Accepted by *Manufacturing & Service Operations Management*
manuscript (MSOM-19-270)

This is the accepted manuscript of the following article: Liang Xu, Yi Zheng, Li Jiang (2021) A Robust Data-Driven Approach for the Newsvendor Problem with Nonparametric Information. Manufacturing & Service Operations Management 24(1):504-523, which has been published in final form at https://doi.org/10.1287/msom.2020.0961.

A Robust Data-Driven Approach for Newsvendor Problem with Non-parametric Information

Liang Xu, Yi Zheng
School of Business Administration, Southwestern University of Finance and Economics, China
Institute of Big Data, Southwestern University of Finance and Economics, China
{arecxuliang1112@swufe.edu.cn, yizheng@smail.swufe.edu.cn}

Li Jiang
Faculty of Business, The Hong Kong Polytechnic University, Hong Kong SAR, China lgtjiang@polyu.edu.hk

Problem definition. For the standard newsvendor problem with an unknown demand distribution, we develop an approach that uses data input to construct a distribution ambiguity set with the non-parametric characteristics of the true distribution, and we use it to make robust decisions.

Academic/Practical relevance. Empirical approach relies on historical data to estimate the true distribution. Although the estimated distribution converges to the true distribution, its performance with limited data is not guaranteed. Our approach generates robust decisions from a distribution ambiguity set that is constructed by data-driven estimators for non-parametric characteristics and includes the true distribution with the desired probability. It fits the situations where data size is small.

Methodology. We apply robust optimization approach with non-parametric information.

Results. Under a fixed method to partition the support of the demand, we construct a distribution ambiguity set, build a protection curve as the proxy for the worst-case distribution in the set, and use it to obtain a robust stocking quantity in closed form. Implementation-wise, we develop an adaptive method to continuously feed data to update partitions with a prespecified confidence level in their unbiasedness and adjust the protection curve to obtain robust decisions. We theoretically and experimentally compare the proposed approach with existing approaches.

Managerial Insights. Our non-parametric approach under adaptive partitioning guarantees that the realized average profit exceeds the worst-case expected profit with a high probability. Using real data sets from Kaggle.com, it can outperform existing approaches in yielding profit rate and stabilizing the generated profits, and the advantages are more prominent as service ratio decreases. Non-parametric information is more valuable than parametric information in profit generation provided that the service requirement is not too high. Moreover, our proposed approach provides a means of combining non-parametric and parametric information in a robust optimization framework.

Key words: Robust optimization; newsvendor; non-parametric information; data-driven decisions

1. Introduction

The newsvendor problem is the foundation to a class of operations and supply chain models. In its standard version, a firm stocks a product before knowing the actual demand and satisfies the realized demand up to stock availability. The firm incurs a marginal cost c and receives a revenue

$p > c$ for each product sold, while unsold products have no salvage value (Natarajan et al. 2018). The profit function is $\Pi(q) = E[p\min\{Z,q\} - cq]$, where Z is the random demand with distribution function $F(\cdot)$. The optimal stocking quantity is the μ_0-quantile of $F(\cdot)$, where $\mu_0 \triangleq 1 - \frac{c}{p}$ is the service ratio. In practice, the true distribution function is unknown and decision makers usually assume some distribution as a rule of thumb. Significant differences can exist between the assumed and the true distributions, jeopardizing the applicability of the obtained policies. A distribution-free newsvendor problem has been introduced to tackle distribution uncertainty. Scarf (1958) derives the stocking quantity that maximizes the minimum expected newsvendor profit when the distribution satisfies given mean and variance. Lariviere and Porteus (1999) adopt a Bayesian approach to dynamically learn the underlying distribution. The performance of these approaches depends on the accuracy of parameter estimates and is therefore not guaranteed to be robust.

Against the backdrop of technology advances that have significantly increased computing power, firms face the challenge of building innovative models to capture and utilize data in decision making. For the newsvendor problem, a class of approaches rely on historical data to estimate the density curve. Rosenblatt (1956) introduces kernel estimation for distribution characteristics. Common kernel functions are uniform and Gaussian. Ruppert and Cline (1994) modify kernel density estimation by transformed data. Huh and Rusmevichientong (2009) propose non-parametric adaptive policies to generate decisions. Despite convergence of the estimated density curve, its performance is not guaranteed when data size is small because the estimated density curve adjusts to data input, varying the stocking quantity and realized profit.

We propose an approach that utilizes the non-parametric characteristics of the distribution – including its support, monotonicity and convexity of its density curve, and specific density values – to partition the support, construct a distribution ambiguity set that includes the true distribution, and build a protection curve to approximate the true density curve. Among the partitioning methods, monotone (full) partitioning, whereby the density curve in a partition is non-decreasing or non-increasing (monotone convex or concave), requires the smallest (largest) set of non-parametric characteristics to construct the distribution ambiguity set, but yields the lowest (highest) worst-case expected profit.

Our main contributions can be summarized as follows. First, to the best of our knowledge, our proposed approach is the first of its kind to use non-parametric characteristics of the distribution to form the distribution ambiguity set. This differs from the previous approaches that use parametric information – including mean and variance – for the same purpose. The protection curve, which serves as the proxy for the worst-case distribution in the distribution ambiguity set, approximates the true density curve and enables us to obtain closed-form expressions for robust stocking quantities. The generated policies are stable and yield less variable realized profits.

Second, we use empirical distribution to form non-parametric estimators to specify partitions, construct distribution ambiguity sets, and build protection curves. We analyze convergence properties of the estimators, investigate the stability of non-parametric characteristics in the formed partitions with increments in data, and develop the criteria for trusting a partitioning method for an interval such that the true distribution restricted on this interval is guaranteed to have the desired non-parametric characteristics with the prespecified confidence level. Our approach feeds data input to adaptively construct the distribution ambiguity set that includes the true distribution with the desired probability, and it yields decision policies with a guaranteed likelihood that the realized average profit exceeds the worst-case expected profit.

Third, we identify a data-size threshold to guide our reliance on empirical distribution or protection curve in data-driven decision models. Specifically, we can trust the distribution ambiguity set based directly on empirical distribution to make robust decisions when data size exceeds the threshold, but trust the distribution ambiguity set based on non-parametric estimators otherwise. This balances the advantage of empirical distribution for its convergence (for a large data size) and that of the non-parametric approach for its stability (for a small data size).

Lastly, for the standard newsvendor problem, we make use of real data sets to show that our non-parametric approach can outperform existing approaches to improve profit rate and stabilize profit generation, and these advantages become more prominent as the service ratio decreases. Non-parametric information is more valuable than its parametric counterpart in yielding the expected profit provided that the service requirement is not too high. Our approach has further advantages over parametric approaches in worst-case performance guarantee with censored data and faster convergence of the data-driven estimators used in the approach. Adding non-parametric characteristics to parametric information to form distribution ambiguity sets by our partitioning method and derive robust decisions can improve the expected profit performance.

The remainder of this paper is organized as follows. Section 2 provides a review of the relevant literature. Section 3 introduces the partitioning method based on non-parametric characteristics, defines distribution ambiguity set and protection curve, and presents the estimators for non-parametric characteristics. Section 4 applies the non-parametric approach to derive robust policies and presents an adaptive partitioning method in implementation. Section 5 studies the value of non-parametric information relative to parametric information, with the worst-case expected profit as the measure. Section 6 presents a series of experiments that justify the effectiveness and practicality of the non-parametric approach. Section 7 concludes the paper with discussions. All the proofs are presented in the Appendix.

2. Literature Review

The basic objective in the standard newsvendor problem is to find the stocking quantity that balances overstocking and understocking risks. The existing research has adopted two classes of approaches to tackle this problem with an unknown demand distribution. Scarf (1958) derives the stocking quantity that maximizes the minimum expected profit for distributions that satisfy given mean and variance. This ushers in a class of parametric approaches to solve distributionally robust newsvendor problems. These approaches typically construct an uncertainty set that includes distributions whose moments or other parameters satisfy certain conditions (e.g. Ben-Tal and Hochman 1976, Popescu and Wu 2007, Zymler et al. 2013). However, it has two main disadvantages. One is that the worst-case distribution can yield too conservative a decision (e.g. Gallego and Moon 1993, Goh and Sim 2010, Wang et al. 2016, Zhu et al. 2013). The other is that the moments are unable to accurately capture the true distribution with limited data input and can be biased with censored data. Wang et al. (2016) identify a problem instance in which Scarf's approach performs less well under exponential distribution. To alleviate the conservatism, some approaches include the information on higher-order moments (Zuluaga et al. 2009). However, closed-form solutions are hard to obtain even for computationally tractable single-item newsvendor problems.

Perakis and Roels (2008) include partial information and derive the stocking quantity that minimizes the newsvendor's maximum regret of not acting optimally. Levi et al. (2011) solve the min-max regret problem with information on the absolute mean spread. Yue et al. (2006) compute the maximum expected value of distribution information (EVDI) that satisfies parametric conditions for a given quantity and introduce a procedure to obtain the quantity that minimizes the maximum EVDI. Natarajan et al. (2018) divide the support into multiple partitions and introduce a second-order statistic on the partitioned demand distribution to capture its asymmetry. As a special case, the second-order statistic reflects semi-variance when the support is divided into two partitions, one before and the other after the demand mean. Experiments show that symmetry information can reduce the expected profit loss, particularly when the true distribution is heavily tailed. We develop the criteria for choosing partitioning methods on the support. Importantly, our approach builds a premise of combining parametric information and non-parametric characteristics to improve expected profit performance.

The second class of approaches make no assumption on the parametric forms of the true distribution. Godfrey and Powell (2001) use a concave adaptive value estimation algorithm to estimate a piecewise-linear concave function for the expected profit based on the information of remaining inventories. Levi et al. (2007) adopt a sampling-driven algorithmic framework to compute policies in single- and multi-period newsvendor problems. Considering distributional robustness, a standard means of constructing ambiguity set is to choose a statistical distance such as ϕ-divergence

(Bayraksan and Love 2015, Jiang and Guan 2016, Sun and Xu 2016) and Wasserstein distance (Mohajerin Esfahani and Kuhn 2018, Zhao and Guan 2018, Gao and Kleywegt 2016), and find those distributions that are close to a nominal distribution like empirical distribution. However, ϕ-divergence is not rich enough to include relevant distributions because it focuses only on bin-by-bin comparison and fails to capture (dis)similarity across points (Rubner et al. 2000). Gao and Kleywegt (2016) consider distributionally robust stochastic optimization (DRSO) with Wasserstein distance and characterize conditions for the existence of worst-case distribution whereby data-driven DRSO problems are tractable. The optimal Wasserstein radius cannot be computed exactly and estimators have direct effects on performances. Mohajerin Esfahani and Kuhn (2018), for instance, use posteriori information to estimate the radius in a data-driven portfolio problem. Wang et al. (2016) present another attempt to allow distributions to deviate from empirical distribution with certain likelihood and derive the optimal decision to maximize the worst-case expected profit. Our approach is much more general for distribution-free newsvendor problems and performs well in practical situations.

As an effort to extend to multi-item newsvendor problems, Gallego and Moon (1993) apply Scarf's bound to the worst-case expected profits of individual items and apply convex optimization with known mean and variance. Ben-Tal et al. (2013) apply ϕ-divergence distance (e.g. chi-squared divergence; Kullback-Leibler divergence) in a multi-item newsvendor problem with information on marginal distribution. These models do not consider demand correlation. Hanasusanto et al. (2015) present a risk-averse multi-item model with information on mean and covariance, and the demand distribution is known to be a mixture of distinct distributions. Natarajan et al. (2018) consider partitioned statistics in a multi-item setting. Our paper provides both analytical and experimental evidence for the value of including non-parametric characteristics to parametric information to derive robust decisions, and explores newsvendor problems with multiple items.

3. Distribution Ambiguity Set, Protection Curve and Non-parametric Estimators

For decision models that have a random variable with an unknown distribution, we propose a data-driven approach that uses non-parametric characteristics of the distribution to construct a distribution ambiguity set that includes all the distributions with the desired characteristics. Based on the constructed distribution ambiguity set, we build a protection curve to serve as a proxy for the worst-case distribution in the set and approximates the true density curve. Our presentation is focused on the situation where the random variable is continuous and has a density curve on a continuous support. The definitions, approaches, and discussions can be modified to the situation where the random variable is discrete and has a mass function on a discrete support.

3.1. Distribution ambiguity set and protection curve

Consider a random variable Z on $[\underline{z},\overline{z}]$. The distribution function is $F(z) \triangleq \mathbb{P}(Z \le z)$, the density function is $f^{(0)}(z) \triangleq \frac{d\mathbb{P}(Z\le z)}{dz}$, and the first- and second-order derivatives of the density function are $f^{(1)}(z) \triangleq \frac{d^2\mathbb{P}(Z\le z)}{dz^2}$ and $f^{(2)}(z) \triangleq \frac{d^3\mathbb{P}(Z\le z)}{dz^3}$. The following Definition 1 defines the protection curve.

Definition 1. A function $c(z)$ is a protection curve for the true density curve $f^{(0)}(z)$ if $\int_y^{\overline{z}} \left[f^{(0)}(z) - c(z)\right] dz \ge 0, \forall y \in [\underline{z},\overline{z}]$.

Proposition 1 presents the property of the protection curve that forms the basis of constructing distribution ambiguity set.

Proposition 1. *Given density curve $f^{(0)}(z)$ of a continuous random variable Z on $[\underline{z},\overline{z}]$ and the protection curve $c(z)$, $\int_{\underline{z}}^{\overline{z}} \min\{z,q\} f^{(0)}(z)\, dz \ge \int_{\underline{z}}^{\overline{z}} \min\{z,q\}\, c(z)\, dz$, where q is stocking quantity.*

When the protection curve is a density curve, i.e., $\int_{\underline{z}}^{\overline{z}} c(z)\, dz = 1$, the underlying random variable is first-order stochastically dominated by the random variable that follows the true distribution. For a given stocking quantity, the expected sales quantity under the true distribution is no less than that under its protection curve, which provides a conservative sales guarantee.

Definition 2. If a density curve $f^{(0)}(z)$ of a continuous random variable Z on $[\underline{z},\overline{z}]$ satisfies $\int_y^{\overline{z}} \left[f^{(0)}(z) - c(z)\right] dz \ge 0,\ \forall y \in [\underline{z},\overline{z}]$, then $f^{(0)}(z)$ belongs to the distribution ambiguity set, for which $c(z)$ serves as a proxy for the worst-case distribution in generating the expected profit.

We construct a distribution ambiguity set to include all the distributions that share the desired non-parametric characteristics with the true distribution, and use the constructed distribution ambiguity set to build the protection curve as a safe approximation to the true density curve.

3.2. Partitioning methods

We divide the support $[\underline{z},\overline{z}]$ for the random variable into mutually exclusive partitions $\bigcup_{i=1}^{m}[\underline{z}_i,\overline{z}_i]$ by one of three methods, which we call monotone, full, and semi-full partitioning. Each method entails certain non-parametric characteristics of the distribution to define partitions. We construct the distribution ambiguity set and build the protection curve for each partition, and unionize them to construct the distribution ambiguity set and build the protection curve on the entire support.

