mathematics

Article

A Deficiency of the Predict-Then-Optimize Framework: Decreased Decision Quality with Increased Data Size

Shuaian Wang and Xuecheng Tian *

Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Hong Kong 999077, China; hans.wang@polyu.edu.hk
* Correspondence: xuecheng-simon.tian@connect.polyu.hk

Abstract: This paper presents an analysis of the decision quality of the predict-then-optimize (PO) framework, an extensively used prescriptive analytics framework in uncertain optimization problems. Our primary aim is to investigate whether an increase in data size invariably leads to better decisions within the PO framework. We focus our analysis on two contextual stochastic optimization problems—one with a non-linear objective function and the other with a linear objective function—under the PO framework. The novelty of our work lies in uncovering a previously unknown relationship: the decision quality can deteriorate with increasing data size in the non-linear case and exhibit non-monotonic behavior in the linear case. These findings highlight a potential pitfall of the PO framework and constitute our main contribution to the field, offering invaluable insights for both researchers and practitioners.

Keywords: data-driven optimization; prescriptive analytics; predict-then-optimize; limited data

MSC: 90-10

Citation: Wang, S.; Tian, X. A Deficiency of the Predict-Then-Optimize Framework: Decreased Decision Quality with Increased Data Size. *Mathematics* **2023**, *11*, 3359. https://doi.org/10.3390/math11153359

Academic Editor: Bo Wang

Received: 30 June 2023
Revised: 24 July 2023
Accepted: 31 July 2023
Published: 31 July 2023

1. Introduction

Uncertainty is pervasive in practical optimization problems, such as routing problems with uncertain travel times and revenue management problems with unpredictable future demand [1–4]. Various frameworks have been developed to model and solve optimization problems involving uncertainty. Bertsimas and Koduri [5] categorize these frameworks into two primary groups based on their reliance on data as a primitive.

The first category includes stochastic programming [6] and robust programming [7,8] literature, which do not treat data as a primitive. These approaches generally predefine distributions for uncertain parameters without using real data. However, it is often challenging for decision makers to know the true distributions of uncertain parameters [9]. In contrast, due to advancements in internet technologies that enable the collection and storage of vast amounts of data, the second category of frameworks—using data as a primitive to characterize uncertainty—has emerged.

This second category can be further subdivided based on the type of data used: historical data of the uncertain parameters themselves or auxiliary data (contextual covariates) that can help predict uncertain parameters. The first subcategory includes frameworks that exclusively use historical data to approximate scenarios or distributions of uncertain parameters, without considering potentially useful auxiliary data. Examples include the sample average approximation (SAA) method [10] and the data-driven distributionally robust optimization framework [11,12]. The second subcategory utilizes various machine learning (ML) techniques to predict uncertain parameters, drawing not only on historical data but also on related auxiliary data. This integration of ML methods and mathematical optimization models is referred to as prescriptive analytics, with the sequential predict-then-optimize (PO) framework being the most widely used approach.

The PO framework initially employs ML models to predict the point values of uncertain parameters using available auxiliary data. It then plugs these point predictions into the downstream optimization problems, thereby transforming uncertain problems into more manageable deterministic problems. In the context of this PO framework, this paper examines the following question:

Does an increase in historical data lead to higher-quality decisions in the PO framework?

Our interest lies in analyzing decision quality across varying data sizes, an area that holds both practical and theoretical significance [13]. Practically, while historical data might be abundant, "relevant" data may be limited due to market heterogeneity. Theoretically, examining the impact of data size on data-driven decisions can provide insights into the robust value of data [13].

Related Literature and Contributions

Our work contributes to two primary streams of literature. First, it relates to contextual stochastic optimization problems, where decision makers observe previous samples of the uncertain parameter, along with features providing additional information about the uncertainty [3,5,14–18]. Besides the widely-used PO framework, several advanced frameworks have emerged, which prescribe decisions directly from data. They include the smart predict-then-optimize framework [16,19], the weighted SAA (wSAA) framework [17,18], the empirical risk minimization framework [14,17], and the kernel optimization framework [5,14,17,20]. Theoretically, these works have demonstrated the asymptotic optimality of these frameworks, i.e., whether the data-driven solutions can converge to the full-information optimal solution as data size approaches infinity [9]. Additionally, some have also shown the generalization bounds, i.e., how well the performance of a prediction model, fitted on finite training data, generalizes from a sample [9,21]. Despite these advancements, there remains a gap in the literature: the impact of data size on the performance of the PO framework is yet to be comprehensively examined. Our study intends to address this gap. For detailed reviews of corresponding prescriptive analytics frameworks for contextual stochastic optimization problems, we refer to Qi and Shen [9] and Tian et al. [19].

