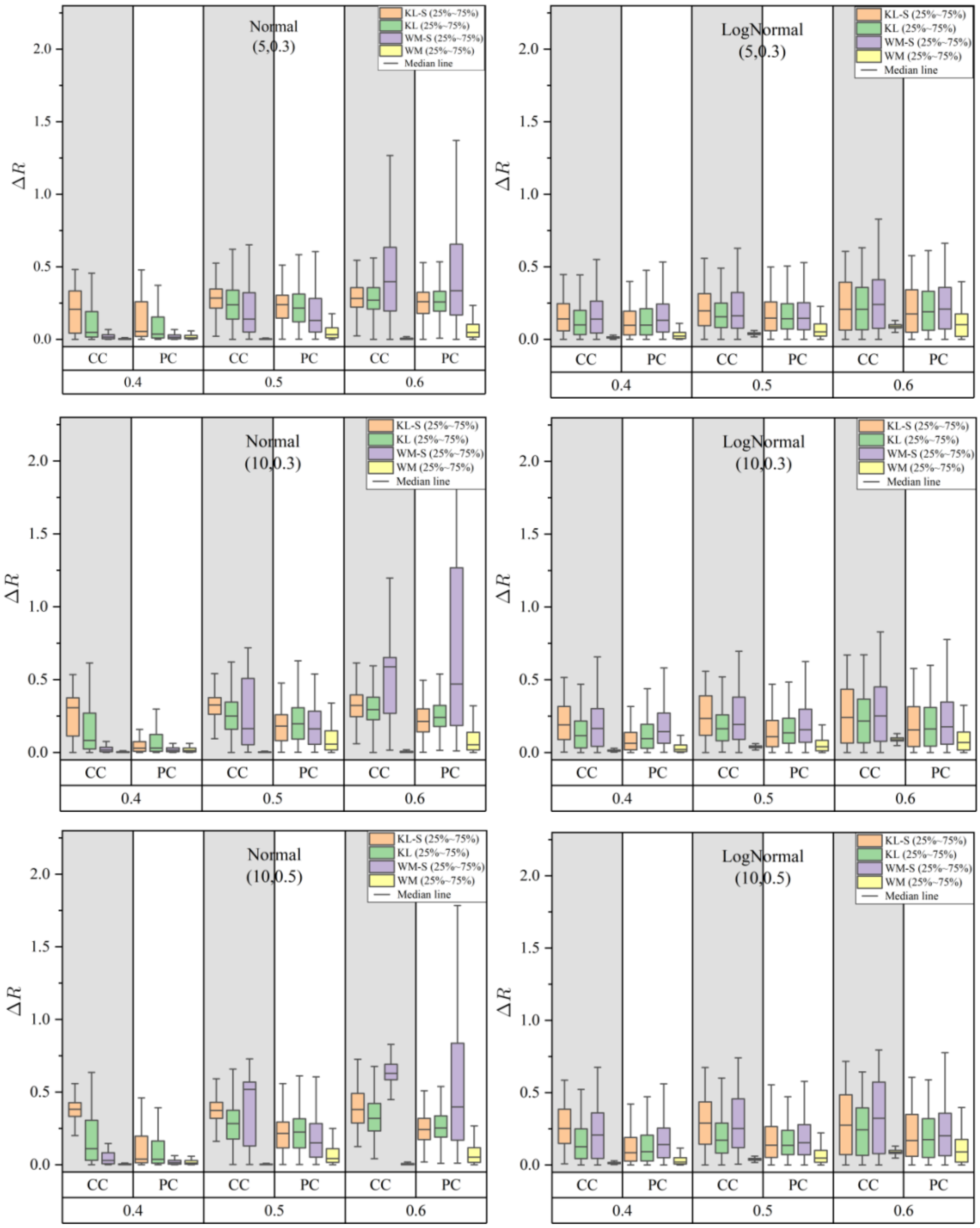


Figure 5: the performances of the RA-DRN models with different censored levels

Further, through the DM test, we investigate whether the outperformance of the RA-DRN models with multisource demand information (KL and WM models) satisfy the statistical test through the Diebold-Mariano (DM) test, compared with the data-driven RA-DRN models with single censored demand information (KL-S and WM-S models). We can observe, for the CO scenario, that the KL and WM model separately outperform the other two models (KL-S and WM-S models) at the 1% significance level, as shown in Table 7.