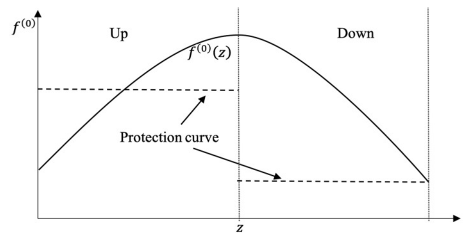
Monotone partitioning

We rely on the monotonicity of the density curve as the main non-parametric characteristic to partition the support. The true density curve in a partition is either non-increasing or non-decreasing. Figure 1 illustrates protection curves under monotone partitioning. Given a partition $[\underline{z}_i,\overline{z}_i]$, where the true density curve is non-increasing and has the lowest density $l_i \triangleq \min_{z\in[\underline{z}_i,\overline{z}_i]} f^{(0)}(z)$, we construct the distribution ambiguity set as follows:

$$\mathcal{P}_{[\underline{z}_i,\overline{z}_i]} = \left\{ \mathbb{P} : f^{(1)}(z) \le 0,\ \min_{z\in[\underline{z}_i,\overline{z}_i]} f^{(0)}(z) = l_i,\ z \in [\underline{z}_i,\overline{z}_i] \right\}, \tag{1}$$

where $f^{(1)}(z) \le 0$ ensures that the density curve is non-increasing and $\min_{z\in[\underline{z}_i,\overline{z}_i]} f^{(0)}(z) = l_i$ ensures that l_i is a tight lower bound to the density. By Definition 1, we verify that $c(z) = l_i$ is a protection curve, which assumes a constant value equal to the lowest density in the non-increasing partition.

Figure 1 **Protection curves under monotone partitioning**



Given a partition $[\underline{z}_i, \overline{z}_i]$ where the true density curve is non-decreasing, we construct the distribution ambiguity set as follows:

$$\mathcal{P}_{[\underline{z}_i,\overline{z}_i]} = \left\{\mathbb{P} : \mathbb{P}\left\{\underline{z}_i \le z \le \overline{z}_i\right\} = P_i, f^{(1)}(z) \ge 0,\ z \in [\underline{z}_i, \overline{z}_i]\right\}, \tag{2}$$

where P_i is the cumulative probability that the random variable assumes a value in this partition and $f^{(1)}(z) \ge 0$ ensures that the density curve is non-decreasing. In this case, we can verify that $c(z) = \frac{P_i}{\overline{z}_i - \underline{z}_i}$ is a protection curve, which assumes a constant value equal to the average density in the non-decreasing partition.

Full partitioning

The protection curve under monotone partitioning is conservative because the method takes the lowest density in a non-increasing partition and the average density in a non-decreasing partition to build protection curves. This to a large extent leads to an underestimate of the chance of occurrence for demand realization when we apply the protection curve to approximate the true density curve. To resolve this conservatism, we adopt full partitioning to refine, by incorporating more non-parametric characteristics, the distribution ambiguity set to build a protection curve that better fits the true density curve. Specifically, we refine a monotone partition into convex or concave subpartitions. Consider a partition $[\underline{z}_i, \overline{z}_i]$. Let the cumulative probability in the partition be $P_i = \mathbb{P}\left\{\underline{z}_i \le z \le \overline{z}_i\right\}$, and the maximum and minimum density values be $u_i = \max_{z\in[\underline{z}_i,\overline{z}_i]} f^{(0)}(z)$ and $l_i = \min_{z\in[\underline{z}_i,\overline{z}_i]} f^{(0)}(z)$, respectively. In what follows, we construct the distribution ambiguity set which depends on the specific shape of the density curve in the partition.

For a non-decreasing and convex partition $[\underline{z}_i, \overline{z}_i]$:

$$\mathcal{P}_{[\underline{z}_i,\overline{z}_i]} = \left\{\mathbb{P} : \mathbb{P}\left\{\underline{z}_i \le z \le \overline{z}_i\right\} = P_i, \min_{z\in[\underline{z}_i,\overline{z}_i]} f^{(0)}(z) = l_i, f^{(1)}(z) \ge 0, f^{(2)}(z) \ge 0\right\}. \tag{3}$$

For a non-increasing and convex partition $[\underline{z}_i, \overline{z}_i]$:

$$\mathcal{P}_{[\underline{z}_i,\overline{z}_i]} = \left\{\mathbb{P} : \mathbb{P}\left\{\underline{z}_i \le z \le \overline{z}_i\right\} = P_i, \max_{z\in[\underline{z}_i,\overline{z}_i]} f^{(0)}(z) = u_i, f^{(1)}(z) \le 0, f^{(2)}(z) \ge 0\right\}. \tag{4}$$

For a non-decreasing and concave partition $[\underline{z}_i, \overline{z}_i]$:

$$\mathcal{P}_{[\underline{z}_i,\overline{z}_i]} = \left\{ \mathbb{P} : \mathbb{P}\left\{\underline{z}_i \leq z \leq \overline{z}_i\right\} = P_i, \max_{z\in[\underline{z}_i,\overline{z}_i]} f^{(0)}(z) = u_i, f^{(1)}(z) \geq 0, f^{(2)}(z) \leq 0 \right\}. \tag{5}$$

For a non-increasing and concave partition $[\underline{z}_i, \overline{z}_i]$:

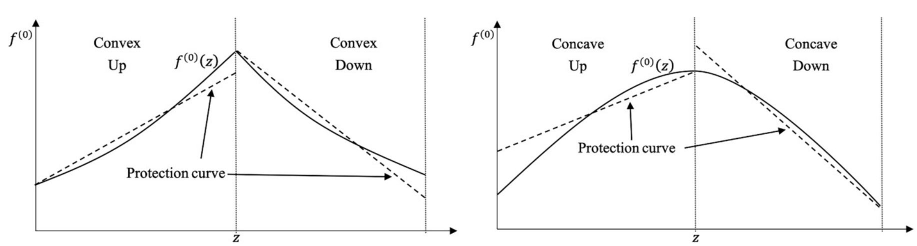
$$\mathcal{P}_{[\underline{z}_i,\overline{z}_i]} = \left\{ \mathbb{P} : \mathbb{P}\left\{\underline{z}_i \leq z \leq \overline{z}_i\right\} = P_i, \min_{z\in[\underline{z}_i,\overline{z}_i]} f^{(0)}(z) = l_i, f^{(1)}(z) \leq 0, f^{(2)}(z) \leq 0 \right\}. \tag{6}$$

The distributions in the distribution ambiguity sets defined in (3)-(6) satisfy the requirements on the cumulative probability, maximum or minimum density value, monotonicity and convexity or concavity of the density curve in respective partitions. The protection curve takes a general form of $c(z) = \alpha_i + \beta_i (z - \underline{z}_i)$, $z \in [\underline{z}_i, \overline{z}_i]$, where coefficients (α_i, β_i) depend on the shape of the density curve and are presented in Table 1. Figure 2 illustrates protection curves under full partitioning. Observe that the density curve in a partition is approximated by a line which intersects with the true curve at its maximum or minimum value and shares the same cumulative probability. Thus, the protection curve is close to the true density curve, whose first-order dominance is ensured.

Table 1 Coefficients for protection curve $c(z) = \alpha_i + \beta_i (z - \underline{z}_i)$.

partition $z \in [\underline{z}_i, \overline{z}_i]$	α_i	β_i
non-decreasing and convex	l_i	$2\frac{P_i - l_i(\overline{z}_i - \underline{z}_i)}{(\overline{z}_i - \underline{z}_i)^2}$
non-increasing and convex	u_i	$-2\frac{u_i(\overline{z}_i - \underline{z}_i) - P_i}{(\overline{z}_i - \underline{z}_i)^2}$
non-decreasing and concave	$\frac{2P_i}{\overline{z}_i - \underline{z}_i} - u_i$	$2\frac{u_i(\overline{z}_i - \underline{z}_i) - P_i}{(\overline{z}_i - \underline{z}_i)^2}$
non-increasing and concave	$\frac{2P_i}{\overline{z}_i - \underline{z}_i} - l_i$	$-2\frac{P_i - l_i(\overline{z}_i - \underline{z}_i)}{(\overline{z}_i - \underline{z}_i)^2}$

Figure 2 Protection curves under full partitioning



Semi-full partitioning

A third alternative for partitioning the support is semi-full partitioning, which combines monotone partitioning and full partitioning. Specifically, the support is divided into non-decreasing monotone partitions and non-increasing full partitions. The distribution ambiguity set for a non-decreasing partition is defined in (2) and those for non-increasing convex and concave partitions are defined in (4) and (6), respectively.

3.3. Estimators for non-parametric characteristics

We are unable to directly partition the support and build the protection curve because the true distribution is unknown. Instead, we utilize available data to estimate the distribution function, density function, and its first- and second-order derivatives, and use these non-parametric estimators to define partitions, cumulative probabilities, maximum and minimum density values.

3.3.1. Data-driven non-parametric estimators

Let the available data $(d_1, \cdots, d_n)$ be independently and identically drawn from a random variable Z with CDF $F(z)$. The empirical distribution function is defined as $\hat{F}(z) \triangleq \frac{\sum_{i=1}^{n} \mathbb{I}\{d_i \le z\}}{n}$, where $\mathbb{I}\{d_i \le z\}$ is the indicator function that assumes one if observation d_i is less than or equal to z but zero otherwise. The empirical density function is $\hat{f}^{(0)}(z) \triangleq \frac{\hat{F}(z+\lambda)-\hat{F}(z-\lambda)}{2\lambda}$, where $\lambda > 0$ is the half width. Both $\hat{F}(z)$ and $\hat{f}^{(0)}(z)$ are stepwise functions. The empirical first- and second-order derivatives are, respectively, $\hat{f}^{(1)}(z) \triangleq \frac{\hat{f}^{(0)}(z+\lambda)-\hat{f}^{(0)}(z-\lambda)}{2\lambda}$ and $\hat{f}^{(2)}(z) \triangleq \frac{\hat{f}^{(1)}(z+\lambda)-\hat{f}^{(1)}(z-\lambda)}{2\lambda}$, which are stepwise functions assuming constants in consecutive intervals. Note that $\hat{F}(z), \hat{f}^{(0)}(z), \hat{f}^{(1)}(z)$, and $\hat{f}^{(2)}(z)$ are estimated on $(-\infty, +\infty)$. The support $[\underline{z}, \overline{z}]$ is estimated to be the maximum interval such that $\hat{f}^{(0)}(z) = 0$ for $z < \underline{z}$ and $z > \overline{z}$.

An interval $[a,b] \subseteq [\underline{z}, \overline{z}]$ is a monotone partition when $\hat{f}^{(1)}(z_1)\hat{f}^{(1)}(z_2) \ge 0, \forall z_1, z_2 \in [a,b]$. A monotone partition $[a,b]$ can be refined into a full partition when $\hat{f}^{(2)}(z_1)\hat{f}^{(2)}(z_2) \ge 0, \forall z_1, z_2 \in [a,b]$. The estimator $\hat{P}$ for the cumulative probability in the partition $\mathbb{P}\{a \le z \le b\}$ is $\hat{P} = \hat{F}(b) - \hat{F}(a)$. The estimators $\hat{l}$ and $\hat{u}$ for the maximum and minimum density values in $[a,b]$ are, respectively, $\hat{l} = \min_{z\in[a,b]} \hat{f}^{(0)}(z)$ and $\hat{u} = \max_{z\in[a,b]} \hat{f}^{(0)}(z)$.

3.3.2. Convergence properties of estimators

Next, we discuss the convergence properties of estimators $\hat{F}(z)$ and $\hat{f}^{(i)}(z), i = 0,1,2$. Obviously, the indicator function $\mathbb{I}\{d_i \le z\}$ follows a Bernoulli distribution with parameter $F(z)$, and:

$$\sqrt{n}\left[\hat{F}(z) - F(z)\right] \xrightarrow{d} N(0, F(z)(1-F(z))), \tag{7}$$

where $\xrightarrow{d}$ indicates convergence in distribution. As proved in the Appendix, the asymptotical distributions for $\hat{f}^{(i)}(z), i = 0,1,2$ are: $\sqrt{n}\left[\hat{f}^{(i)}(z) - f^{(i)}(z)\right] \xrightarrow{d} N(0, \sigma_i)$, where $\sigma_0 \le \frac{1}{2\sqrt{2}\lambda}, \sigma_1 \le \frac{1}{(2\lambda)^2}$ and $\sigma_2 \le \frac{1}{4\sqrt{2}(\lambda)^3}$. With confidence level $1-\omega$, the pointwise estimate error of $\hat{F}(z)$ relative to $F(z)$ and that of $\hat{f}^{(i)}(z)$ relative to $f^{(i)}(z)$ are, respectively, as follows:

$$\mathbb{P}\{|\hat{F}(z) - F(z)| > g^P(n,z,\omega)\} \le \omega; \mathbb{P}\left\{\left|\hat{f}^{(i)}(z) - f^{(i)}(z)\right| > g_i^P(n,z,\omega)\right\} \le \omega, i = 0,1,2, \tag{8}$$

where $g^P(n,z,\omega) = \frac{\sigma\varphi^{-1}(1-\frac{\omega}{2})}{\sqrt{n}}$, $g_i^P(n,z,\omega) = \frac{\sigma_i\varphi^{-1}(1-\frac{\omega}{2})}{\sqrt{n}}$, $\varphi^{-1}(\cdot)$ is the inverse cumulative function of $N(0,1)$, and $\sigma = \sqrt{F(z)(1-F(z))}$.

Uniform estimate errors of $\hat{f}^{(i)}(z)$ relative to $f^{(i)}(z)$ in any interval $[a,b]\subseteq[\underline{z},\overline{z}]$ are:

$$\mathbb{P}\left\{\max_{z\in[a,b]}\left|\hat{f}^{(i)}(z)-f^{(i)}(z)\right|>g_i(n,\omega)\right\}\leq\omega, i=0,1,2, \tag{9}$$

where $g_0(n,\omega)=\frac{D_n(\omega)}{\lambda}+\frac{\lambda K_0}{2}$, $g_1(n,\omega)=\frac{D_n(\omega)}{\lambda^2}+\frac{2\lambda K_1}{3}$, $g_2(n,\omega)=\frac{D_n(\omega)}{\lambda^3}+\frac{7\lambda K_2}{8}$, K_i is the upper bound for $\left|f^{(i+1)}(z)\right|$, and $D_n(\omega)$ is the Kolmogorov-Smirnov statistic in Waterman and Whiteman (1978). It can be verified that $g_i^p(n,z,\omega)$ decreases with data size and has a convergence ratio of $O(\frac{1}{\sqrt{n}})$, while $g_i(n,\omega)$ comprises two parts: one part decreases with data size at a convergence ratio of $O(\frac{1}{\sqrt{n}})$ and the other part is independent of data size and non-decreases with half width.