Additionally, our work contributes to a class of literature that investigates the influence of data size on data-driven decisions. However, much of this literature focuses solely on previous samples of uncertainty, without considering contextual features. Our main idea is closely related to a recent study by Besbes and Mouchtaki [13], which characterizes the full spectrum of performances achievable by SAA across data sizes. Surprisingly, their analysis revealed that the worst-case relative regret is not monotonically related to the number of samples available, suggesting that an influx of information can potentially degrade decision quality. Other studies concerning bounds on probabilistic guarantees of the relative regret of the SAA include Levi et al. [22] and Cheung and Simchi-Levi [23]. Recently, Wang and Tian [24] investigate the decision performance of the wSAA framework across data sizes. In contrast, our work examines the decision performance of the PO framework across data sizes. To provide a more comprehensive overview, we summarize these research backgrounds and their key findings in Table 1.

To achieve our research aim, we show two contextual stochastic optimization problems, one with a linear objective function and the other with a non-linear objective function. Our analysis reveals the following intriguing finding:

The decision quality of the PO framework can strictly decrease with the increase in data size for some data sizes in cases with linear objective functions, and for all data sizes in cases with non-linear objective functions.

Hence, the main contribution of our work is to reveal that the PO framework may not always appropriately accumulate information, sometimes even "destroying" valuable information in the process. This paper presents this observation through two straightforward examples, laying the groundwork for future research to potentially achieve the optimal decision performance of the PO framework as a function of data size.

Table 1. Summary of research backgounds and findings in the literature.

Research Area	Key Findings	Gap in Literature	Related Works
Contextual stochastic optimization problems	Various frameworks developed that utilize data to prescribe decisions; asymptotic optimality and generalization bounds of these frameworks demonstrated	Lack of investigation into the impact of data size on the performance of the PO framework	[3,5,14–18,21]
Data size's influence on data-driven decisions	Performance of SAA across data sizes studied; non-monotonic relationship between worst-case relative regret and number of samples revealed	Focus is primarily on previous samples of uncertainty, with little consideration for contextual features	[13,22,23]

The remainder of this paper is organized as follows. In Section 2, we define a general problem setting for the PO framework. In Section 3, we present our findings through a contextual stochastic optimization problem with a linear objective, and in Section 4, we demonstrate our findings through a contextual stochastic optimization problem with a non-linear objective. Section 5 concludes and outlines future research directions.

2. Problem Setting

Following the notation of Bertsimas and Kallus [18], we consider a stochastic optimization problem with a given cost function $c(\boldsymbol{y};\boldsymbol{z})$, where $\boldsymbol{y} \in \mathcal{Y} \subset \mathbb{R}^{d_y}$ represents the random variable that affects the value of the cost function, and $\boldsymbol{z} \in \mathcal{Z} \subset \mathbb{R}^{d_z}$ represents the decision variable, characterized by a set of constraints. Let $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^{d_x}$ denote the contextual information that is related to the distribution of $\boldsymbol{y}$. Suppose we have a new observation of auxiliary data $\boldsymbol{x}_0 \in \mathcal{X} \subset \mathbb{R}^{d_x}$. The contextual stochastic optimization problem can be mathematically formulated as follows:

$$\boldsymbol{z}^*(\boldsymbol{x}_0) \in \arg\min_{\boldsymbol{z}\in\mathcal{Z}} \mathbb{E}_{\boldsymbol{y}}\left[c(\boldsymbol{y};\boldsymbol{z})\middle|\boldsymbol{x}=\boldsymbol{x}_0\right]. \quad (1)$$

A historical dataset $\{(\boldsymbol{x}_i,\boldsymbol{y}_i)\}_{i=1}^{N}$ is available to solve the optimization problem (1), where $\boldsymbol{x}_i$ are the historical auxiliary data, and $\boldsymbol{y}_i$ are the historical realizations of $\boldsymbol{y}$. Given this dataset, the PO framework first builds an ML model to predict the value of $\boldsymbol{y}$ based on the new observation $\boldsymbol{x}_0$, denoted by $\hat{\boldsymbol{y}}_0$ [16,18]. It then plugs $\hat{\boldsymbol{y}}_0$ into the optimization problem (1) to solve

$$\min_{\boldsymbol{z}\in\mathcal{Z}} c(\hat{\boldsymbol{y}}_0;\boldsymbol{z}) \quad (2)$$

to obtain the approximate solution. In this paper, we adopt the most commonly used mean squared error (MSE) loss to train the ML model, defined as follows:

$$L_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i\right\|_2^2. \quad (3)$$

3. Properties of the PO Framework in the Non-Linear Case

In this section, we consider a contextual stochastic optimization problem with a non-linear objective. We use this example to demonstrate that the PO framework's prediction of unknown parameters can lead to a strictly monotonic decrease in decision quality as the amount of data increases.