Table 7: the results of the DM test with different censored levels

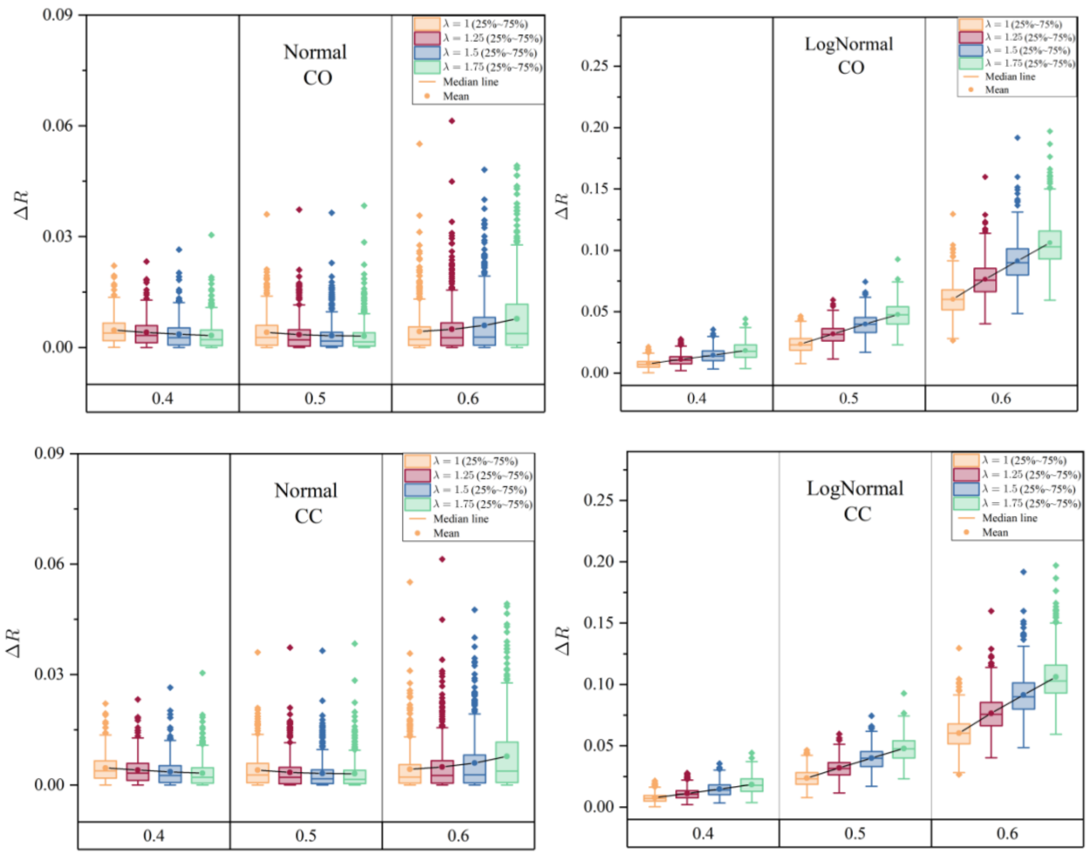
Information level	Distribution	Normal			Lognormal		
	β	0.4	0.5	0.6	0.4	0.5	0.6
	Model1 vs. Model 2	DM value$^{\text{Significance}}$					
1	Demand censoring level (5,0.3)						
CC	KL-S vs. KL	8.21***	3.11***	2.10**	4.57***	-1	-1.00
	WM-S vs. WM	3.51***	15.93***	9.76***	13.99***	16.13***	17.48***
PC	KL-S vs. KL	-1	-1.75	-1	-1	-1	-0.90
	WM-S vs. WM	1.20	13.67***	10.09***	10.62***	3.58***	6.65***
2	Demand censoring level (10,0.3)						
CC	KL-S vs. KL	9.40***	7.11***	4.75***	-1.00	-1.00	-1.00
	WM-S vs. WM	5.75***	18.30***	13.33***	16.41***	17.59***	18.14***
PC	KL-S vs. KL	-3.71	-5.42	-1	-1	-2.57	-1
	WM-S vs. WM	2.85***	13.76***	13.74***	10.19***	5.54***	5.53***
3	Demand censoring level (10,0.5)						
CC	KL-S vs. KL	-1.00	4.20***	-1.00	-1.00	-1.00	-1.07
	WM-S vs. WM	11.65***	29.82***	27.01***	18.87***	20.16***	19.78***
PC	KL-S vs. KL	-1	-1	-1	-2.52	2.48***	1.95**
	WM-S vs. WM	2.23**	27.01***	12.03***	9.82***	6.10*	6.10***

In conclusion, the incremental value of external demand information becomes higher as censored levels aggravate. That is, DMs can integrate external demand information to make more robust order decisions and avoid the adverse effect of demand censoring, especially at highly censored levels. It provides risk-averse DMs with robust means to cope with limited inaccurate demand information.

4.5 The effect of risk aversion

In this section, we go deep into how the risk aversion of DMs affects decision accuracy. Based on the above analysis, the WM model exhibits the highest decision accuracy and the strongest robustness. To further validate the performance of the WM model, we focus on the effect of risk aversion coefficients on its performance. Note that the other parameters setting is the same as Section 4.3.

Figure 6 depicts the differences in the WM model between risk-neutral $\lambda = 1$ and risk-averse DMs $\lambda \in \{1.25, 1.5, 1.75\}$. From Figure 6, it is evident that given information cases, risk-averse DMs make more conservative ordering decisions. Moreover, the effect of risk aversion get more and more obvious as the residual loss β increases in most scenarios. The result illustrates that using finer-grained demand information can effectively hedge conservation induced by both high risk-aversion coefficients and high residual loss. The higher the risk aversion coefficients, the more incentive DMs gather more detailed demand information. In addition, when the tail of distribution becomes heavier, DMs with lower risk aversion coefficients make more accurate order decisions.



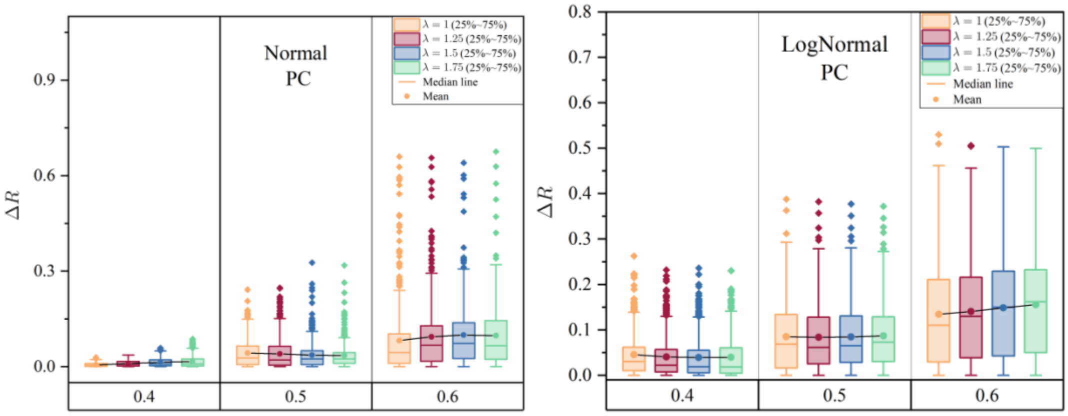
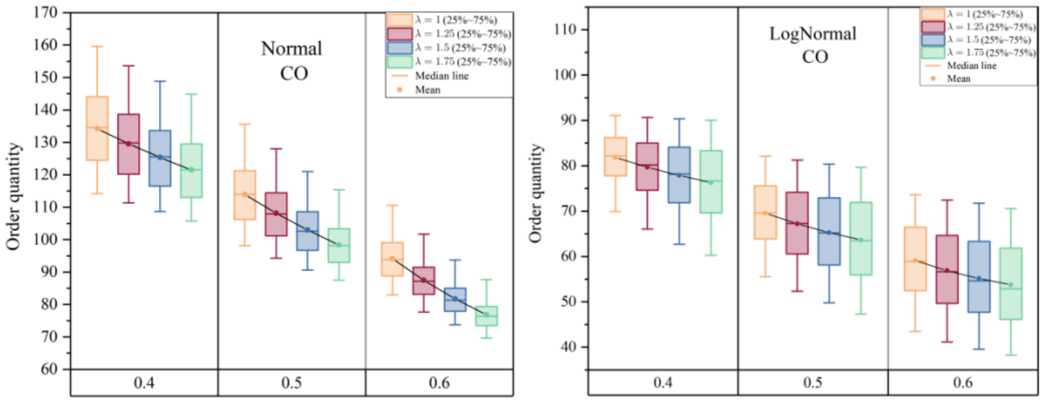