For an unbiased partition $[a,b]$, the estimator $\hat{P}$ for cumulative probability is based on $\hat{F}(b)$ and $\hat{F}(a)$ and hence has the same convergence ratio as $\hat{F}(z)$, while the estimators $\hat{l}$ and $\hat{u}$ for the minimum and maximum density values are based on $\hat{f}^{(0)}(a)$ or $\hat{f}^{(0)}(b)$, depending on the monotonicity of $[a,b]$, and hence have the same convergence ratio as $\hat{f}^{(0)}(z)$. We discuss the biasedness of the partitions formed by non-parametric estimators and the means of handling it in Section 4.2.

3.3.3. Stability of non-parametric characteristics

Let $\hat{F}_n(z)$ and $\hat{f}_n^{(0)}(z)$ be the empirical distribution function and empirical density function with data size n. Recall that they are piecewise linear functions assuming constants in consecutive intervals. After a new data $d_{n+1}\in[\underline{z},\overline{z}]$ is revealed, they are updated to:

$$\hat{F}_{n+1}(z)=\begin{cases}\frac{n\hat{F}_n(z)}{n+1} & z\in[\underline{z},d_{n+1})\\ \frac{n\hat{F}_n(z)+1}{n+1} & z\in[d_{n+1},\overline{z}]\end{cases}, \hat{f}_{n+1}^{(0)}(z)=\begin{cases}\frac{n\hat{f}_n^{(0)}(z)}{n+1} & z\in[\underline{z},d_{n+1}-\lambda)\cup[d_{n+1}+\lambda,\overline{z}]\\ \frac{2n\lambda\hat{f}_n^{(0)}(z)+1}{(2\lambda)(n+1)} & z\in[d_{n+1}-\lambda,d_{n+1}+\lambda)\end{cases}.$$

In addition to a scale adjustment of $\frac{n}{n+1}$ due to an increase in data size, the empirical distribution function can shift by a factor of $\frac{1}{n+1}$, and the empirical density function can shift by a factor of $\frac{1}{2\lambda(n+1)}$. The empirical first- and second-order derivatives are updated to:

$$\hat{f}_{n+1}^{(1)}(z)=\begin{cases}\frac{n\hat{f}_n^{(1)}(z)}{n+1} & z\in[\underline{z},d_{n+1}-2\lambda)\cup[d_{n+1}+2\lambda,\overline{z}]\\ \frac{4n\lambda^2\hat{f}_n^{(1)}(z)+1}{(2\lambda)^2(n+1)} & z\in[d_{n+1}-2\lambda,d_{n+1})\\ \frac{4n\lambda^2\hat{f}_n^{(1)}(z)-1}{(2\lambda)^2(n+1)} & z\in[d_{n+1},d_{n+1}+2\lambda)\end{cases}, \text{and}$$

$$\hat{f}_{n+1}^{(2)}(z)=\begin{cases}\frac{n\hat{f}_n^{(2)}(z)}{n+1} & z\in[\underline{z},d_{n+1}-3\lambda)\cup[d_{n+1}+3\lambda,\overline{z}]\\ \frac{8n\lambda^3\hat{f}_n^{(2)}(z)+1}{(2\lambda)^3(n+1)} & z\in[d_{n+1}-3\lambda,d_{n+1}-\lambda)\cup[d_{n+1}+\lambda,d_{n+1}+3\lambda)\\ \frac{8n\lambda^3\hat{f}_n^{(2)}(z)-2}{(2\lambda)^3(n+1)} & z\in[d_{n+1}-\lambda,d_{n+1}+\lambda)\end{cases}.$$

Lemma 1 presents preliminary properties of the empirical first- and second-order derivatives.

LEMMA 1. *Given data size n, $\left|\hat{f}_n^{(1)}(z)\right|\geq\frac{1}{(2\lambda)^2 n}$ for any z with $\left|\hat{f}_n^{(1)}(z)\right|>0$, and $\left|\hat{f}_n^{(2)}(z)\right|\geq\frac{1}{(2\lambda)^3 n}$ for any z with $\left|\hat{f}_n^{(2)}(z)\right|>0$.*

Hence, whenever the empirical first-order (second-order) derivative is non-zero, its norm is bounded from below by $\frac{1}{(2\lambda)^2 n}$ ($\frac{1}{(2\lambda)^3 n}$), which depends on and decreases with data size and is scaled by half width. Recall that, given data size n, an interval $[a,b]$ formed by non-parametric estimators is a monotone partition when $\hat{f}_n^{(1)}(z_1)\,\hat{f}_n^{(1)}(z_2) \geq 0, \forall z_1, z_2 \in [a,b]$. Proposition 2 reveals how a new data input influences the monotonicity of the formed partitions.

PROPOSITION 2. *Let $B = \cup_{i=1}^{s} B_i$, where $B_i = \left\{z \in B_i \subseteq [\underline{z}, \overline{z}] : \hat{f}_n^{(1)}(z) = 0\right\}$ belongs to a monotone partition with data size n. As a new data d_{n+1} is revealed and the empirical density curve is updated from $\hat{f}_n^{(0)}(z)$ to $\hat{f}_{n+1}^{(0)}(z)$, the monotonicity of $[a,b]$, which is non-increasing (non-decreasing) under $\hat{f}_n^{(0)}(z)$, will change if $[a,b] \cap B \neq \emptyset$ and there exists a $z \in [a,b] \cap B$ such that $\hat{f}_{n+1}^{(1)}(z) > 0$ ($\hat{f}_{n+1}^{(1)}(z) < 0$); otherwise, the monotonicity of the partition will not change.*

Similarly, Proposition 3 reveals how a new data input influences the convexity of partitions, where a partition $[a,b]$ formed by non-parametric estimators is convex or concave when $\hat{f}_n^{(2)}(z_1)\,\hat{f}_n^{(2)}(z_2) \geq 0, \forall z_1, z_2 \in [a,b]$.

PROPOSITION 3. *Let $C = \cup_{i=1}^{s} C_i$, where $C_i = \left\{z \in C_i \subseteq [\underline{z}, \overline{z}] : \hat{f}_n^{(2)}(z) = 0\right\}$ belongs to a convex or concave partition with data size n. As a new data d_{n+1} is revealed and the empirical density curve is updated from $\hat{f}_n^{(0)}(z)$ to $\hat{f}_{n+1}^{(0)}(z)$:*

The convexity of a partition $[a,b]$ under $\hat{f}_n^{(0)}(z)$ will change if 1) $[a,b] \cap \{C \setminus [d_{n+1}-\lambda, d_{n+1}+\lambda)\} \neq \emptyset$ and there exists a $z \in [a,b] \setminus [d_{n+1}-\lambda,\ d_{n+1}+\lambda) \cap C$ such that $\hat{f}_{n+1}^{(2)}(z) < 0$; or 2) $[a,b] \cap [d_{n+1}-\lambda, d_{n+1}+\lambda) \neq \emptyset$ and there exists a $z \in (a,b) \cap [d_{n+1}-\lambda, d_{n+1}+\lambda)$ such that $\hat{f}_n^{(2)}(z) < \frac{2}{(2\lambda)^3 n}$; otherwise, the convexity of the partition will not change.

The concavity of a partition $[a,b]$ under $\hat{f}_n^{(0)}(z)$ will change if $[a,b] \cap C \neq \emptyset$ and there exists a $z \in [a,b] \cap C$ such that $\hat{f}_{n+1}^{(2)}(z) > 0$; otherwise, the concavity of the partition will not change.

Propositions 2 and 3 provide sufficient and necessary conditions for the characteristics of the partitions formed by non-parametric estimators to adjust with data input. Together with the updating procedure, they send the message that our data-driven approach provides a stable instrument to absorb data input and partition the support. Recall that, after a new data d_{n+1} is revealed, the empirical distribution function on the entire support is updated either by a scale adjustment or by a scale adjustment and a shift factor. However, the monotonicity of an interval remains unaffected if the empirical first-order derivative in it is updated by a scale adjustment only. Our results show that the monotonicity changes only when the empirical first-order derivative assumed a value of zero in $[d_{n+1}-2\lambda, d_{n+1}+2\lambda)$, and is updated by both a scale adjustment and a shift factor. Similarly, the convexity of an interval remains unaffected if the empirical second-order derivative in it is updated by a scale adjustment only. The convexity changes only when the empirical second-order derivative assumed a value of zero in $[d_{n+1}-3\lambda, d_{n+1}-\lambda) \cup [d_{n+1}+\lambda, d_{n+1}+3\lambda)$ or assumed a very small value in $[d_{n+1}-\lambda, d_{n+1}+\lambda)$, and is updated by both a scale adjustment and a shift factor.

4. Closed-form Robust Policies and Implementation

We first apply the non-parametric approach under given partitioning method to obtain closed-form stocking quantities that maximize the worst-case expected profit. We then explore the biasedness of the partitions formed by non-parametric estimators with available data and develop an adaptive procedure to build proper protection curves and obtain robust decisions.

4.1. Fixed partitioning

The standard newsvendor problem finds the stocking quantity q that maximizes the expected profit $\Pi(q) = E[\pi(q,Z)]$, where $\pi(q,Z) = p\min\{Z,q\} - cq$. The random demand Z is continuous and follows an unknown distribution with density curve $f^{(0)}(z)$ on support $[\underline{z},\overline{z}]$. Under a given partitioning method, the support $[\underline{z},\overline{z}]$ is partitioned into $\cup_{i=1}^{m}[\underline{z}_i,\overline{z}_i]$, with $\overline{z}_i = \underline{z}_{i+1}, i = 1,\dots,m, \underline{z}_1 = \underline{z}, \overline{z}_m = \overline{z}$, and the density curve in a partition $[\underline{z}_i,\overline{z}_i]$ satisfies certain non-parametric characteristics. By total expectation, the expected profit is expressed as follows:

$$\Pi(q) = \sum_{i=1}^{m} E\left[\pi(q,Z)|z \in [\underline{z}_i,\overline{z}_i]\right]\mathbb{P}\left\{z \in [\underline{z}_i,\overline{z}_i]\right\}. \tag{10}$$

Let $\mathcal{P}_{[\underline{z}_i,\overline{z}_i]}$ be the distribution ambiguity set on $[\underline{z}_i,\overline{z}_i]$ and $\mathcal{P} = \bigcup_{i=1}^{m}\mathcal{P}_{[\underline{z}_i,\overline{z}_i]}$ be the distribution ambiguity set on the entire support. We derive the stocking quantity that maximizes the worst-case expected profit for the distributions within $\mathcal{P}$ as follows:

$$\max_{q}\min_{f^{(0)}(z)\in\mathcal{P}} \Pi(q). \tag{11}$$

The protection curve serves as the proxy for the worst-case distribution in the distribution ambiguity set. Under a fixed partitioning method, the optimal solution to problem (11) is obtained by the corresponding protection curve. Let $F_c(\cdot)$ be the function for the cumulative area under the protection curve. Obviously, $F_c(\cdot)$ is a non-decreasing function.

We first consider monotone partitioning, whereby the density curve in each partition is non-decreasing or non-increasing, and denote the protection curve for a partition $[\underline{z}_i,\overline{z}_i]$ by $c_i(z) = \ell_i$. The distribution ambiguity set $\mathcal{P}_{[\underline{z}_i,\overline{z}_i]}$ is defined in (1) for a non-increasing partition or in (2) for a non-decreasing partition. Proposition 4 offers a closed-form solution for problem (11) in this case.

PROPOSITION 4. *Under monotone partitioning, let* $p_i = \sum_{j=1}^{i}\ell_j\left(\overline{z}_j - \underline{z}_j\right), i = 1,2,\dots,m$ *and* $r = p_m \le 1$. *With* $\mu_c \triangleq r - \frac{c}{p}$, *a robust solution to the standard newsvendor problem is* $q_c = F_c^{-1}(\mu_c) = \underline{z}_{i+1} + \frac{\mu_c - p_i}{\ell_{i+1}}$, *when* $p_i \le \mu_c \le p_{i+1}, i = 1,2,\dots,m-1$.

Note that $\ell_j\left(\overline{z}_j - \underline{z}_j\right)$ is the area under the protection curve $c(z) = \ell_j$ in partition $[\underline{z}_j,\overline{z}_j]$, and p_i is the cumulative area under the protection curve from partition 1 to i. Obviously, $p_i \le p_{i+1}, i = 1,2,\dots,m-1$ and $F_c(\overline{z}) = p_m = r \le 1$. The condition $p_i \le \mu_c \le p_{i+1}$ identifies the partition $i+1$

that contains μ_c, which is the service ratio applicable to the protection curve and differs from the newsvendor service ratio μ_0.

We next consider full partitioning, in which case the protection curve for a partition $[\underline{z}_i, \overline{z}_i]$ takes the general form of $c(z) = \alpha_i + \beta_i (z - \underline{z}_i)$ and the distribution ambiguity sets $\mathcal{P}_{[\underline{z}_i, \overline{z}_i]}$ for different shapes of partitions are defined in (3) to (6).

Proposition 5. *Under full partitioning, let* $p_i = \sum_{j=1}^{i} [\alpha_j(\overline{z}_j - \underline{z}_j) + \frac{\beta_j}{2}(\overline{z}_j - \underline{z}_j)^2], i = 1, 2, \ldots, m,$ *where* α_j *and* β_j *are given in Table 1. A robust solution to the standard newsvendor problem is* $q_c = F_c^{-1}(\mu_c)$, *where* $F_c(z) = p_i + [\alpha_{i+1}(z - \underline{z}_{i+1}) + \frac{\beta_{i+1}}{2}(z - \underline{z}_{i+1})^2]$, $p_i \le \mu_c \le p_{i+1}, i = 1, 2, \ldots, m,$ *and* $\mu_c = \mu_0 = 1 - \frac{c}{p}$.

By Proposition 5, the optimal quantity is obtained as the μ_0-quantile of $F_c(\cdot)$. This is similar to the determination of the optimal stocking quantity with a known distribution $F(\cdot)$, in which case, the optimal quantity is the μ_0-quantile of $F(\cdot)$. Note that the protection curve constructed under full partitioning satisfies $F_c(\overline{z}) = 1$, i.e., the protection curve is a density curve.

We next discuss the solution method under semi-full partitioning, which combines the methods under monotone and full partitioning and can be modified to obtain robust solutions under adaptive partitioning (to be discussed later). Consider $[\underline{z}_1, \overline{z}_i] = \cup_{j=1}^{i} [\underline{z}_j, \overline{z}_j]$, among which k are non-decreasing monotone partitions and $u = i - k$ are non-increasing full partitions. Let the indices for the k monotone partitions be $\{D_1^i, D_2^i, \ldots, D_k^i\}$ and those for the u full partitions be $\{I_1^i, I_2^i, \ldots, I_u^i\}$, with $\{D_1^i, D_2^i, \ldots, D_k^i\} \cup \{I_1^i, I_2^i, \ldots, I_u^i\} = \{1, 2, \ldots, i\}$.