We consider an optimization problem involving a decision on whether to allocate production resources considering the uncertainty of demand. The decision variable is denoted by $z \in \{0,1\}$, where $z = 1$ denotes the allocation of resources, and 0 otherwise. We assume that the resource capacity is infinite. The cost associated with resource allocation is

1 unit. We denote the uncertain demand as $\tilde{y}$. If at least one product is sold (i.e., $\tilde{y} \geq 1$ and $z = 1$), the company obtains a reward of 10,000 units from its parent company. However, no reward is provided if no items are sold (i.e., $\tilde{y} < 1$). The objective is to maximize the profit. Now, we have a new observation $\boldsymbol{x}_0$ that is related to the distribution of $\tilde{y}$. Given this, the contextual stochastic optimization problem can be formulated mathematically as follows:

$$\max_{z \in \{0,1\}} \mathbb{E}\big[z[\mathbb{I}(\tilde{y} \geq 1) \times 10000 - 1]|\boldsymbol{x}_0\big] = \max_{z \in \{0,1\}} z\mathbb{E}\big[\mathbb{I}(\tilde{y} \geq 1) \times 10000 - 1|\boldsymbol{x}_0\big]. \tag{4}$$

Here, $\mathbb{I}(\tilde{y} \geq 1)$ is the indicator function that equals 1 when $\tilde{y} \geq 1$ is satisfied, and 0 otherwise. The expectation, $\mathbb{E}$, is taken over the distribution of $\tilde{y}$ conditioned on $\boldsymbol{x}_0$. The objective function represents the expected profit from the decision to allocate resources, taking into account the reward and the cost of resources.

Assume that we have access to a historical dataset $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$, where the features of all samples are identical, $\boldsymbol{x}_1 = \ldots = \boldsymbol{x}_N$. The historical targets y_i $(i \in \{1, \ldots, N\})$ each have a 50% chance of being 0 or 1. We further assume that $\boldsymbol{x}_0 = \boldsymbol{x}_1 = \ldots = \boldsymbol{x}_N$. Hence, the actual conditional distribution of $\tilde{y}$ given $\boldsymbol{x}_0$ follows a Bernoulli distribution, with its probability mass function expressed as follows:

$$\mathrm{P}(\tilde{y} = k|\boldsymbol{x}_0) = \begin{cases} 0.5, & \text{if } k = 0, \\ 0.5, & \text{if } k = 1. \end{cases} \tag{5}$$

If we knew this actual distribution of $\tilde{y}$, we could solve the following problem:

$$\begin{aligned} &\max_{z \in \{0,1\}} z\big[\mathrm{P}(\tilde{y} \geq 1) \times (10000 - 1) + \mathrm{P}(\tilde{y} < 1) \times (0 - 1)\big] \\ =& \max_{z \in \{0,1\}} z\big[0.5 \times (10000 - 1) + 0.5 \times (0 - 1)\big]. \end{aligned} \tag{6}$$

The solution is to allocate resources ($z = 1$), yielding an expected profit of 4999 units. However, in practice, it is difficult to know the actual distribution of $\tilde{y}$. We apply the PO framework: we use historical data to build an ML model that predicts the expected value of $\tilde{y}$, denoted by $\hat{y}$, given the new observation $\boldsymbol{x}_0$, and then use the prediction $\hat{y}$ to make the decision z to maximize the expected profit by solving the following approximate problem:

$$\max_{z \in \{0,1\}} \mathbb{E}\big[z[\mathbb{I}(\hat{y} \geq 1) \times 10000 - 1]|\boldsymbol{x}_0\big] = \max_{z \in \{0,1\}} z\mathbb{E}\big[\mathbb{I}(\hat{y} \geq 1) \times 10000 - 1|\boldsymbol{x}_0\big], \tag{7}$$

where resources will be allocated ($z = 1$) if and only if the predicted $\hat{y} \geq 1$.