Figure 6: The performance of the WM model with different risk aversion degree

Interestingly, for light-tailed distribution (normal distribution) and low residual loss $\beta \in [0.4, 0.5]$, risk aversion coefficients have a positive effect on decision accuracy obtained by the WM model. The consequence is that the higher the risk aversion coefficients, the fewer ordering products. Its behavior of less ordering can provide DMs with more informed order decisions with low residual loss. It implies that for the stable demand market (light-tailed demand distribution), DMs selling products with low residual loss should have a highly risk-averse attitude and adopt a less-order strategy.

In addition, as Figure 7 shows, given demand distribution and residual loss, the order quantity with limited demand information is decreasing in risk aversion coefficients $\frac{\partial q}{\partial \lambda} < 0$. This conclusion is the same as the existing literature (Nagarajan and Shechter, 2014).



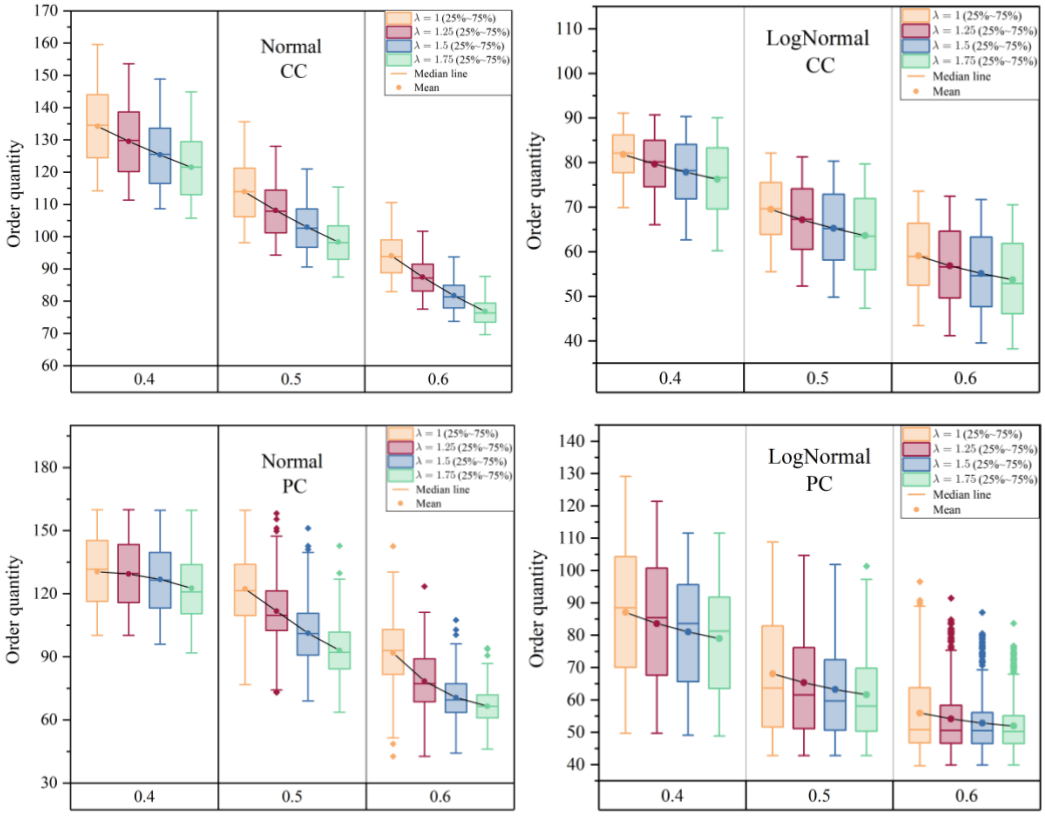


Figure 7: The order quantity with different risk aversion coefficients

5. Conclusions and Implications

In this study, we propose an integration of a distributionally robust optimization framework into the risk-averse newsvendor problem, accounting for demand censoring. Based on the phi-divergence measure and the Wasserstein measure, multisource external demand information is integrated into uncertainty sets, and then is effectively exploited to enhance the decision accuracy of risk-averse DMs with limited historical censored sales data. A series of numerical experiments are designed to explore the critical role of external demand information and the impact of the censoring indicator. Through these experiments, we scrutinize the effects of varying levels of demand censoring and risk aversion coefficients on the order decisions, thereby providing actionable insights into the decision-making process.

(1) The Value of External Demand Information

Incorporating external demand information can stabilize the decision accuracy of risk-averse DMs facing demand censoring. Across various censored levels, the performances of the RA-DRN models with local-external synergy remain consistent. This suggests that the proposed models with local-external synergy possess stronger robustness and can effectively avoid the adverse effect of demand censoring. As the censored level increases, the incremental value of external demand information on order decisions becomes more pronounced. Interestingly, DMs with access to finer-grained external demand information can treat censored demand as uncensored demand.

(2) The Effect of Risk Aversion

The effect of risk aversion on order decisions is not universally negative. In most cases, risk-averse DMs make more conservative order decisions, which can lead to reduced decision accuracy. Compared with risk-neutral DMs, risk-averse DMs show a preference for gathering external demand information with a higher granularity. The negative effect of risk aversion coefficients on order decision intensifies as the tail of the underlying demand distribution becomes heavier. In contrast, for light-tailed distribution and low residual loss, the conservative order quantity is beneficial to improve decision accuracy.