Proposition 6. *Under semi-full partitioning, let* $p_i = \sum_{j=D_1^i}^{D_k^i} \ell_j(\overline{z}_j - \underline{z}_j) + \sum_{j=I_1^i}^{I_u^i} [\alpha_j(\overline{z}_j - \underline{z}_j) + \frac{\beta_j}{2}(\overline{z}_j - \underline{z}_j)^2]$, $i = 1, \ldots, m$, *and* $r = p_m$. *A robust solution to the standard newsvendor problem is*

$$q_c = \begin{cases} H_c^{-1}(\mu_c), & p_i \le \mu_c \le p_{i+1}, f^{(1)}(z) \le 0, z \in [\underline{z}_{i+1}, \overline{z}_{i+1}] \\ \underline{z}_{i+1} + \frac{\mu_c - p_i}{\ell_{i+1}}, & p_i \le \mu_c \le p_{i+1}, f^{(1)}(z) \ge 0, z \in [\underline{z}_{i+1}, \overline{z}_{i+1}] \end{cases},$$

where $H_c(z) = p_i + \left[\alpha_{i+1}(z - \underline{z}_{i+1}) + \frac{\beta_{i+1}}{2}(z - \underline{z}_{i+1})^2\right]$ *and* $\mu_c = r - \frac{c}{p}$.

Proposition 6 states the robust stocking quantity obtained under semi-full partitioning. With p_i being the cumulative area under the protection curve up to partition i, the condition $p_i \le \mu_c \le p_{i+1}$ locates the partition that contains $\mu_c = r - \frac{c}{p}$. When the identified partition is a full partition, the optimal quantity is $H_c^{-1}(\mu_c)$, where $H_c(\cdot)$ is an auxiliary function to $F_c(\cdot)$. When it is a monotone partition, however, the optimal stocking quantity is obtained by interpolation.

Lemma 2. *Given a random variable on support* $[\underline{z}, \overline{z}]$ *and stocking quantity* q, *let* $\Pi_{mono}(q)$, $\Pi_{semi}(q)$, *and* $\Pi_{full}(q)$ *be the expected profits generated from the protection curves under monotone, semi-full, and full partitioning, respectively, then* $\Pi_{mono}(q) \le \Pi_{semi}(q) \le \Pi_{full}(q)$.

Lemma 2 states that, among the three partitioning methods, for a given stocking quantity, monotone partitioning yields the lowest worst-case expected profit because it generates the largest distribution ambiguity set with the least requirements on non-parametric characteristics, while full partitioning yields the highest worse-case expected profit despite the least stability of its distribution ambiguity set that requires the most characteristics. Semi-full partitioning balances stability and profitability in data-driven robust models.

4.2. Adaptive partitioning

Implementation-wise, we form non-parametric estimators based on data input to partition the support and construct the distribution ambiguity set. An important issue is how likely the true distribution belongs to the constructed distribution ambiguity set. To address this issue, we explore the unbiasedness in the monotonicity and convexity of the partitions formed by non-parametric estimators. It lays the groundwork to develop a procedure that adaptively feeds data input to divide, with a prespecified confidence level, the support into unbiased partitions, construct the distribution ambiguity set, and build the protection curve to obtain a robust stocking quantity.

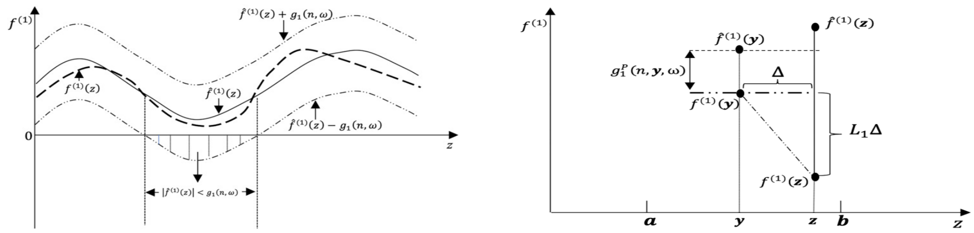
4.2.1. Unbiasedness in monotonicity and convexity of formed partitions

Consider a subset $[a,b]\subseteq[\underline{z},\overline{z}]$. Given confidence level $1-\omega$, the monotonicity of $[a,b]$ is unbiased if $\mathbb{P}\left\{\hat{f}^{(1)}(z)\,f^{(1)}(z)\geq 0:\forall z\in[a,b]\right\}\geq 1-\omega$, where $\hat{f}^{(1)}(z)$ and $f^{(1)}(z)$ are the empirical and true first-order derivatives, and the convexity of $[a,b]$ is unbiased if $\mathbb{P}\left\{\hat{f}^{(2)}(z)\,f^{(2)}(z)\geq 0:\forall z\in[a,b]\right\}\geq 1-\omega$, where $\hat{f}^{(2)}(z)$ and $f^{(2)}(z)$ are the empirical and true second-order derivatives. Proposition 7 provides sufficient conditions to guarantee the unbiasedness in the monotonicity or convexity of a subset based on uniform convergence of empirical first- and second-order derivatives.

Proposition 7. **(Uniform Convergence)** *Suppose* $\mathbb{P}\left\{\max_{z\in[a,b]}\left|\hat{f}^{(i)}(z)-f^{(i)}(z)\right|>g_i(n,\omega)\right\}\leq\omega$*, where* $g_i(n,\omega),i=1,2$ *are defined in (9) given* n *and* ω*. Let* Ω_i *be the largest subset of* $[a,b]$ *that* $\mathbb{P}\left\{\hat{f}^{(i)}(z)\,f^{(i)}(z)\geq 0:\forall z\in\Omega_i\right\}\geq 1-\omega$ *and* $U_i=\left\{z\in[a,b]:\left|\hat{f}^{(i)}(z)\right|\geq g_i(n,\omega)\right\}$*, then* $U_i\subseteq\Omega_i$*.*

Proposition 7 states that, under uniform convergence, for given data size n and confidence level $1-\omega$, provided that the norm of the empirical first-order (second-order) derivative on $[a,b]$ exceeds threshold $g_1(n,\omega)$ ($g_2(n,\omega)$), which is the uniform estimate error of the empirical relative to the true first-order (second-order) derivative, the monotonicity (convexity) of the subset is guaranteed. As the norm for the empirical first- or second-order derivative increases ($\left|\hat{f}^{(1)}(z)\right|$ or $\left|\hat{f}^{(2)}(z)\right|$ has a higher value), the monotonicity or convexity is more likely to be guaranteed. Note that uniform convergence is a strong condition. In Figure 3.a), even when the condition based on uniform convergence in Proposition 7 is violated, it can still hold that $\hat{f}^{(1)}(z)\,f^{(1)}(z)\geq 0$, i.e., the monotonicity of the partition formed by non-parametric estimators is unbiased.

Figure 3 **Point convergence of empirical first-order derivative**



(a) (b)

We next apply point convergence properties, which are discussed in Section 3.3.2, to obtain an alternative means of guaranteeing the unbiasedness in the monotonicity or convexity of a subset.

Proposition 8. **(Point Convergence)** *Suppose (1) $f^{(i)}(z)$ satisfies the Lipschitz condition, i.e., $\left|f^{(i)}(z_1)-f^{(i)}(z_2)\right|\leq L_i\left|z_1-z_2\right|,\forall z_1,z_2\in[a,b]$ and (2) $\mathbb{P}\left\{\left|\hat{f}^{(i)}(y)-f^{(i)}(y)\right|>g_i^P(n,y,\omega)\right\}\leq\omega$ for a certain $y\in[a,b]$, where $g_i^P(n,y,\omega),i=1,2$ are defined in (8) given n and ω. Let Ω_i be the largest subset of $[a,b]$ such that $\mathbb{P}\left\{\hat{f}^{(i)}(z)f^{(i)}(z)\geq 0:\forall z\in\Omega_i\right\}\geq 1-\omega$ and $U_i^P=\left\{z\in[a,b]:\left|\hat{f}^{(i)}(z)\right|\geq g_i^P(n,y,\omega)+|\hat{f}^{(i)}(z)-\hat{f}^{(i)}(y)|+L_i|z-y|\right\}$, then $U_i^P\subseteq\Omega_i$.*

By Proposition 8, when the true first-order (second-order) derivative satisfies the Lipschitz condition, any point convergence at $y\in[a,b]$, together with the satisfaction of the requirement on the norm of the empirical first-order (second-order) derivative, guarantees the monotonicity (convexity) of the partition formed by non-parametric estimators (see Figure 3.b).

4.2.2. Distribution ambiguity set

We first introduce adaptive partitioning under uniform convergence of non-parametric estimators and then modify it to apply point convergence of the estimators. Given data input and confidence level $1-\omega$, we apply Proposition 7 to obtain sets $U_i,i=1,2$ so that $[\underline{z},\overline{z}]=(U_1\cap U_2)\cup(U_1\setminus U_2)\cup\overline{U}_1$. Recall that the monotonicity of the partitions in U_1 and the convexity of the partitions in U_2 are guaranteed with confidence level $1-\omega$, but the monotonicity of the partitions in $\overline{U}_1$ is not guaranteed with the desired confidence level. Let $(U_1\cap U_2)\cup(U_1\setminus U_2)\cup\overline{U_1}=\cup_{i=1}^m[\underline{z}_i,\overline{z}_i]$, where $[\underline{z}_i,\overline{z}_i],i=1,2,\ldots,m$ are the partitions. We construct the distribution ambiguity set to contain the true distribution for each partition as follows:

For a partition $[\underline{z}_i,\overline{z}_i]\subseteq\overline{U}_1$:

$$\mathcal{P}_{[\underline{z}_i,\overline{z}_i]}=\{\mathbb{P}:f^{(0)}(z)\geq\underline{f}^{(0)}(z),\forall z\in[\underline{z}_i,\overline{z}_i]\}.\tag{12}$$

For a non-increasing monotone partition $[\underline{z}_i,\overline{z}_i]\subseteq U_1\setminus U_2$, where $\hat{f}^{(1)}(z)\leq 0,\forall z\in[\underline{z}_i,\overline{z}_i]$:

$$\mathcal{P}_{[\underline{z}_i,\overline{z}_i]}=\{\mathbb{P}:\bar{l}_i\geq l_i\geq\underline{l}_i,f^{(1)}(z)\leq 0,\forall z\in[\underline{z}_i,\overline{z}_i]\}.\tag{13}$$

For a non-decreasing monotone partition $[\underline{z}_i,\overline{z}_i]\subseteq U_1\setminus U_2$, where $\hat{f}^{(1)}(z)\geq 0,\forall z\in[\underline{z}_i,\overline{z}_i]$:

$$\mathcal{P}_{[\underline{z}_i,\overline{z}_i]}=\{\mathbb{P}:\overline{P}_i\geq P_i\geq\underline{P}_i,f^{(1)}(z)\geq 0,\forall z\in[\underline{z}_i,\overline{z}_i]\}.\tag{14}$$

For a non-decreasing convex partition $[\underline{z}_i, \overline{z}_i] \subseteq U_1 \cap U_2$, where $\hat{f}^{(1)}(z) \geq 0,\ \hat{f}^{(2)}(z) \geq 0,\ \forall z \in [\underline{z}_i, \overline{z}_i]$:

$$\mathcal{P}_{[\underline{z}_i, \overline{z}_i]} = \{\mathbb{P} : \bar{l}_i \geq l_i \geq \underline{l}_i, \overline{P}_i \geq P_i \geq \underline{P}_i, f^{(1)}(z) \geq 0, f^{(2)}(z) \geq 0, \forall z \in [\underline{z}_i, \overline{z}_i]\}. \tag{15}$$

For a non-increasing convex partition $[\underline{z}_i, \overline{z}_i] \subseteq U_1 \cap U_2$, where $\hat{f}^{(1)}(z) \leq 0,\ \hat{f}^{(2)}(z) \geq 0,\ \forall z \in [\underline{z}_i, \overline{z}_i]$:

$$\mathcal{P}_{[\underline{z}_i, \overline{z}_i]} = \{\mathbb{P} : \overline{u}_i \geq u_i \geq \underline{u}_i, \overline{P}_i \geq P_i \geq \underline{P}_i, f^{(1)}(z) \leq 0, f^{(2)}(z) \geq 0, \forall z \in [\underline{z}_i, \overline{z}_i]\}. \tag{16}$$

For a non-decreasing concave partition $[\underline{z}_i, \overline{z}_i] \subseteq U_1 \cap U_2$, where $\hat{f}^{(1)}(z) \geq 0,\ \hat{f}^{(2)}(z) \leq 0,\ \forall z \in [\underline{z}_i, \overline{z}_i]$:

$$\mathcal{P}_{[\underline{z}_i, \overline{z}_i]} = \{\mathbb{P} : \overline{u}_i \geq u_i \geq \underline{u}_i, \overline{P}_i \geq P_i \geq \underline{P}_i, f^{(1)}(z) \geq 0, f^{(2)}(z) \leq 0, \forall z \in [\underline{z}_i, \overline{z}_i]\}. \tag{17}$$

For a non-increasing concave partition $[\underline{z}_i, \overline{z}_i] \subseteq U_1 \cap U_2$, where $\hat{f}^{(1)}(z) \leq 0,\ \hat{f}^{(2)}(z) \leq 0,\ \forall z \in [\underline{z}_i, \overline{z}_i]$:

$$\mathcal{P}_{[\underline{z}_i, \overline{z}_i]} = \{\mathbb{P} : \bar{l}_i \geq l_i \geq \underline{l}_i, \overline{P}_i \geq P_i \geq \underline{P}_i, f^{(1)}(z) \leq 0, f^{(2)}(z) \leq 0, \forall z \in [\underline{z}_i, \overline{z}_i]\}. \tag{18}$$

These distribution ambiguity sets on the partitions are formed by non-parametric estimators and estimate errors. Specifically, $\underline{f}^{(0)}(z) = \hat{f}^{(0)}(z) - g_0(n, \omega)$, $\overline{u}_i = \hat{u}_i + g_0(n, \omega)$, $\underline{u}_i = \hat{u}_i - g_0(n, \omega)$, $\bar{l}_i = \hat{l}_i + g_0(n, \omega)$, $\underline{l}_i = \hat{l}_i - g_0(n, \omega)$, $\overline{P}_i = \hat{P}_i + g_0(n, \omega)(\overline{z}_i - \underline{z}_i)$, and $\underline{P}_i = \hat{P}_i - g_0(n, \omega)(\overline{z}_i - \underline{z}_i)$. The distribution ambiguity set $\mathcal{P}$ on the support $[\underline{z}, \overline{z}]$ can be obtained by unionizing the distribution ambiguity sets $\mathcal{P}_{[\underline{z}_i, \overline{z}_i]}$ on all partitions $[\underline{z}_i, \overline{z}_i], i = 1, 2, ..., m$. Proposition 9 states the probability that the true distribution belongs to $\mathcal{P}$.