Therefore, for $n \in \{1, ..., N\}$ historical data samples, the model will decide to allocate resources if and only if all historical sample targets are equal to 1, i.e., $y_1 = ... = y_n = 1$ (recall that we use MSE to train the model, and the historical targets follow a Bernoulli distribution alternating between 0 and 1). This situation has a probability of $(\frac{1}{2})^n$. In such a case, the model predicts $\hat{y} = 1$, and the decision problem yields $z = 1$. The expected profit is calculated as $(10000 - 1) \times 0.5 + (-1) \times 0.5 = 4999$. In all other scenarios in which $\hat{y} < 1$, the decision problem is solved as

$$\max_{z \in \{0,1\}} z[\mathbb{I}(\hat{y} \geq 1) \times 10000 - 1] = \max_{z \in \{0,1\}} z[0 - 1], \tag{8}$$

leading to $z = 0$, meaning that no resources should be allocated and the expected profit is 0. Thus, the total expected profit under the PO framework with n random data samples is $4999 \times (\frac{1}{2})^n$. In comparison to the optimal expected profit under full information, which stands at 4999 units, the PO framework, even with the acquisition of more data, paradoxically results in a strictly decreased expected profit, computed as $4999 \times (\frac{1}{2})^n$.

To verify this finding, we further conduct a simulation analysis, the results of which are shown in Figure 1. In this simulation, different data sizes are tested to see how the decision quality changes with increasing data size in the non-linear case. In our analysis,

Figure 1 suggests a strictly monotonic relationship between the decision quality and the data size. As a result, we can conclude that the decision quality of the PO framework can strictly decrease as the data size increases. This finding reveals a significant deficiency of the PO framework when dealing with certain types of decision problems under uncertainty.

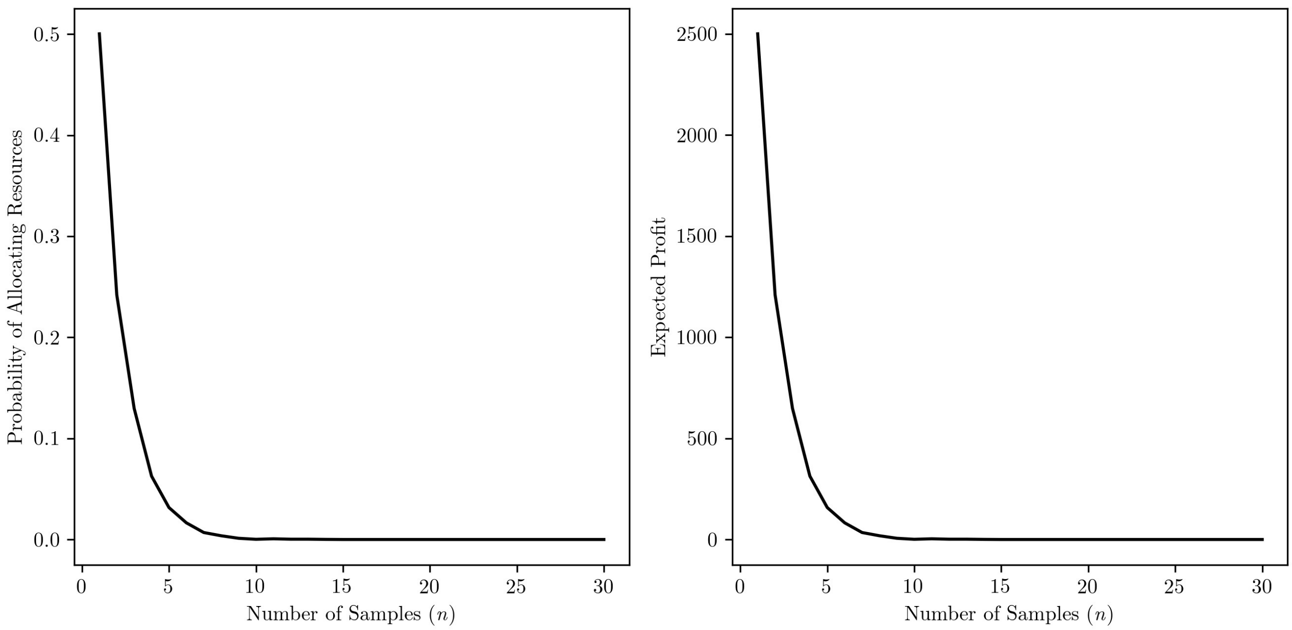

Figure 1. The decision quality across data sizes for the non-linear case.