(3) Dual Role of the Censoring Indicator

The censored indicator exhibits the dual effect. Without finer-grained external demand information, DMs should use the KM estimator to exploit the censored indicator and describe demand censoring, which can reduce censoring-induced errors. Conversely, when incorporating finer-grained external demand information, the censored indicator information reduces the decision accuracy. The phenomenon is because the value of censored indicator information may be ambiguous and uncertain.

In summary, we provide a thorough insight into risk-averse DMs. According to the characteristics of collected demand information and risk aversion, DMs can implement appropriate approaches to weaken the effects associated with demand censoring and risk aversion. Our approaches provide suitable order decisions for risk-averse DMs with demand censoring, which can be extended to decision-making for both ordering and pricing problems. Moreover, demand and yield randomness could also be considered the newsvendor problem of risk-averse DMs with demand censoring simultaneously.

Reference

Amjad, M., & Shah, D. (2017, February). Trading bitcoin and online time series prediction. In NIPS 2016 time series workshop (pp. 1-15). PMLR.

Scarf, H. (1958). A min-max solution of an inventory problem. Studies in the Mathematical Theory of Inventory and Production.

Ban, G. Y. (2020). Confidence intervals for data-driven inventory policies with demand censoring. *Operations Research*, 68(2), 309-326.

Ban, G. Y., & Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research*, 67(1), 90-108.

Bayraksan, G., & Love, D. K. (2015). Data-driven stochastic programming using phi-divergences. In The operations research revolution (pp. 1-19). INFORMS.

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

Bensoussan, A., Cakanyıldırım, M., & Sethi, S. (2009). Censored newsvendor model revisited with unnormalized probabilities. *Journal of Industrial Management and Optimization*, 5(2).

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

Bisi, A., Dada, M., & Tokdar, S. (2011). A censored-data multiperiod inventory problem with newsvendor demand distributions. *Manufacturing & Service Operations Management*, 13(4), 525-533.

Bisi, A., Kalsi, K., & Abdollahian, G. (2015). A non-parametric adaptive algorithm for the censored newsvendor problem. *IIE Transactions*, 47(1), 15-34.

Bu, J., Simchi-Levi, D., & Wang, L. (2022). Offline pricing and demand learning with censored data. *Management Science*, 69(2), 885-903.

Chang, S., Li, J., Sethi, S. P., & Wang, X. (2024). Risk hedging for VaR-constrained newsvendors. *Transportation Research Part E: Logistics and Transportation Review*, 181, 103365.

Chen, B., Chao, X., & Shi, C. (2021). Nonparametric learning algorithms for joint pricing and inventory control with lost sales and censored demand. *Mathematics of Operations Research*, 46(2), 726-756.

Chen, B., Chao, X., & Wang, Y. (2020). Data-based dynamic pricing and inventory control with censored demand and limited price changes. *Operations Research*, 68(5), 1445-1456.

Chen, B., Wang, Y., & Zhou, Y. (2023). Optimal policies for dynamic pricing and inventory control with nonparametric censored demands. *Management Science*, 70(5), 3362-3380.

Chen, Z., & Xie, W. (2021). Regret in the newsvendor model with demand and yield randomness. *Production and Operations Management*, 30(11), 4176-4197.

Chengyi Lyu, Huanan Zhang, Linwei Xin (2024) UCB-Type Learning Algorithms with Kaplan–Meier Estimator for Lost-Sales Inventory Models with Lead Times. *Operations Research*, 72(4):1317-1332.

Cheramin, M., Cheng, J., Jiang, R., & Pan, K. (2022). Computationally efficient approximations for distributionally robust optimization under moment and Wasserstein ambiguity. *INFORMS Journal on Computing*, 34(3), 1768-1794.

Das, B., Dhara, A., & Natarajan, K. (2021). On the heavy-tail behavior of the distributionally robust newsvendor. *Operations research*, 69(4), 1077-1099.

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

Ding, J., Huh, W. T., & Rong, Y. (2024). Feature-based inventory control with censored demand. *Manufacturing & Service Operations Management*, 26(3), 1157-1172.

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

Eeckhoudt, L., Gollier, C., & Schlesinger, H. (1995). The risk-averse (and prudent) newsboy. *Management Science*, 41(5), 786-794.

Gao, R. (2022). Finite-sample guarantees for Wasserstein distributionally robust optimization: Breaking the curse of dimensionality. *Operations Research*, 71(6), 2291-2306.

Gao, R., & Kleywegt, A. (2022). Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research*, 48(2), 603-655.

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

Ha, A. Y., Luo, H., & Shang, W. (2022). Supplier encroachment, information sharing, and channel structure in online retail platforms. *Production and Operations Management*, 31(3), 1235–1251.

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

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

Jain, A., Rudi, N., & Wang, T. (2015). Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need. *Operations Research*, 63(1), 134-150.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282), 457-481.

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.

Li, G., Tian, L., & Zheng, H. (2021). Information sharing in an online marketplace with co-opetitive sellers. *Production and Operations Management*, 30(10), 3713-3734.

Liu, C., & Zhu, W. (2024). Newsvendor conditional value-at-risk minimisation: A feature-based approach under adaptive data selection. *European Journal of Operational Research*, 313(2), 548-564.

Lin, S., Chen, Y., Li, Y., & Shen, Z. J. M. (2022). Data-driven newsvendor problems regularized by a profit risk constraint. *Production and Operations Management*, 31(4), 1630-1644.

Long, X., & Nasiry, J. (2015). Prospect theory explains newsvendor behavior: The role of reference points. *Management Science*, 61(12), 3009-3012.

Mersereau, A. J. (2015). Demand estimation from censored observations with inventory record inaccuracy. *Manufacturing & Service Operations Management*, 17(3), 335-349.

Nagarajan, M., & Shechter, S. (2014). Prospect theory and the newsvendor problem. *Management Science*, 60(4), 1057-1062.

Natarajan K, Sim M, & Uichanco J. (2017). Asymmetry and Ambiguity in Newsvendor Models. *Management Science*, 64(7), 3146-3167.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382-2393.