PROPOSITION 9. *Given data size n and confidence level $1 - \omega$, under uniform convergence of non-parametric estimators, the distribution ambiguity set $\mathcal{P}$ includes the true distribution on the entire support with a probability of at least $(1 - \omega)(1 - 2\omega)$, which is nonnegative when $\omega \leq 0.5$.*

Thus, under uniform convergence, the true distribution has a guaranteed probability $(1 - \omega)(1 - 2\omega)$ of belonging to the distribution ambiguity set formed by non-parametric estimators and their errors. However, estimate error $g_0(n, \omega)$ can be large, inflating the distribution ambiguity set. As $g_i^P(n, y, \omega)$ can be smaller than $g_i(n, \omega)$ in a given interval, we next apply point convergence properties of non-parametric estimators to guide partitioning and improve the worst-case expected profit. Note that partitioning under point convergence guarantees the unbiasedness of non-parametric characteristics within partitions, while that under uniform convergence guarantees the unbiasedness on the entire support.

Under point convergence, to guarantee the monotonicity or convexity of a formed partition with confidence, the norm of $\hat{f}^{(i)}(z)$ should satisfy $|\hat{f}^{(i)}(z)| \geq g_i^P(n, y, \omega) + |\hat{f}^{(i)}(z) - \hat{f}^{(i)}(y)| + L_i|z - y|$ for a certain y in the partition. We first apply full partitioning directly based on $\hat{f}^{(1)}(z)$ and $\hat{f}^{(2)}(z)$ to obtain $[\underline{z}, \overline{z}] = \cup_{j=1}^{k} [\underline{z}_j, \overline{z}_j]$. The Lipshitz condition holds for $f^{(i)}(z), i = 0, 1, 2$ provided that $f^{(i+1)}(z)$

is continuous in each bounded interval $[\underline{z}_j, \overline{z}_j], j = 1, 2, \ldots, k$, enabling us to apply Proposition 8 with $a = \underline{z}_j$ and $b = \overline{z}_j$ to obtain $U^P_{1,j} \subseteq [\underline{z}_j, \overline{z}_j]$ and $U^P_{2,j} \subseteq [\underline{z}_j, \overline{z}_j]$, wherein the monotonicity and convexity are guaranteed with confidence level $1 - w$. With $U^P_1 = \cup^k_{j=1} U^P_{1,j}$ and $U^P_2 = \cup^k_{j=1} U^P_{2,j}$, we partition the support into $[\underline{z}, \overline{z}] = (U^P_1 \cap U^P_2) \cup (U^P_1 \setminus U^P_2) \cup \overline{U^P_1} = \cup^m_{i=1} [\underline{z}_i, \overline{z}_i]$.

With estimate errors $g^P_i(n, z, \omega), i = 0, 1, 2$ and $g^P(n, z, \omega)$, the distribution ambiguity set $\mathcal{P}^P_{[\underline{z}_i, \overline{z}_i]}$ on a partition $[\underline{z}_i, \overline{z}_i]$ can be formed similarly as that under uniform convergence with the following adjustments in (12) to (18): $\underline{f}^{(0)}(z) = \hat{f}^{(0)}(z) - g^P_0(n, y, \omega) - L_0(\overline{z}_i - \underline{z}_i)$, $\overline{l}_i = \hat{l}_i + g^P_0(n, \overline{z}_i, \omega)$, $\underline{l}_i = \hat{l}_i - g^P_0(n, \overline{z}_i, \omega)$, $\overline{P}_i = \hat{P}_i + g^P(n, \underline{z}_i, \omega) + g^P(n, \overline{z}_i, \omega)$, $\underline{P}_i = \hat{P}_i - g^P(n, \underline{z}_i, \omega) - g^P(n, \overline{z}_i, \omega)$, $\overline{u}_i = \hat{u}_i + g^P_0(n, \underline{z}_i, \omega)$, and $\underline{u}_i = \hat{u}_i - g^P_0(n, \underline{z}_i, \omega)$. Proposition 10 states the probability that the true distribution restricted on $[\underline{z}_i, \overline{z}_i]$ belongs to the distribution ambiguity set $\mathcal{P}^P_{[\underline{z}_i, \overline{z}_i]}$ and the probability that the true distribution on the support belongs to $\mathcal{P}^P$, which is the union of all $\mathcal{P}^P_{[\underline{z}_i, \overline{z}_i]}, i = 1, 2, \ldots, m$.

Proposition 10. *Given data size n and confidence level $1 - \omega$, under point convergence of non-parametric estimators, the distribution ambiguity set $\mathcal{P}^P_{[\underline{z}_i, \overline{z}_i]}$ includes the true distribution in partition $[\underline{z}_i, \overline{z}_i]$ with a probability of at least $(1 - 2\omega)(1 - 3\omega)$, and the distribution ambiguity set $\mathcal{P}^P$ includes the true distribution on the support $[\underline{z}, \overline{z}] = \cup^m_{i=1} [\underline{z}_i, \overline{z}_i]$ with a probability of at least $1 - m(5\omega - 6\omega^2)$, which is nonnegative when $\omega \le \omega_0 \triangleq \frac{5\sqrt{m} - \sqrt{25m - 24}}{12\sqrt{m}}$.*

Proposition 10 states $1 - m(5\omega - 6\omega^2)$ as a lower bound for the probability that the true density belongs to distribution ambiguity set $\mathcal{P}^P$. It bounds the likelihood that the realized average profit exceeds the worst-case expected profit within the ambiguity set. When partition number m is small, the threshold ω_0 is not too conservative, and the stated probability bound provides a reasonable indicator for a given $\omega < \omega_0$. However, when partition number m is large, the confidence level to guide support partitioning needs to be sufficiently high (ω is sufficiently low) to ensure that the stated probability bound has indicative value. Experiments in Section 6.1 reveal that, by applying point convergence of non-parametric estimators in adaptive partitioning, the realized average profit exceeds the worst-case expected profit with a probability close to one. It implies a high chance that the true distribution belongs to the distribution ambiguity set constructed by our approach.

4.2.3. Protection curve and stocking quantities

Under uniform-convergence-based adaptive partitioning, the support is divided into partitions in subsets $\overline{U_1}$, $U_1 \setminus U_2$, and $U_1 \cap U_2$, i.e., $[\underline{z}, \overline{z}] = (U_1 \cap U_2) \cup (U_1 \setminus U_2) \cup \overline{U_1}$, with prespecified confidence level. For the distribution ambiguity set $\mathcal{P}_{[\underline{z}_i, \overline{z}_i]}$ on a partition $[\underline{z}_i, \overline{z}_i]$, we derive protection curve $c(z) = \underline{f}^{(0)}(z)$ when $[\underline{z}_i, \overline{z}_i] \subseteq \overline{U}_1$, and $c(z) = \underline{l}_i$ or $c(z) = \frac{\underline{P}_i}{\overline{z}_i - \underline{z}_i}$ when $[\underline{z}_i, \overline{z}_i] \subseteq U_1 \setminus U_2$. For a partition $[\underline{z}_i, \overline{z}_i] \subseteq U_1 \cap U_2$, the protection curve takes a general form of $c(z) = \alpha'_i + \beta'_i (z - \underline{z}_i)$, where coefficients (α'_i, β'_i) are shown in Table 2. The protection curve for distribution ambiguity set $\mathcal{P}$ on the support is the union of those for $\mathcal{P}_{[\underline{z}_i, \overline{z}_i]}, i = 1, 2, ..., m$ on all partitions.

Table 2 **Coefficients for protection curve** $c(z) = \alpha_i' + \beta_i'(z - \underline{z}_i)$

partition $[\underline{z}_i, \overline{z}_i]$	α_i	β_i
non-decreasing and convex	$\bar{l}_i$	$2\frac{\underline{P}_i - \bar{l}_i(\overline{z}_i - \underline{z}_i)}{(\overline{z}_i - \underline{z}_i)^2}$
non-increasing and convex	$\overline{u}_i$	$-2\frac{\overline{u}_i(\overline{z}_i - \underline{z}_i) - \underline{P}_i}{(\overline{z}_i - \underline{z}_i)^2}$
non-decreasing and concave	$\frac{2\underline{P}_i}{\overline{z}_i - \underline{z}_i} - \underline{u}_i$	$2\frac{\underline{u}_i(\overline{z}_i - \underline{z}_i) - \underline{P}_i}{(\overline{z}_i - \underline{z}_i)^2}$
non-increasing and concave	$\frac{2\underline{P}_i}{\overline{z}_i - \underline{z}_i} - \underline{l}_i$	$-2\frac{\underline{P}_i - \underline{l}_i(\overline{z}_i - \underline{z}_i)}{(\overline{z}_i - \underline{z}_i)^2}$

Under point-convergence-based adaptive partitioning, the protection curve for distribution ambiguity set $\mathcal{P}^P$ can be built by the same procedure as that under uniform convergence except that parameters $\underline{f}^{(0)}(z), \underline{l}_i, \bar{l}_i, \underline{u}_i, \overline{u}_i, \underline{P}_i, \overline{P}_i$ shall be replaced by those applicable under point convergence. It should be noted that the protection curve thus built is similar to the one developed under semi-full partitioning, which is a combination of flat lines and lines with a slope. We can apply the solution method in Proposition 6 to obtain robust quantities by adaptive partitioning. The likelihood that the realized average profit exceeds the worst-case expected profit can be guaranteed, given the probability that the true distribution belongs to the constructed distribution ambiguity set, as presented in Propositions 9 and 10 for uniform and point convergence respectively.

4.3. Non-parametric approach versus empirical approach

The empirical density function converges to the true density function as data size increases but the protection curve provides a safe protection when data size is small. By the convergence property stated in (9), the empirical density function satisfies $\mathbb{P}\left\{\max_{z\in[\underline{z},\overline{z}]}\left|\hat{f}^{(0)}(z) - f^{(0)}(z)\right| > g_0(n,\omega)\right\} \leq \omega$, where $g_0(n,\omega) = \frac{D_n(\omega)}{\lambda} + \frac{K_0\lambda}{2}$ and $|f^{(1)}(z)| \leq K_0, \forall z$. Let $\varepsilon = g_0(n,\omega)$. An ε-distribution ambiguity set is defined as $\mathcal{P}(\varepsilon) = \left\{\mathbb{P} : \max_{z\in[\underline{z},\overline{z}]}\left|\hat{f}^{(0)}(z) - f^{(0)}(z)\right| \leq \varepsilon\right\}$, and $\max_q \min_{f\in\mathcal{P}(\varepsilon)} [E(\pi(q,z))]$ yields a stocking quantity q^* that maximizes the worst-case expected profit when the true distribution belongs to $\mathcal{P}(\varepsilon)$. The realized average profit under this quantity exceeds the worst-case expected profit in $\mathcal{P}(\varepsilon)$, which is $\int_{\underline{z}}^{\overline{z}} \pi(q^*,z)(\hat{f}^{(0)}(z) - g_0(n,\omega))dz$, with a probability of at least $1-\omega$. To see this, note that $\mathbb{P}\{\int_{\underline{z}}^{\overline{z}} \pi(q^*,z)(\hat{f}^{(0)}(z) - f^{(0)}(z))dz \leq \int_{\underline{z}}^{\overline{z}} \pi(q^*,z) g_0(n,\omega)dz\} \geq \mathbb{P}\{\int_{\underline{z}}^{\overline{z}} \pi(q^*,z) \max_{z\in[\underline{z},\overline{z}]}(\hat{f}^{(0)}(z) - f^{(0)}(z))dz \leq \int_{\underline{z}}^{\overline{z}} \pi(q^*,z) g_0(n,\omega)dz\} = \mathbb{P}\{\max_{z\in[\underline{z},\overline{z}]}(\hat{f}^{(0)}(z) - f^{(0)}(z)) \leq g_0(n,\omega)\} \geq 1-\omega$. It holds that $\mathbb{P}\{\int_{\underline{z}}^{\overline{z}} \pi(q^*,z) f^{(0)}(z)dz \geq \int_{\underline{z}}^{\overline{z}} \pi(q^*,z)(\hat{f}^{(0)}(z) - g_0(n,\omega))dz\} \geq 1-\omega$.

Following a similar logic, we can show that $1-\omega$ also bounds the probability that, under the stocking quantity obtained by our approach with adaptive partitioning, the realized average profit exceeds the worst-case expected profit within $\mathcal{P}(\varepsilon)$. This is regardless of the convergence properties (uniform vs point) that we apply and the number of partitions. Thus, we obtain an alternative means of evaluating the effectiveness of our approach. As data size increases and ω and λ approach zero, $\mathcal{P}(\varepsilon)$ converges to containing only the true distribution. Given the same confidence level $1-\omega$, experiments in Section 6.1 reveal the existence of a threshold such that ε-distribution ambiguity set $\mathcal{P}(\varepsilon)$, which is based on empirical density function, is preferred to obtain robust decisions for its convergence performance when data size exceeds the threshold, but distribution ambiguity set $\mathcal{P}^P$, which utilizes point convergence properties, is preferred to obtain safe robust decisions otherwise.

5. Value of Non-parametric Information

Next, we investigate the value of non-parametric characteristics relative to parametric information in data-driven robust decision models through a comparative study. To that end, we consider a discrete model in which a random variable follows a mass function. A continuous model can hardly yield closed-form stocking quantities because the worst-case distribution in the ambiguity set with non-parametric and parametric information cannot be specified. In a discrete model, the stocking quantities can be enumerated to overcome this difficulty.