Perakis, G., & Roels, G. (2008). Regret in the newsvendor model with partial information. *Operations research*, 56(1), 188-203.

Rahimian, H., Bayraksan, G., & Homem-de-Mello, T. (2019). Identifying effective scenarios in distributionally robust stochastic programs with total variation distance. Math*ematical Programming*, 173(1), 393-430.

Rudi, N., & Drake, D. (2014). Observation bias: The impact of demand censoring on newsvendor level and adjustment behavior. *Management Science*, 60(5), 1334-1345.

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

Sayın, F., Karaesmen, F., & Özekici, S. (2014). Newsvendor model with random supply and financial hedging: Utility-based approach. *International Journal of Production Economics*, 154, 178-189.

Shinde, A., Mehta, P., & Amit, R. K. (2021). An experimental investigation of newsvendor decisions under ambiguity. *International Journal of Production Research*, 59(19), 5960-5971.

Surti, C., Celani, A., & Gajpal, Y. (2020). The newsvendor problem: The role of prospect theory and feedback. *European Journal of Operational Research*, 287(1), 251-261.

van der Laan, N., Teunter, R. H., Romeijnders, W., & Kilic, O. A. (2022). The data-driven newsvendor problem: Achieving on-target service-levels using distributionally robust chance-constrained optimization. *International Journal of Production Economics*, 249, 108509.

Wang, C. X., Webster, S., & Suresh, N. C. (2009). Would a risk-averse newsvendor order less at a higher selling price?. *European Journal of Operational Research*, 196(2), 544-553.

Wang, S., & Delage, E. (2024). A column generation scheme for distributionally robust multi-item newsvendor problems. *INFORMS Journal on Computing*.

Xue, W., Ma, L., Liu, Y., & Lin, M. (2022). Value of inventory pooling with limited demand information and risk aversion. *Decision Sciences*, 53(1), 51-83.

Yang, C. H., Ma, X., & Talluri, S. (2019). Optimal acquisition decision in a remanufacturing system with partial random yield information. *International Journal of Production Research*, 57(6), 1624-1644.

Yang, C. H., Su, X. L., & Wu, P. (2023). A data-driven distributionally newsvendor problem for edge-cloud collaboration in intelligent manufacturing systems. *Engineering Applications of Artificial Intelligence*, 126, 106995.

Zhang, H., Chao, X., & Shi, C. (2020). Closing the gap: A learning algorithm for lost-sales inventory systems with lead times. *Management Science*, 66(5), 1962-1980.

Zhang, Y., & Siemsen, E. (2019). A meta-analysis of newsvendor experiments: Revisiting the pull-to-center asymmetry. *Production and Operations Management*, 28(1), 140-156.

Zhu, Z., Zhang, J., & Ye, Y. (2013). Newsvendor optimization with limited distribution information. *Optimization methods and software*, 28(3), 640-667.

Appendix

Proof of Proposition 1

Referred to Zhang et al. (2020) and Yang et al. (2023), the Lagrangian function of the inner minimization problem about Eq. (11):

$$\begin{aligned} &K(q,t_m,\lambda_e,\lambda_c,\eta) \\ &= -(\lambda_e\rho_e+\lambda_c\rho_c)-\eta-\sum_{m=1}^{M}\hat{p}_m^e[-t_m(\pi_m+\eta)]-\lambda_e\phi(t_m)-\lambda_c\vartheta_m\phi(\frac{t_m}{\vartheta_m}) \end{aligned} \tag{A.1}$$

where $t_m=\frac{p_m}{\hat{p}_m^{e,PC}}$, $\vartheta_m=\frac{\hat{p}_m^c}{\hat{p}_m^{e,PC}}$, and

$$\pi_m=\left((\lambda-1)\alpha\min\left\{\bar{d}_m,\frac{\beta}{\alpha}q\right\}+(\lambda-1)(1-\alpha)\min\left\{\bar{d}_m,\frac{1-\beta}{1-\alpha}q\right\}+\min\{\bar{d}_m,q\}-\lambda\beta q-\lambda(1-\alpha)d_m\right).$$

The dual problem of Eq. (11) can be written as:

$$\begin{aligned} &\max_{q,\lambda_e,\lambda_c,\eta}\ \min_{t_m\geq 0} K(q,t_m,\lambda_e,\lambda_c,\eta) \\ &s.t. q,\lambda_e,\lambda_c\geq 0 \end{aligned} \tag{A.2}$$

Under the condition of the KL divergence $\phi(t)=t\log t-t+1$, we take the first-order derivative of the inner minimization problem of Eq. (A.2) with respect to t_m. The optimal solution of t_m as follows:

$$t_m^* = e^{Z_m}$$

where $Z_m = e^{\frac{-\pi_m + \lambda_c \log \vartheta_m}{\lambda_e + \lambda_c}}$.

Combining $t_m = \frac{p_m}{\hat{p}_m^{e,PC}}$ with $\sum_{m=1}^{M} p_m = 1$, we can obtain $t_m^* = e^{Z_m} \big/ \mathrm{E}_{\hat{P}_e^{PC}}\left(e^{\mathbf{Z}}\right)$ and $\eta = \left(\lambda_e + \lambda_c\right) \log E_{\hat{P}_e^{PC}}\left(e^{\mathbf{Z}}\right)$ ($\mathbf{Z} = \left\{Z_1, Z_2, \ldots, Z_M\right\}$). Then, putting t_m^* and η into the problem (A.2), we have:

$$\begin{aligned} &\max_{q,\lambda_e,\lambda_c} -\left(\lambda_e \rho_e + \lambda_c \rho_c\right) - \left(\lambda_e + \lambda_c\right) \log \mathrm{E}_{\hat{P}_e^{PC}}\left(e^{\mathbf{Z}}\right) \\ &\quad s.t. \quad q, \lambda_e, \lambda_c \geq 0 \end{aligned}$$

Q.E.D.

Proof of Proposition 2

For the RA-DRN-WM model, we first examine the inner maximum problem.

$$\begin{aligned} &\min_F EU_F(q) \\ &s.t. W_z\left(F, \hat{P}_e^{PC}\right) \leq \rho_W^e \\ &\quad W_z\left(F, \hat{F}_c\right) \leq \rho_W^c \end{aligned} \tag{A.3}$$

From the Lagrangian duality theory, we obtain:

$$\begin{aligned} G_W\left(\lambda_e, \lambda_c\right) = &\min_F \max_{\lambda_e,\lambda_c \geq 0} EU_F(q) \\ &-\lambda_e\left(\rho_W^e - W_z\left(F, \hat{P}_e\right)\right) - \lambda_c\left(\rho_W^c - W_z\left(F, \hat{F}_c\right)\right) \end{aligned} \tag{A.4}$$

Then, based on the minimax inequality:

$$\begin{aligned} G_W\left(\lambda_e, \lambda_c\right) \geq &\max_{\lambda_e,\lambda_c \geq 0} -\lambda_e \rho_W^e - \lambda_c \rho_W^c \\ &+\min_F\left[EU_F(q) + \lambda_e W_z\left(F, \hat{P}_e\right) + \lambda_c W_z\left(F, \hat{F}_c\right)\right] \end{aligned} \tag{A.5}$$

where λ_e and λ_c are Lagrangian multipliers associated with Wasserstein distance constraints.

According to the c-transformation proposed by Ambrosio et al. (2013), Eq. (8) and Eq. (9) can be recast as:

$$W_z\left(F, \hat{P}_e\right) = \max_{\varphi_e} \int_{\Xi} \varphi_e(x) F(dx) + \int_{\Xi} \varphi_c^1\left(\hat{x}_e\right) \hat{P}_e\left(d\hat{x}_e\right) \tag{A.6}$$

$$W_z\left(F, \hat{F}_c\right) = \max_{\varphi_c} \int_{\Xi} \varphi_c(x) f(x) dx + \int_{\Xi} \varphi_c^2\left(\hat{x}_c\right) \hat{f}_c\left(\hat{x}_c\right) d\hat{x}_c \tag{A.7}$$

So, we can obtain:

$$
\begin{aligned}
A &= \min_F \left[EU_F(q) + \lambda_e W_z\left(F, \hat{P}_e\right) + \lambda_c W_z\left(F, \hat{F}_c\right) \right] \\
&= \min_F EU_F(q) + \lambda_e \max_{\varphi_e} \int_\Xi \varphi_e(x) F(dx) + \int_\Xi \varphi_c^1(\hat{x}_e) \hat{P}_e(d\hat{x}_e) \\
&\quad + \lambda_c \max_{\varphi_c} \int_\Xi \varphi_c(x) f(x) dx + \int_\Xi \varphi_c^2(\hat{x}_c) \hat{f}_c(\hat{x}_c) d\hat{x}_c \\
&= \min_F EU_F(q) - \lambda_e \min_{\varphi_e} \int_\Xi -\varphi_e(x) F(dx) - \int_\Xi \varphi_c^1(\hat{x}_e) \hat{P}_e(d\hat{x}_e) \\
&\quad - \lambda_c \min_{\varphi_c} \int_\Xi -\varphi_c(x) f(x) dx - \int_\Xi \varphi_c^2(\hat{x}_c) \hat{f}_c(\hat{x}_c) d\hat{x}_c
\end{aligned} \tag{A.8}
$$

Setting $\varphi_e(x) = \frac{\Pi(q)}{2\lambda_e}$ and $\varphi_c(x) = \frac{\Pi(q)}{2\lambda_c}$, we can obtain a lower bound of A that:

$$
\begin{aligned}
&\min_F \left[EU_F(q) + \lambda_e W_z\left(F, \hat{P}_e\right) + \lambda_c W_z\left(F, \hat{F}_c\right) \right] \\
&\geq \lambda_e \int_\Xi \varphi_c^1(\hat{x}_e) \hat{P}_e(d\hat{x}_e) + \lambda_c \int_\Xi \varphi_c^2(\hat{x}_c) \hat{f}_c(\hat{x}_c) d\hat{x}_c
\end{aligned} \tag{A.9}
$$

Combining Eq. (A.6) and Eq. (A.7) with Eq. (A.9), we conclude that:

$$
\begin{aligned}
G_W(\lambda_e, \lambda_c) \geq \max_{\lambda_e, \lambda_c \geq 0} - \lambda_e \rho_W^e - \lambda_c \rho_W^c + + \lambda_e \int_\Xi \min_{x \in \Xi} \left\{ d(x, \hat{x}_e) - \frac{\Pi(q)}{2\lambda_e} \right\} \hat{P}_e(d\hat{x}_e) \\
+ \lambda_c \int_\Xi \min_{x \in \Xi} \left\{ d(x, \hat{x}_c) - \frac{\Pi(q)}{2\lambda_c} \right\} \hat{f}_c(\hat{x}_c) d\hat{x}_c
\end{aligned} \tag{A.10}
$$

The problem (12) is equivalent to the following convex optimization problem:

$$
\begin{aligned}
\max_{q, \lambda_e, \lambda_c \geq 0} - \lambda_e \rho_W^e - \lambda_c \rho_W^c + \int_\Xi \min_{x \in \Xi} \left\{ \lambda_e d(x, \hat{x}_e) - \frac{\Pi(q)}{2} \right\} \hat{P}_e(d\hat{x}_e) \\
+ \int_\Xi \min_{x \in \Xi} \left\{ \lambda_c d(x, \hat{x}_c) - \frac{\Pi(q)}{2} \right\} \hat{f}_c(\hat{x}_c) d\hat{x}_c
\end{aligned} \tag{A.11}
$$

Let $u_e(q, x) = \inf_{x \in \Xi} \left[\lambda_e d(x, \hat{x}_e) - \frac{\Pi(q)}{2} \right]$, $u_c(q, x) = \inf_{x \in \Xi} \left[\lambda_c d(x, \hat{x}_c) - \frac{\Pi(q)}{2} \right]$.

We analyze the optimal solution of $u_e(q, x)$ and $u_c(q, x)$. Through observing the structure of mathematical expression $u_e(q, x)$, we explore its optimal solution in four cases (i.e., $\frac{1-\beta}{1-\alpha} q \leq \hat{x}_e$, $q \leq \hat{x}_e < \frac{1-\beta}{1-\alpha} q$, $\frac{\beta}{\alpha} q \leq \hat{x}_e < q$, and $\hat{x}_e < \frac{\beta}{\alpha} q$).