Let random demand Z be discrete with an unknown mass function on $[\underline{z}, \overline{z}]$. Suppose the support has m partitions, denoted as $(\overline{z}_{i-1}, \overline{z}_i], i = 1, 2, \ldots, m$, with $[\underline{z}, \overline{z}] = (\overline{z}_0, \overline{z}_1] \cup (\overline{z}_1, \overline{z}_2] \cup \ldots \cup (\overline{z}_{m-1}, \overline{z}_m]$, where $\overline{z}_0 < \underline{z}$ is a nominal value with $\mathbb{P}\{Z \le \overline{z}_0\} = 0$. In partition $(\overline{z}_{i-1}, \overline{z}_i]$, the random demand assumes value from a finite set $\{z_i^1, z_i^2, \ldots, z_i^n\}$, with $\overline{z}_{i-1} = z_i^0 < z_i^1 = \underline{z}_i < z_i^2 < \ldots < z_i^n = \overline{z}_i$, i.e., $\underline{z}_i = z_i^1$ is the lowest demand in $(\overline{z}_{i-1}, \overline{z}_i]$ and $\overline{z}_{i-1} = z_{i-1}^n$ is the largest value of the preceding partition. Let the parametric information be the first and second moments (e_1^i, e_2^i) for the random variable in the partition, and the non-parametric information be based on any protection curve $c(z)$ that satisfies $\mathbb{P}\{z \in [z_i^y, \overline{z}_i]\} - \int_{z_i^{y-1}}^{\overline{z}_i} c(z)\, dz \ge 0, y \in \{1, 2, \ldots n\}$. The distribution ambiguity set, which includes both types of information in partition $(\overline{z}_{i-1}, \overline{z}_i]$, is defined as follows:

$$\mathcal{P}^{pn}_{(\overline{z}_{i-1}, \overline{z}_i]} = \{\mathbb{P} : E_{\mathbb{P}}\left[z^j \middle| z \in (\overline{z}_{i-1}, \overline{z}_i]\right] \mathbb{P}\{z \in (\overline{z}_{i-1}, \overline{z}_i]\} = e_j^i, j = 1, 2; \mathbb{P}\{z \in [z_i^y, \overline{z}_i]\} - \int_{z_i^{y-1}}^{\overline{z}_i} c(z)\, dz \ge 0, y \in \{1, \cdots, n\}\}. \tag{19}$$

Suppose that the stocking quantity is chosen from a finite set $Q = \{q_1, q_2, \ldots, q_a\}$. The following problem maximizes the expected newsvendor profit:

$$\max \sum_{k=1}^{a} \Pi(q_k)\, x_k, \text{subject to} : \sum_{k=1}^{a} x_k = 1, x_k \in \{0, 1\}, \tag{20}$$

where $\Pi(q_k) = E[p \min\{z, q_k\} - cq_k]$. In the formulation, x_k is a binary indicator variable, with $x_k = 1$ if q_k is chosen but $x_k = 0$ otherwise. By total expectation, the newsvendor profit under a given stocking quantity q_k can be written as follows:

$$\Pi(q_k) = p \sum_{i=1}^{m} E[\min\{z, q_k\} | z \in (\overline{z}_{i-1}, \overline{z}_i]]\, \mathbb{P}\{z \in (\overline{z}_{i-1}, \overline{z}_i]\} - cq_k. \tag{21}$$

For partition $(\overline{z}_{i-1}, \overline{z}_i]$, the worst-case expected profit in the distribution ambiguity set, which is defined in (19), can be written as:

$$\inf_{\mathbb{P} \in \mathcal{P}^{pn}_{(\overline{z}_{i-1}, \overline{z}_i]}} E[\min\{z, q_k\} | z \in (\overline{z}_{i-1}, \overline{z}_i]] \mathbb{P}\{z \in (\overline{z}_{i-1}, \overline{z}_i]\} = \max_{s_1, s_2, u_1, \ldots, u_n \ge 0} \Upsilon(s_1, s_2, u_1, \ldots, u_n), \tag{22}$$

where $\Upsilon(s_1, s_2, u_1, \ldots, u_n) = \sum_{j=1}^{2} e_j^i s_j + \sum_{l=1}^{n} u_l \int_{z_i^{l-1}}^{\overline{z}_i} c(z)\,dz$, s_1 and s_2 are the dual variables associated with e_1^i and e_2^i, and $u_1, \ldots, u_n$ are those associated with $\int_{z_i^{l-1}}^{\overline{z}_i} c(z)\,dz, l = 1, \ldots, n$. These dual variables satisfy the following inequalities:

$$s_1 z_i^l + s_2 \left(z_i^l\right)^2 + \sum_{j=1}^{l} u_j \leq \min\left\{z_i^l, q_k\right\}, l \in \{1, \ldots, n\}. \tag{23}$$

We can construct the distribution ambiguity set solely with parametric or non-parametric information by using the relevant type of information in (19), and use $\mathcal{P}^p$ and $\mathcal{P}^n$ to denote the respective distribution ambiguity sets. It should be noted that $\mathcal{P}^p$ is similar in form to the distribution ambiguity set constructed in Natarajan et al. (2018), but the partitions are developed by our non-parametric approach. With non-parametric information included in $\mathcal{P}^p$, $\mathcal{P}^{pn}$ is smaller than $\mathcal{P}^p$ because of the additional constraints to form the set, and the objective function value for problem (22) under $\mathcal{P}^{pn}$ is no worse than that under $\mathcal{P}^p$. Let the stocking quantities that maximize the worst-case expected profits under $\mathcal{P}^p$ and $\mathcal{P}^n$ be q_p^* and q_n^*, respectively. We use $G_\pi \triangleq \left[\inf_{\mathbb{P} \in \mathcal{P}^n} \Pi(q_n^*) - \inf_{\mathbb{P} \in \mathcal{P}^p} \Pi\left(q_p^*\right)\right]$ to measure the value of non-parametric relative to parametric information in yielding the worst-case expected profit. Our experiments reveal that non-parametric information is more valuable ($G_\pi > 0$) unless the service ratio is very high.

5.1. Value of non-parametric information for given stocking quantity

To shed more light on the relative value of using non-parametric information in constructing the distribution ambiguity set for robust decision making, we compare the expected newsvendor profits obtained under two types of information for a given stocking quantity. Let $\pi_i \triangleq \int_{\overline{z}_{i-1}}^{\overline{z}_i} c(z)\,dz$, where $c(z)$ is the protection curve, be the non-parametric characteristics, and e_1^i and e_2^i be the parametric information to construct the distribution ambiguity set for $(\overline{z}_{i-1}, \overline{z}_i]$. For a given stocking quantity q_k, we reformulate problem (19) as follows:

$$\inf_{\mathbb{P} \in \mathcal{P}^{pn}_{(\overline{z}_{i-1}, \overline{z}_i]}} E\left[\min\{z, q_k\} | z \in (\overline{z}_{i-1}, \overline{z}_i]\right] \mathbb{P}\left\{z \in (\overline{z}_{i-1}, \overline{z}_i]\right\} = \max_{s_1, s_2, u_1 \geq 0} \Upsilon(s_1, s_2, u_1), \tag{24}$$

where $\Upsilon(s_1, s_2, u_1) = e_1^i s_1 + e_2^i s_2 + \pi_i u_1$, and s_1, s_2, u_1 are the associated dual variables.

The dual variables satisfy the following conditions:

$$s_1 \underline{z}_i + s_2 \underline{z}_i^2 + u_1 \leq \min\{\underline{z}_i, q_k\}; s_1 \overline{z}_i + s_2 \overline{z}_i^2 + u_1 \leq \min\{\overline{z}_i, q_k\}; s_1, s_2, u_1 \geq 0. \tag{25}$$

Note that $e_1^i = \sum_{l=1}^{n} z_i^l p_i^l$, where p_i^l is the probability mass for scenario z_i^l. Provided the protection curve $c(z)$ satisfies $\int_{\overline{z}_{i-1}}^{\overline{z}_i} c(z)\,dz = \sum_{l=1}^{n} p_i^l$, the following inequalities hold: $\frac{e_1^i}{\overline{z}_i} < \pi_i < \frac{e_1^i}{\underline{z}_i}$ and $\frac{e_2^i}{\overline{z}_i^2} < \pi_i < \frac{e_2^i}{\underline{z}_i^2}$. Hence, the magnitude of the non-parametric information is bounded by the moments scaled by the inverse of the lower and upper boundaries for the partition. The candidate optimal solutions

Table 3 **Candidate optimal solutions.**

q_k	(u_1, s_1, s_2)
$q_k \le \overline{z}_{i-1}$	$(q_k, 0, 0), \left(0, \frac{q_k}{\overline{z}_i}, 0\right), \left(0, 0, \frac{q_k}{\overline{z}_i^2}\right)$
$\overline{z}_{i-1} < q_k < \overline{z}_i$	$(\underline{z}_i, 0, 0), \left(0, \frac{q_k}{\overline{z}_i}, 0\right), \left(0, 0, \frac{q_k}{\overline{z}_i^2}\right), \left(\frac{\underline{z}_i \overline{z}_i - \underline{z}_i q_k}{\overline{z}_i - \underline{z}_i}, \frac{q_k - \underline{z}_i}{\overline{z}_i - \underline{z}_i}, 0\right), \left(\frac{\underline{z}_i \overline{z}_i^2 - \underline{z}_i^2 q_k}{\overline{z}_i^2 - \underline{z}_i^2}, 0, \frac{q_k - \underline{z}_i}{\overline{z}_i^2 - \underline{z}_i^2}\right)$
$q_k \ge \overline{z}_i$	$(\underline{z}_i, 0, 0), (0, 1, 0), \left(0, 0, \frac{1}{\overline{z}_i}\right), \left(\frac{\overline{z}_i \underline{z}_i}{\overline{z}_i + \underline{z}_i}, 0, \frac{1}{\overline{z}_i + \underline{z}_i}\right)$

to (u_1, s_1, s_2) are listed in Table 3 for different values of q_k relative to partition boundaries. The proofs are provided in the Appendix.

In Table 3, at least one dual variable is zero, indicating that the corresponding parametric or non-parametric information is inconsequential to the objective function value. Lemma 3 states and compares the optimal solutions to u_1^*, which is associated with the non-parametric information, and (s_1^*, s_2^*), which are associated with the parametric information.

LEMMA 3. *According to formulation (24), given a stocking quantity q_k and partition $(\overline{z}_{i-1}, \overline{z}_i]$:*

1) When $q_k \le \overline{z}_{i-1} : u_1^* \ge s_1^* = s_2^* = 0$.

2) When $\overline{z}_{i-1} < q_k < \overline{z}_i$:

a) $s_2^* = 0$ if $e_1^i (\overline{z}_i + \underline{z}_i) - e_2^i \ge \overline{z}_i \underline{z}_i \pi_i$, in which case, $u_1^* \ge s_1^*$ if $\frac{q_k}{\underline{z}_i} \le \frac{1+\overline{z}_i}{1+\underline{z}_i}$.

b) $s_1^* = 0$ if $e_1^i (\overline{z}_i + \underline{z}_i) - e_2^i < \overline{z}_i \underline{z}_i \pi_i$, in which case, $u_1^* \ge s_2^*$ if $\frac{q_k}{\underline{z}_i} \le \frac{1+\overline{z}_i^2}{1+\underline{z}_i^2}$.

3) When $q_k \ge \overline{z}_i$:

a) $s_2^* = 0$ if $e_1^i (\overline{z}_i + \underline{z}_i) - e_2^i \ge \overline{z}_i \underline{z}_i \pi_i$, in which case, $u_1^* = 0$.

b) $s_1^* = 0$ if $e_1^i (\overline{z}_i + \underline{z}_i) - e_2^i < \overline{z}_i \underline{z}_i \pi_i$, in which case, $u_1^* \ge s_2^*$ if $\overline{z}_i \underline{z}_i \ge 1$.

Lemma 3 states that, when the stocking quantity is below the lower partition boundary (case 1), non-parametric information dominates in yielding the worst-case expected profit, while neither of the two moments is consequential to the objective function value. Otherwise, either the first or the second moment is influential, depending on their scale relative to each other. When the stocking quantity falls in the partition (case 2), non-parametric information is more valuable than the parametric information that is influential if the scale of stocking quantity is small. As the stocking quantity inches above the upper boundary (case 3), whenever the second moment is inconsequential to the objective function value, so is the non-parametric information. However, when the first moment is inconsequential, the non-parametric information is more valuable than the second moment when the geometric mean of the lower and upper partition boundaries is not too small.

PROPOSITION 11. *Consider a random demand Z with a mass function on support $[\underline{z}, \overline{z}]$ which has m partitions. Let parametric information be e_1^i, e_2^i and non-parametric information be π_i for all partitions $(\overline{z}_{i-1}, \overline{z}_i]$, $i = 1, 2, \ldots, m$. Given a stocking quantity q_k, from the perspective of yielding* *the worst-case expected profit: 1) if $q_k < \underline{z}_1$, then non-parametric information is more valuable; 2) if $\underline{z}_1 \le q_k < \overline{z}_m$, then, let $\underline{z}_i \le q_k < \overline{z}_i$ ($q_k = \overline{z}_i, i \ne m$) for some i, non-parametric information is more valuable if $\sum_{j=1}^{i} \pi_j \underline{z}_j + \sum_{j=i+1}^{m} \pi_j q_k > \sum_{j=1}^{i-1} e_1^j + \sum_{j=i}^{m} \frac{e_1^j q_k}{\overline{z}_j}$ ($\sum_{j=1}^{i} \pi_j \underline{z}_j + \sum_{j=i+1}^{m} \pi_j q_k > \sum_{j=1}^{i} e_1^j + \sum_{j=i+1}^{m} \frac{e_1^j q_k}{\overline{z}_j}$), while parametric information is more valuable otherwise; 3) if $q_k \ge \overline{z}_m$, then parametric information is more valuable.*

On the entire support, non-parametric information dominates in yielding the worst-case expected profit when the chosen stocking quantity is below the lower boundary of the support (and hence always unable to fully satisfy demand), while parametric information dominates when it is above the upper boundary of the support (and hence always able to fully satisfy demand). When it falls in a certain partition, non-parametric information is more valuable when the weighted cumulative area of all partitions under the protection curve is larger than the scaled first moment, while the parametric information is more valuable otherwise. Notably, the second moment is inconsequential to comparison outcomes.

5.2. Remarks on multi-item setting

Our proposed non-parametric approach can be adapted to the situation when the newsvendor manages multiple items. With non-parametric characteristics of the distribution to form the distribution ambiguity set and build the protection curve, we can apply the approach to each item individually provided that marginal data is available. It is also feasible to partition the support by our non-parametric approach, and include both non-parametric and parametric information in each partition to form the distribution ambiguity set and derive robust decisions. To that end, in addition to the first and second moments for each item within each partition as shown in (19), we add more information on the covariance of demands among items to improve performance.