When $\frac{1-\beta}{1-\alpha} q \leq \hat{x}_e$, $u_e(q, x)$ is broken down into two parts (i.e., $\lambda_e |x - \hat{x}_e|$ and $-\frac{\Pi(q)}{2}$) for analysis, as shown in Figure A.1. We can obtain that the change of $u_e(q, x)$ function as random variable x increases, which be outlined as:

$$u_e(q,x)=\begin{cases}-\lambda_e(x-\hat{x}_e)-\frac{\lambda}{2}(\alpha x-\beta q), \hat{x}_e>\frac{\beta}{\alpha}q>x\\ -\lambda_e(x-\hat{x}_e)-\frac{1}{2}(\alpha x-\beta q), \hat{x}_e>q>x>\frac{\beta}{\alpha}q\\ -\lambda_e(x-\hat{x}_e)-\frac{1}{2}((1-\beta)q-(1-\alpha)x), \hat{x}_e>\frac{1-\beta}{1-\alpha}q>x>q\\ -\lambda_e(x-\hat{x}_e)-\frac{\lambda}{2}((1-\beta)q-(1-\alpha)x), \hat{x}_e>x>\frac{1-\beta}{1-\alpha}q\\ \lambda_e(x-\hat{x}_e)-\frac{\lambda}{2}((1-\beta)q-(1-\alpha)x), x>\hat{x}_e>\frac{1-\beta}{1-\alpha}q\end{cases}$$

Through analyzing coefficient of random variable x, we can find that the function $u_e(q,x)$ can be minimized when λ_e satisfies the condition $\lambda_e \geq \frac{\lambda}{2}(1-\alpha)$. From Figure A.1, we can find that $x=\hat{x}_e$ condition makes $u_e(q,x)$ get minimum. So $u_e^*(q,x)=-\frac{1}{2}\Pi(q)$.

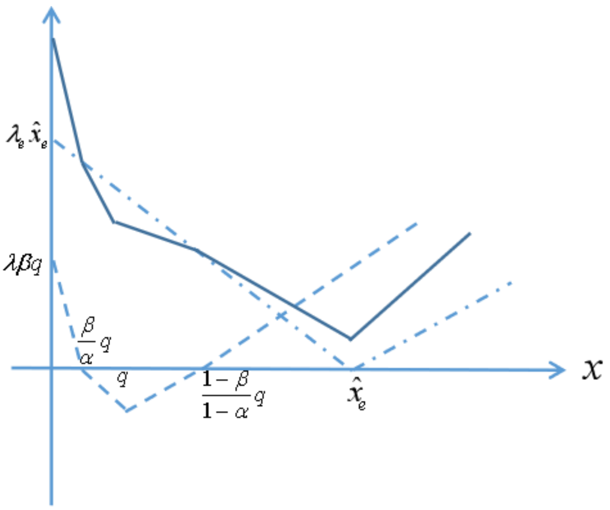


Figure A.1: The first case for function $u_e(q,x)$

Noting that dotted lines represent $\lambda_e|x-\hat{x}_e|$ and $-\frac{\Pi(q)}{2}$, respectively, and the solid line represents $u_e(q,x)$.

When $q \leq \hat{x}_e < \frac{1-\beta}{1-\alpha}q$, we can get:

$$u_e(q,x)=\begin{cases}-\lambda_e(x-\hat{x}_e)-\frac{\lambda}{2}(\alpha x-\beta q),\hat{x}_e>\frac{\beta}{\alpha}q>x\\ -\lambda_e(x-\hat{x}_e)-\frac{1}{2}(\alpha x-\beta q),\hat{x}_e>q>x>\frac{\beta}{\alpha}q\\ -\lambda_e(x-\hat{x}_e)-\frac{1}{2}((1-\beta)q-(1-\alpha)x),\frac{1-\beta}{1-\alpha}q>\hat{x}_e>x>q\\ \lambda_e(x-\hat{x}_e)-\frac{1}{2}((1-\beta)q-(1-\alpha)x),\frac{1-\beta}{1-\alpha}q>x>\hat{x}_e>q\\ \lambda_e(x-\hat{x}_e)-\frac{\lambda}{2}((1-\beta)q-(1-\alpha)x),x>\frac{1-\beta}{1-\alpha}q>\hat{x}_e\end{cases}$$

Then, we find that the function $u_e(q,x)$ can be minimized when λ_e also satisfies the condition $\lambda_e\ge\frac{(1-\alpha)}{2}$. In addition, $u_e^*(q,x)=-\frac{1}{2}\Pi(q)$ with $x=\hat{x}_e$, as shown in Figure A.2.

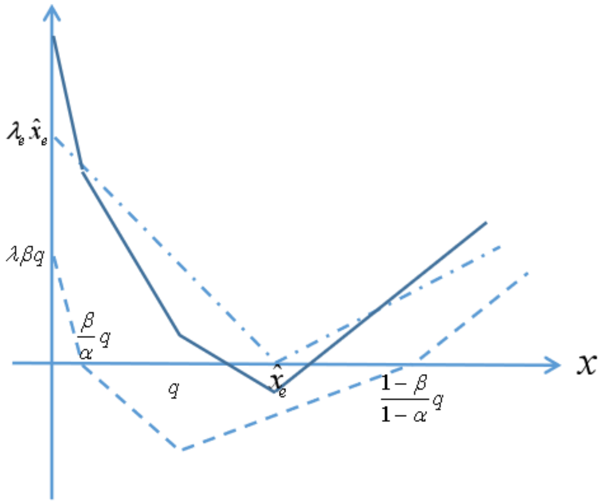


Figure A.2: The second case for function $u_e(q,x)$

When $\frac{\beta}{\alpha}q\le\hat{x}_e<q$, we attain:

$$u_e(q,x)=\begin{cases}-\lambda_e(x-\hat{x}_e)-\frac{\lambda}{2}(\alpha x-\beta q),\hat{x}_e>\frac{\beta}{\alpha}q>x\\ -\lambda_e(x-\hat{x}_e)-\frac{1}{2}(\alpha x-\beta q),q>\hat{x}_e>x>\frac{\beta}{\alpha}q\\ \lambda_e(x-\hat{x}_e)-\frac{1}{2}(\alpha x-\beta q),q>x>\hat{x}_e>\frac{\beta}{\alpha}q\\ \lambda_e(x-\hat{x}_e)-\frac{1}{2}((1-\beta)q-(1-\alpha)x),\frac{1-\beta}{1-\alpha}q>x>q>\hat{x}_e\\ \lambda_e(x-\hat{x}_e)-\frac{\lambda}{2}((1-\beta)q-(1-\alpha)x),x>\frac{1-\beta}{1-\alpha}q>\hat{x}_e\end{cases}$$

When λ_e satisfies the condition $\lambda_e\ge\frac{\alpha}{2}$, the function $u_e(q,x)$ can be minimized. In addition,

$u_e^*(q,x) = -\frac{1}{2}\Pi(q)$ with $x = \hat{x}_e$, as shown in Figure A.3.

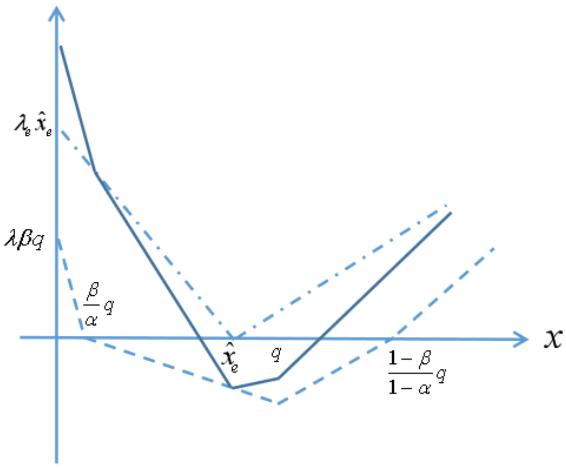


Figure A.3. The third case for function $u_e(q,x)$

When $\hat{x}_e < \frac{\beta}{\alpha}q$, we can get:

$$u_e(q,x) = \begin{cases} -\lambda_e(x-\hat{x}_e) - \frac{\lambda}{2}(\alpha x - \beta q), \frac{\beta}{\alpha}q > \hat{x}_e > x \\ \lambda_e(x-\hat{x}_e) - \frac{\lambda}{2}(\alpha x - \beta q), \frac{\beta}{\alpha}q > x > \hat{x}_e \\ \lambda_e(x-\hat{x}_e) - \frac{1}{2}(\alpha x - \beta q), q > x > \frac{\beta}{\alpha}q > \hat{x}_e \\ \lambda_e(x-\hat{x}_e) - \frac{1}{2}((1-\beta)q - (1-\alpha)x), \frac{1-\beta}{1-\alpha}q > x > q > \hat{x}_e \\ \lambda_e(x-\hat{x}_e) - \frac{\lambda}{2}((1-\beta)q - (1-\alpha)x), x > \frac{1-\beta}{1-\alpha}q > \hat{x}_e \end{cases}$$

Then, we can find that the function $u_e(q,x)$ can be minimized when λ_e also satisfies the condition $\lambda_e \geq \frac{\lambda}{2}\alpha$. In addition, $u_e^*(q,x) = -\frac{1}{2}\Pi(q)$ with $x = \hat{x}_e$, as shown in Figure A.4.

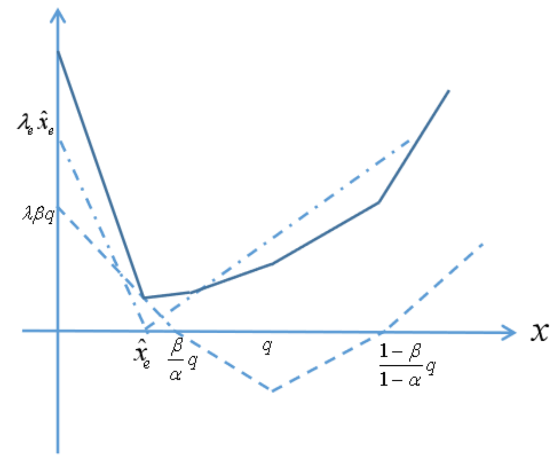


Figure A.4. The fourth case for function $u_e(q,x)$

In conclusion, $u_e^*(q,x) = -\frac{1}{2}\Pi(q)$ with $\lambda_e \geq \frac{1}{2}\max\{\lambda\alpha, \lambda(1-\alpha)\}$.

Similar to the above analysis process, we obtain that $u_c^*(q,x) = -\frac{1}{2}\Pi(q)$ with $\lambda_c \geq \frac{1}{2}\max\{\lambda\alpha, \lambda(1-\alpha)\}$.

So, the problem (A.14) can be transferred as:

$$
\begin{aligned}
&\max_{q\geq 0,\lambda_c,\lambda_e} -\lambda_c\rho_W^c - \lambda_e\rho_W^e + \frac{1}{2}EU_{\hat{F}_c}(q) + \frac{1}{2}EU_{\hat{P}_e}(q) \\
&s.t. \lambda_c \geq \frac{1}{2}\max\{\lambda\alpha, \lambda(1-\alpha)\} \\
&\quad\ \lambda_e \geq \frac{1}{2}\max\{\lambda\alpha, \lambda(1-\alpha)\}
\end{aligned}
\tag{A.15}
$$

From Eq. (A.15), the order quantities q and the dual variables (λ_e and λ_c) are separated. We can obtain the optimal dual variables $\lambda_e = \lambda_c = \frac{1}{2}\max\{\lambda\alpha, \lambda(1-\alpha)\} = \frac{\nu}{2}$. Thus, Eq. (A.15) can be rewritten as:

$$
\max_{q\geq 0} -\frac{\nu}{2}\rho_W^c - \frac{\nu}{2}\rho_W^e + \frac{1}{2}EU_{\hat{F}_c}(q) + \frac{1}{2}EU_{\hat{P}_e}(q) \tag{A.16}
$$

Q.E.D.