Specifically, we consider multivariate discrete random demand $Z = (Z_1, Z_2, \cdots, Z_T)$ for T items, with realizations $z = (z_1, z_2, \cdots, z_T)$. Suppose Z has X possible realizations $\left\{z^{(1)}, z^{(2)}, \ldots z^{(X)}\right\}$ with an unknown joint mass function $f\left(z^{(x)}\right) = f^x$, where $z^{(x)} \in \mathbb{R}^T, x = 1, 2, ..., X$. The stocking quantities $q = (q_1, q_2, \cdots, q_T)$ for the T items belong to a finite feasible set Q. An item $t \in \{1, 2, ..., T\}$ has price p_t and cost c_t. Let $p = (p_1, p_2, \cdots, p_T)$. Unsold items at the end of period have no salvage value. Let $\Theta = \{\Theta_1, \Theta_2, \ldots, \Theta_T\}$, where $\Theta_t, t = 1, ..., T$ is the set of partitions on the support of demand for item t, and $\Theta_t^i = (\overline{z}_{t,i-1}, \overline{z}_{t,i}]$, $i = 1, \ldots, |\Theta_t|$ be a partition in Θ_t. We assume $z_t \in \{z_{t,i}^1, z_{t,i}^2, \cdots, z_{t,i}^n\}$ in Θ_t^i, with $\overline{z}_{t,i-1} = z_{t,i}^0 < z_{t,i}^1 = \underline{z}_{t,i} < z_{t,i}^2 < \ldots < z_{t,i}^n = \overline{z}_{t,i}$. For the parametric information, let e_t be the mean of marginal demand for item t on the support, $(e_{t,1}^i, e_{t,2}^i)$ be the first and second moments of marginal demand for item t in partition Θ_t^i, and σ_{t_1,t_2} be the covariance between the demands for any two items t_1 and t_2. For the non-parametric information, let $c_t(z), t = 1, 2, ..., T$ be the protection curve based on non-parameteric characteristics of the marginal distribution of the demand for each item t. With both parametric and non-parametric information to construct distribution ambiguity set $\mathcal{P}_{\Theta}^{pn}$, the robust newsvendor problem is formulated as:

$$\max_{q\in Q}\min_{f\in\mathcal{P}_{\Theta}^{pn}} E_f\left(\sum_{t=1}^{T} p_t \min\{z_t,q_t\} - \sum_{t=1}^{T} c_t q_t\right)$$

We can apply the discrete-choice model to solve this robust optimization problem. Given stocking quantities q, the inner problem can be formulated as:

$$\inf_{f\in\mathcal{P}_{\Theta}^{pn}} E\left[\sum_{t=1}^{T} p_t \min\{z_t,q_t\}\right] = \min \sum_{x=1}^{X} (p\cdot \min\{z^{(x)},q\}) f^x$$

$$s.t.\ \sum_{l=1}^{n}((z_{t,i}^{l})^j \sum_{x=1}^{X}(f^x \mathbb{I}\{z_t^{(x)}=z_{t,i}^{l}\})) \geq e_{t,j}^{i}, j=1,2, \forall i=1,2,\ldots,|\Theta_t|, \forall t=1,\ldots,T;$$

$$\sum_{l=y}^{n}\sum_{x=1}^{X}(f^x \mathbb{I}\{z_t^{(x)}=z_{t,i}^{l}\}) \geq \int_{z_{t,i}^{y-1}}^{\overline{z}_{t,i}} c_t(z)dz, y\in\{1,2,\ldots n\}, \forall i=1,2,\ldots,|\Theta_t|, \forall t=1,\ldots,T;$$

$$\sum_{x=1}^{X}(z_{t_1}^{(x)}-e_{t_1})(z_{t_2}^{(x)}-e_{t_2}) f^x = \sigma_{t_1,t_2}, \forall t_1,t_2=1,\ldots,T; f^1,\ldots,\ f^X \geq 0,$$

where the indicator function $\mathbb{I}\{z_t^{(x)}=z_{t,i}^{l}\}$ equals one when the tth attribute of $z^{(x)}$ equals $z_{t,i}^{l}$ and zero otherwise, and $\sum_{x=1}^{X} f^x \mathbb{I}\{z_t^{(x)}=z_{t,i}^{l}\}$ equals the marginal mass probability for realization $z_{t,i}^{l}$. The first set of constraints are on the first and second moments of the marginal demand for each item in all its partitions. The second set of constraints use the non-parametric information based on the protection curve of the marginal demand for each item in its partitions. The third set of constraints are on the covariance between the demands for any two items t_1 and t_2.

This multivariate discrete-choice model can be solved similarly as that for the model with a single item, except that attention should be given to demand covariance. When the constraints on covariance are removed, the model can be solved by considering each item individually and the robust quantity obtained by the single-item model is also optimal to the multi-item model.

6. Experiments

We perform a series of experiments to test the effectiveness and practicability of the proposed data-driven non-parametric approach.

6.1. Non-parametric approach applied to arbitrary distribution

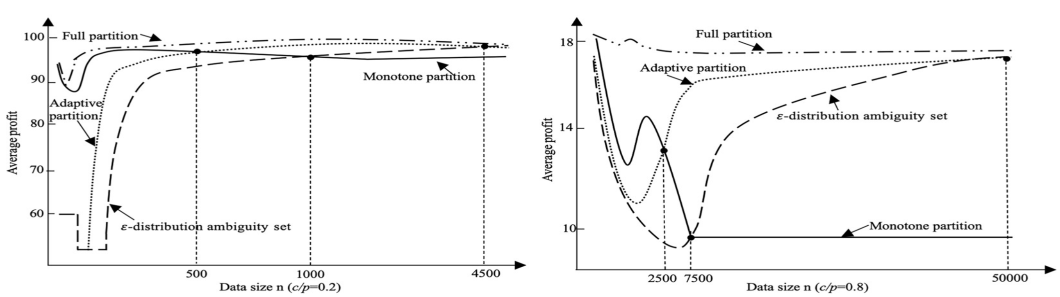
In this experiment, we randomly draw observations from an arbitrary distribution with a multi-modal density curve. First, we compare the applications of point and uniform convergence of non-parametric estimators in adaptively forming partitions by our approach. The unbiased monotone and convex partitions for a given confidence level can be obtained by Propositions 7 and 8. We choose confidence level $1-\omega \in \{0.8, 0.85, 0.9, 0.95\}$. Let $N(U_i^P), i = 1, 2$ be the data size at which the "unbiased" range U_i^P occupies a certain target proportion of the support under point convergence. Let $N(U_i), i = 1, 2$ be the counterpart for U_i under uniform convergence. The results when half width $\lambda = 0.75, 1, 1.25$ with target proportions 50% and 75% are presented in Table 4.

Table 4 Threshold levels for data size to guarantee monotonicity and convexity

ω	Target proportion for U_1 or U_1^P				Target proportion for U_2 or U_2^P			
	≥50%		≥75%		≥50%		≥75%	
	λ=1				λ=1			
	$N(U_1^P)$	$N(U_1)$	$N(U_1^P)$	$N(U_1)$	$N(U_2^P)$	$N(U_2)$	$N(U_2^P)$	$N(U_2)$
0.05	600	11,000	4,100	29,000	800	17,000	13,400	131,000
0.10	200	8,900	3,900	28,100	800	14,000	10,200	11,3000
0.15	200	7,900	3,000	28,100	600	12,000	8,100	80,000
0.2	200	7,500	2,700	28,100	200	11,000	7,500	7,6000
	λ=0.75				λ=0.75			
0.05	500	24,200	28,100	70,000	2,100	17,1000	41,400	730,000
0.10	500	20,500	28,100	59,100	1,600	17,1000	4,700	580,000
0.15	200	19,300	28,100	50,700	1,600	16,0000	4,300	500,000
0.2	200	16,000	28,100	42,500	1,300	14,3000	4,100	450,000
	λ=1.25				λ=1.25			
0.05	600	2,200	2,200	40,000	300	7,500	700	55,000
0.10	200	2,200	2,200	35,000	300	6,200	600	47,000
0.15	200	2,200	2,200	31,100	200	5,300	500	40,000
0.2	200	2,200	2,200	21,200	200	2,200	400	37,000

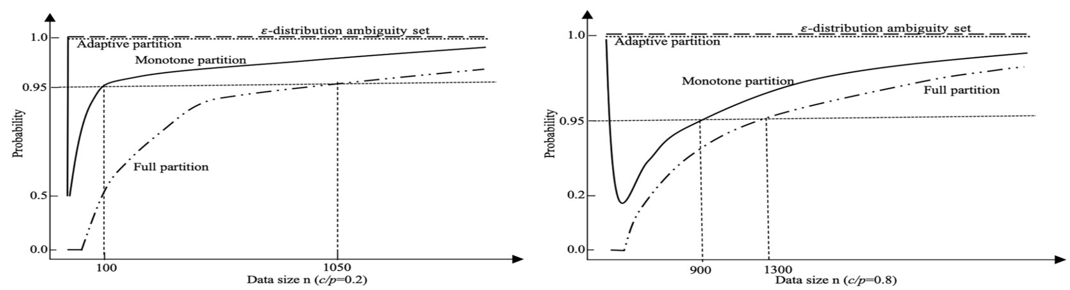
Observe that $N(U_i^P)$ is significantly smaller than $N(U_i)$, indicating the superiority of point convergence over uniform convergence in the amount of data required to guarantee the unbiasedness of the formed partitions with a prespecified confidence level. Table 4 reveals that the choice of half width λ impacts the scale of $N(U_1^P)$ relative to that of $N(U_2^P)$. $N(U_1^P) << N(U_2^P)$ when $\lambda < 1$ in most circumstances, while $N(U_1^P) > N(U_2^P)$ when $\lambda > 1$. A half width slightly above one is favoured to balance the accuracy of and confidence in guaranteeing unbiasedness, and only a small data size is needed to ensure the unbiasedness of partitions formed on 50%-75% of the support.

Figure 4 Comparison on profit performance, fixed/adaptive partitioning vs empirical distribution

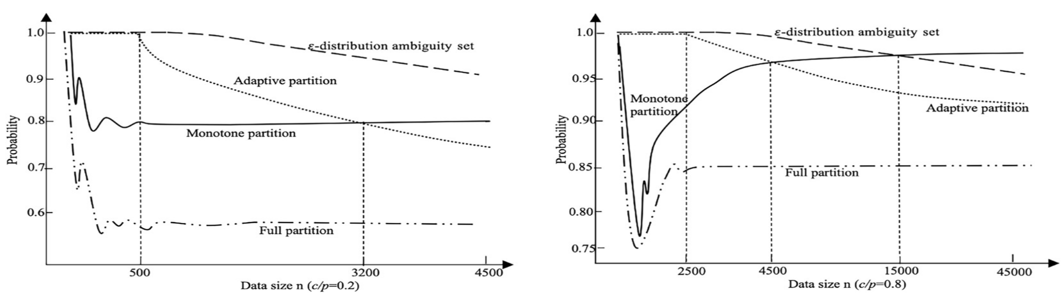


Next, using the randomly generated data to form distribution ambiguity sets and obtain stocking quantities by our approach under various partitioning methods and the ε-distribution ambiguity set, as stated in Section 4.3, we compare profit performance across the methods. Particularly, we apply fixed monotone and full partitioning and adaptive partitioning under point convergence with $\omega = 0.05$ and half width $\lambda = 1.1$. With an initial set of 50 data, after a new data is generated in a period, we use all the available data to obtain a stocking quantity by each method, and match the obtained quantity with all the data up to the period to calculate the average profit. The results with selling price $p = 10$ are shown in Figure 4. Adaptive partitioning underperforms fixed monotone partitioning when data size is small, i.e., $n < 500$ $(2,500)$ at $\frac{c}{p} = 0.2$ (0.8), and yields an average profit closer to that under fixed full partitioning as data size increases. The profit obtained by using ε-distribution ambiguity set is small with limited data input and increases as more data becomes available. At $\frac{c}{p} = 0.2$, our approach under adaptive partitioning outperforms the approach using ε-distribution ambiguity set when data size is below 4,500. At $\frac{c}{p} = 0.8$, the relevant threshold increases to 50,000. In general, this threshold increases and, therefore, our approach under adaptive partitioning is more appealing as the cost ratio increases.

Figure 5 Comparison on probability guarantee, fixed/adaptive partitioning vs empirical distribution



(a) Proportion that the average profit exceeds the worst-case expected profit



(b) Proportion that the single realized profit exceeds the worst-case expected profit

We further examine the proportion of periods when the average profit up to the period exceeds the worst-case expected profit. Figure 5.a) shows that adaptive partitioning attains a proportion close to 1, much higher than the theoretical bound established in Proposition 10. With the stocking quantity executed on a long-term basis, decision makers who are concerned about the chance that the average realized profit exceeds the worst-case expected profit may apply fixed full partitioning when data size is large. As revealed in the experiment, the probability is above 0.95 when data size exceeds 1,300 (1,050) at $\frac{c}{p} = 0.8$ (0.2). With the stocking quantity executed on a short-term basis, we examine the proportion of periods when the single-period realized profit exceeds the worst-case expected profit. Figure 5.b) shows that adaptive partitioning outperforms fixed partitioning to protect the realized profit when data size is small, i.e., $n < 3,200$ $(4,500)$ at $\frac{c}{p} = 0.2$ (0.8). From the perspective of the converging value of this measure, fixed monotone (full) partitioning attains the highest (lowest) value while adaptive partitioning attains a value in between.

Moreover, we use the same data set to compare the performance of our approach with the worst-case expected profit in $\mathcal{P}(\varepsilon)$ under the stocking quantity obtained by adaptive partitioning. Our approach generates a higher average realized profit in almost all periods and a higher single-period realized profit in more than 95% of the periods. These results are better than those when we compare realized profits with the worst-case expected profit in the constructed ambiguity set $\mathcal{P}^P$. Thus, our approach provides a high probability guarantee in yielding expected profit performance.

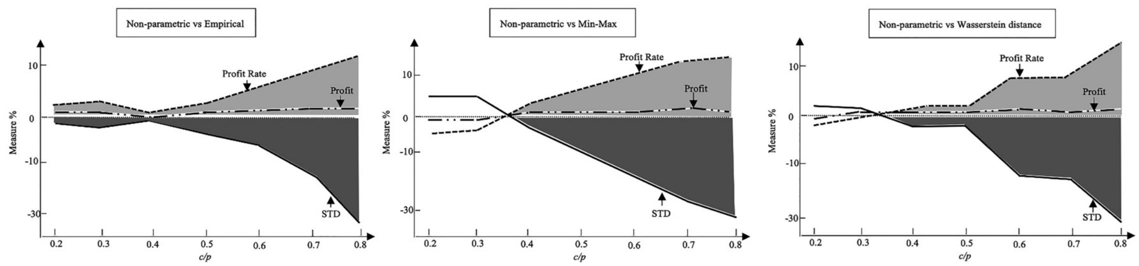
6.2. Comparison with existing robust approaches based on real data set

Kaggle.com is an online community owned by Google LLC, and it allows users to publish data sets in a web-based data-science environment. To group data conveniently and compare our approach with existing approaches for a small data size, we randomly choose a set of 500 data from Kaggle.com for item 1 in store 4 from January 1, 2013 to May 15, 2014, and denote the data set as $\{d_1, ...d_{500}\}$. The data can be retrieved from `https://www.kaggle.com/c/demand-forecasting-kernels-only/data`. As detailed information on cost and price is unavailable, we vary the service ratio from set $\{0.2, 0.3, ..., 0.8\}$ and experiment with each scenario.

The approach which uses empirical distribution directly provides a first benchmark. Gallego and Moon (1993) show that Scarf (1958)'s min-max approach, which is a known parametric approach, generates close-to-optimal stocking quantities when the random variable follows a normal distribution. The empirical distribution in our experiment closely resembles a normal distribution, motivating us to compare our approach to Scarf (1958)'s approach. We also use Wasserstein distance method in Gao and Kleywegt (2016) in comparison because it outperforms, as far as we know, many existing approaches that consider distribution distance. Let the cost function for p-order Wasserstein distance be $d^p(z_1, z_2) = |z_1 - z_2|,\ p = 1, z_1, z_2 \in [\underline{z}, \overline{z}]$. The radius of Wasserstein ball is estimated with confidence level 95%.

For the single-period newsvendor problem, we use $\{d_1, ...d_{250}\}$ to obtain an optimal stocking quantity by our approach under semi-full partitioning and that for each of the other approaches. Matching the obtained stocking quantity to every data in $\{d_{251}, ...d_{500}\}$ yields 250 single-period profits, based on which we obtain the performance in profit, profit rate (defined as $\frac{profit}{cost}$), and standard deviation of the profit (STD). The measures used in comparison are $\frac{M_{nonpara} - M_{benchmark}}{M_{benchmark}} \cdot 100\%$, where M indicates the performance, which can be profit, profit rate, or STD, and the benchmark refers to the approach with which we compare our non-parametric approach. By this definition, our approach outperforms in profit or profit rate when the corresponding measure is positive, and in STD when the corresponding measure is negative.

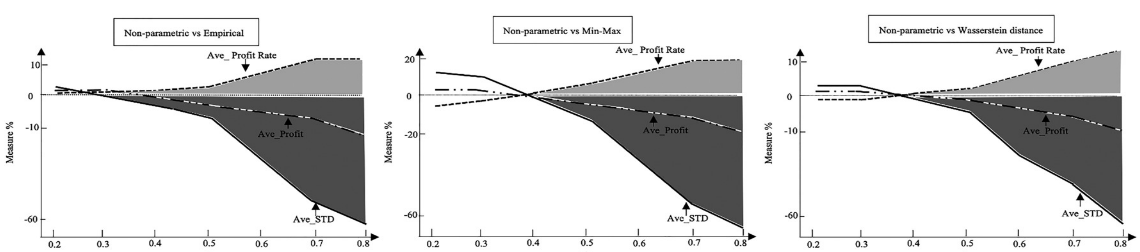
Figure 6 Results of relative performance with data set from Kaggle.com



Note. The shade area represents where our approach outperforms other approaches in profit rate or STD

Figure 6 reveals that our non-parametric approach yields no significant improvement in profit compared with the other approaches, but it can outperform in profit rate and STD. Specifically, our approach outperforms the empirical approach in the three measures. For instance, at a service ratio of 0.2 ($\frac{c}{p} = 0.8$), it yields 0.83% and 12.41% improvements in profit and profit rate, respectively, and a 31.32% reduction in STD. Our approach outperforms the min-max approach provided that the service ratio is not high ($\frac{c}{p} > 0.3$). At a service ratio of 0.3 ($\frac{c}{p} = 0.7$), it achieves 2.08% and 14.04% improvements in profit and profit rate, respectively, and a 28.20% reduction in STD. Compared with the Wasserstein distance method, our approach also performs better when the service ratio is not high ($\frac{c}{p} > 0.3$). At a service ratio of 0.2 ($\frac{c}{p} = 0.8$), on top of a 0.62% improvement in profit, it achieves a 11.80% improvement in profit rate and a 30.62% reduction in STD. Generally, the advantages of our approach in improving profit rate and reducing profit variability become more prominent as the service ratio decreases (value of $\frac{c}{p}$ increases).

Figure 7 Results of average performance for various input sizes using data set from Kaggle.com



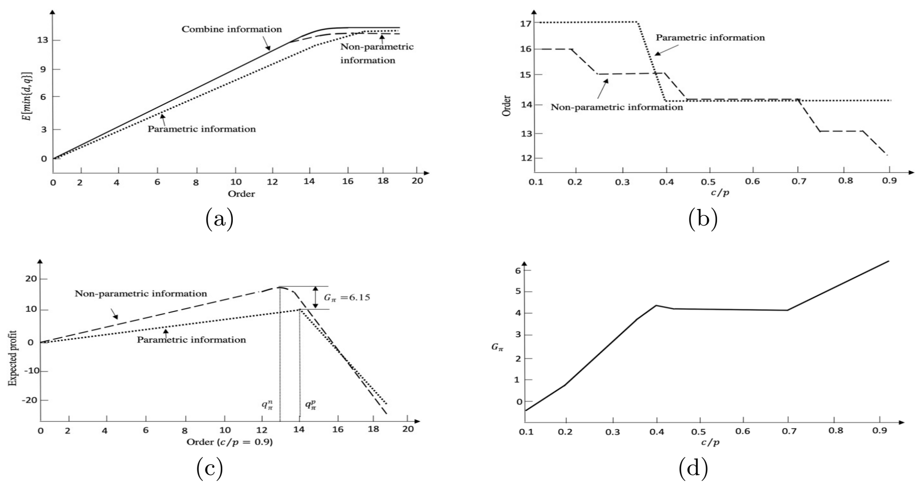
To investigate the effects of input size on relative profit performance, we apply a rolling method to obtain the optimal stocking quantities using data $\{d_1, \ldots, d_{50}\}, \{d_1, \ldots, d_{100}\}, \ldots, \{d_1, \ldots, d_{450}\}$ from the same data set, with input sizes $50, 100, ..., 450$. In each scenario, the obtained stocking quantity is matched to the next 50 unused data to calculate profits. Then, we aggregate the results in all scenarios to obtain the average performance across various sizes of data input. Figure 7 reveals a similar pattern as that in Figure 6 on the relative performance of our approach with respect to the other three approaches. One exception occurs to the average profit performance when the service ratio is low ($\frac{c}{p} > 0.4$), in which case, our approach underperforms the other approaches. Nevertheless, it achieves significant improvements in average profit rate and STD. For instance, at a service ratio of 0.3 ($\frac{c}{p} = 0.7$), compared with the Wasserstein distance method, despite a 7.79% reduction in average profit, it achieves a 8.66% gain in average profit rate and a 44.34% reduction in average STD. Thus, our approach provides an effective instrument to managers who value high return on investment and stability in profit generation.

6.3. Value of non-parametric information

A stream of existing works use parametric information like mean and variance to form the distribution ambiguity set. By contrast, our approach uses non-parametric characteristics. In this experiment, we combine the two types of information to construct distribution ambiguity sets by full partitioning, and obtain optimal stocking quantities and worst-case expected profits to perform a comparative examination. We construct a mass function for a discrete random demand on $[12, 17]$ and let candidate stocking quantities be $Q = \{0, 1, \ldots, 19\}$. The worst-case expected sales quantities for given stocking quantities are shown in Figure 8.a). The three curves correspond to when only non-parametric, only parametric, and both types of information are used. Non-parametric information generates more sales when the stocking quantity is low (< 16) but less sales otherwise than parametric information. At low stocking quantities, adding parametric to non-parametric information is inconsequential to sales generation, but adding non-parametric to parametric information increases sales. It is more prominent as stocking quantity increases. At high stocking quantities (> 12), compared with using parametric or non-parametric information alone, combining the two types of information yields the most sales. The optimal stocking quantities obtained by non-parametric and parametric information (q_π^n,q_π^p) are shown in Figure 8.b). Observe that the two types of information yield the same stocking quantity when the service ratio is intermediate ($\frac{c}{p} \in [0.45, 0.7]$). Parametric information yields a larger stocking quantity in most other circumstances. With the obtained quantities, Figure 8.c) shows that, at a low service ratio of 0.1 ($\frac{c}{p} = 0.9$), non-parametric information yields a higher expected profit with an improvement of $G_\pi = 6.15$. Figure 8.d) further shows that non-parametric information outperforms its parametric counterpart in yielding the expected profit provided the service ratio is not too high, and the advantage of non-parametric information is more prominent as the service ratio decreases.

Finally, we compare our approach with the second-order statistics approach by Natarajan et al. (2018), which is a parametric approach using the first moment on the entire support and the second moment in each partition. Their approach defines exogenous partitions such as equal-length

Figure 8 Value of non-parametric information



partitions and partitions with decreasing length. For comparison purpose, we generate the same number of partitions by the two approaches. To experiment with the 2-partition method, we use an asymmetric distribution for a discrete variable on $[12,17]$ and generate full partitions $A1$ and $A2$ by our approach but partitions $B1 = [12,e)$ and $B2 = [e,17]$, where e is the demand mean, by Natarajan et al. (2018)'s approach, with $|A1| \neq |B1|$ and $|A2| \neq |B2|$. We engage in a similar procedure to experiment with the 3-partition method.

Figure 9 Value of non-parametric partitioning

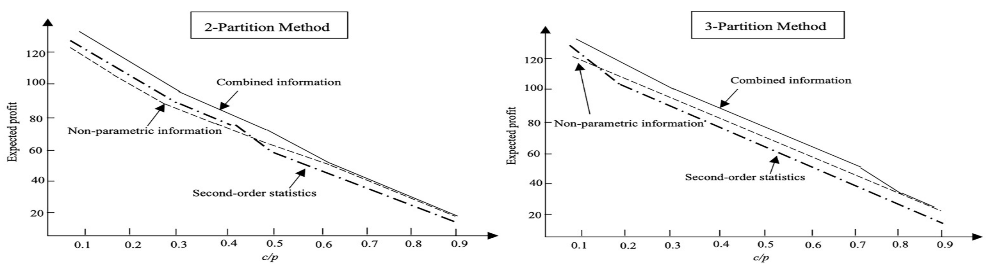


Figure 9 reveals that our non-parametric approach outperforms Natarajan et al. (2018)'s second-order statistics approach unless the service ratio is high. Specifically, our approach outperforms their 2-partition method when the service ratio is below 0.6 ($\frac{c}{p} \geq 0.4$), and it outperforms their 3-partition method when the service ratio is below 0.8 ($\frac{c}{p} \geq 0.2$). Moreover, it is valuable to include parametric information – including the first and second moments – to the non-parametric characteristics in the partitions to construct the distribution ambiguity set and build the protection curve for robust decision making. This improves the profit performance relative to that by Natarajan et al. (2018)'s second-order statistics approach when the service ratio is high, in which case, our approach with only non-parametric information performs less well.

7. Concluding Remarks

We have proposed a data-driven non-parametric approach and applied it to well-known newsvendor problems. In decision settings that have a random demand with an unknown distribution, prior efforts were dedicated to estimating the distribution by available (often times censored) data. To avoid the biasedness that may occur, we utilize the non-parametric characteristics of the distribution to build a protection curve and use it to make decisions. Our approach has two advantages over traditional approaches using empirical distribution. One is that the constructed protection curve provides a safe approximation to the true density curve and the realized profits exhibit low variability. The other is that, even when data size is small, our approach can adaptively update the protection curve by data input with a given confidence level in the unbiasedness of the formed partitions to have the desired characteristics. Sections 4.3 and 6.1 provide theoretical analysis and experimental evidence for choosing between protection curve and empirical distribution to obtain robust policies.

Our approach can outperform existing approaches to significantly improve profit rate and stabilize profit generation. Using non-parametric information is more effective than parametric information in yielding expected profit when the service requirement is not too high. Moreover, it has advantages in worst-case performance guarantee and convergence ratio. In reality, data is typically censored. This causes parametric statistics – including sample mean and sample variance – to be biased, leading to a loss of worst-case guarantee of the expected profit. However, the protection curve built with censored data, under a proper partitioning method for a given confidence level of unbiasedness, still protects the true density curve. To see this, let X and Y be, respectively, the censored and true demand. For a given stocking quantity q, $X = \min\{q, Y\}$, so that $\mathbb{P}\{X \leq d\} \leq \mathbb{P}\{Y \leq d\}, \forall d$. Hence, Y is first-order stochastically dominated by X. A protection curve $c(z)$ that satisfies $\int_y^{\overline{z}} [f_X(z) - c(z)]\, dz \geq 0, \forall y \in [\underline{z}, \overline{z}]$ must also satisfy $\int_y^{\overline{z}} [f_Y(z) - c(z)]\, dz \geq 0$. Therefore, our non-parametric approach, when applied to censored data, can still yield a lower bound for the expected profit under the true distribution.

Our approach also boasts a faster convergence ratio than parametric approaches provided that the true distribution is not over-stable. Taking the sample mean as an example for parameteric information. According to the central limit theorem, the sample mean $\frac{\sum_{i=1}^{n} d_i}{n}$ asymptotically follows normal distribution with standard deviation $\frac{\sigma}{\sqrt{n}}$, where σ is the standard deviation of the true distribution. We showed in Section 3.3.2 that the empirical first-order derivative asymptotically follows normal distribution with standard deviation $\hat{\sigma}_1 \leq \frac{1}{(2\lambda)^2\sqrt{n}}$, where λ is the half width. This implies that the empirical first-order derivative converges faster than the sample mean if the half width is large ($\lambda \geq 1$) and the true variance is not too small ($\sigma > 1/4$). Moreover, we reveal, theoretically in Section 5.1 and experimentally in Section 6.3, that our proposed approach provides a means of combining non-parametric and parametric information in a robust optimization framework.

Acknowledgments

The authors are grateful to Professor Brian Tomlin, an associate editor, and two anonymous referees for their comments and guidance that have significantly improved the presentation and content of the paper. The research of Liang Xu is supported in part by the National Natural Science Foundation of China under Grants #U1811462 and #71971177. The research of Li Jiang is supported in part by the National Natural Science Foundation of China under Grants #71772157 and #71803211.

References

Bayraksan G, Love DK (2015) Data-driven stochastic programming using Phi-divergences. *The Operations Research Revolution*, 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.

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

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

Gao R, Kleywegt AJ (2016) Distributionally robust stochastic optimization with Wasserstein distance. *arXiv preprint arXiv:1604.02199* .

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.

Goh J, Sim M (2010) Distributionally robust optimization and its tractable approximations. *Operations research* 58(4-part-1):902–917.

Hanasusanto GA, Kuhn D, Wallace SW, Zymler S (2015) Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming* 152(1-2):1–32.

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

Jiang R, Guan Y (2016) Data-driven chance constrained stochastic program. *Mathematical Programming* 158(1-2):291–327.

Lariviere MA, Porteus EL (1999) Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science* 45(3):346–363.

Levi R, Perakis G, Uichanco J (2011) Regret optimization for stochastic inventory models with spread information. *Submitted to Operations Research, revision requested* .

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.

Mohajerin Esfahani P, Kuhn D (2018) Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 171(1-2):115–166.

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.

Popescu I, Wu Y (2007) Dynamic pricing strategies with reference effects. *Operations research* 55(3):413–429.

Rosenblatt M (1956) Remarks on Some Nonparametric Estimates of a Density Function. *Annals of Mathematical Statistics* 27(3):832–837.

Rubner Y, Tomasi C, Guibas LJ (2000) The earth mover's distance as a metric for image retrieval. *International journal of computer vision* 40(2):99–121.

Ruppert D, Cline DB (1994) Bias reduction in kernel density estimation by smoothed empirical transformations. *The Annals of Statistics* 185–210.

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

Sun H, Xu H (2016) Convergence analysis for distributionally robust optimization and equilibrium problems. *Mathematics of Operations Research* 41(2):377–401.

Wang Z, Glynn PW, Ye Y (2016) Likelihood robust optimization for data-driven problems. *Computational Management Science* 13(2):241–261.

Waterman M, Whiteman D (1978) Estimation of probability densities by empirical density functions. *International Journal of Mathematical Education in Science and Technology* 9(2):127–137.

Yue J, Chen B, Wang MC (2006) Expected value of distribution information for the newsvendor problem. *Operations research* 54(6):1128–1136.

Zhao C, Guan Y (2018) Data-driven risk-averse stochastic optimization with Wasserstein metric. *Operations Research Letters* 46(2):262–267.

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

Zuluaga LF, Peña J, Du D (2009) Third-order extensions of Lo's semiparametric bound for European call options. *European Journal of Operational Research* 198(2):557–570.

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