In the prior example, we expose a scenario where, for a given decision variable z, the objective function is discontinuous of the unknown parameter $\tilde{y}$. For instance, this might occur when $z = 1$. An essential question arising from this situation is whether a similar decrease in prediction quality could occur even when the objective function exhibits Lipschitz continuity for all $z \in \{0, 1\}$. Let us consider a modified scenario. Assume that the parent company decides to alter its reward mechanism. Under this new regime, no reward is granted when the sales volume falls within the range of 0 to $1 - \epsilon$, where $0 < \epsilon < 1$. When the sales volume lies between $1 - \epsilon$ and 1, the reward linearly escalates from 0 to 10,000 units. For a sales volume that exceeds the 1 unit mark, the reward remains at 10,000 units. This modifies the reward function into a Lipschitz continuous function of $\tilde{y}$. Mathematically, we can represent this new reward function as follows:

$$r(\tilde{y}) = \begin{cases} 0, & \text{if } 0 \le \tilde{y} < 1 - \epsilon, \\ 10000 \times \frac{\tilde{y} - (1-\epsilon)}{\epsilon}, & \text{if } 1 - \epsilon \le \tilde{y} < 1, \\ 10000, & \text{if } \tilde{y} \ge 1. \end{cases} \quad (9)$$

We find that for a sufficiently small ϵ approaching zero, the predictive performance of the PO framework could still exhibit a strictly decreasing trend as the amount of data surges. To formalize this, we can construct the optimization problem as a function of the reward and the number of historical data samples n, shown as follows:

$$\max_{z \in \{0,1\}} \mathbb{E}\big[z[r(\hat{y}) - 1] | \boldsymbol{x}_0\big], \quad (10)$$

where $\hat{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$, $n \in \{1, ..., N\}$. When $\hat{y} < 1 - \frac{9999}{10000}\epsilon \approx 1 - \epsilon$, we will choose $z = 0$ to maximize the expected profit. However, as n grows, $\hat{y}$ will converge to 0.5 and always be less than $1 - \epsilon$, given a sufficiently small ϵ. Thus, with more data, the decision z will always be 0, and the expected profit will be equivalent to $4999 \times (\frac{1}{2})^n$ for a large n.

This outcome uncovers a critical flaw of the PO framework when dealing with specific types of decision problems under uncertainty. It persists even when the objective function is a Lipschitz continuous function of $\tilde{y}$. It reinforces our prudence towards the application of the PO framework in decision problems under uncertainty.

4. Properties of the PO Framework in the Linear Case

In this section, we consider a contextual stochastic optimization problem with a linear objective. We use this example to illustrate how the PO framework's prediction of the expected value of the unknown parameter could lead to a non-monotonic decrease in decision quality as the amount of data increases.

We consider a resource allocation problem to be where a single resource can be allocated to two products. The first product earns a unit profit of 1, while the unit profit for the second product is uncertain, denoted by $\tilde{y}$. We denote by $z \in \{0,1\}$ the decision variable, where $z = 1$ represents allocating all the resource to product 1, and 0 otherwise. The objective is to maximize the expected total profit. Now, we have a new observation $\boldsymbol{x}_0$ that is related to the distribution of $\tilde{y}$. Given this, the contextual stochastic optimization problem can be formulated mathematically as follows:

$$\max_{z \in \{0,1\}} \mathbb{E}\big[z \times 1 + (1-z) \times \tilde{y} | \boldsymbol{x}_0\big]. \tag{11}$$

Assume that we have access to a historical dataset $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$, where the features of all samples are identical, $\boldsymbol{x}_1 = ... = \boldsymbol{x}_N$. The historical targets y_i $(i \in \{1, ..., N\})$ each have a 50% chance of being 0 or 2.00000001. We further assume that $\boldsymbol{x}_0 = \boldsymbol{x}_1 = ... = \boldsymbol{x}_N$. Hence, the actual conditional distribution of $\tilde{y}$ given $\boldsymbol{x}_0$ is

$$\mathrm{P}(\tilde{y} = k | \boldsymbol{x}_0) = \begin{cases} 0.5, & \text{if } k = 0, \\ 0.5, & \text{if } k = 2.00000001. \end{cases} \tag{12}$$

If we knew this actual distribution of $\tilde{y}$, we could solve the following problem:

$$\begin{aligned} &\max_{z \in \{0,1\}} \Big\{ z \times 1 + (1-z) \times \big[\mathrm{P}(\tilde{y} = 0) \times 0 + \mathrm{P}(\tilde{y} = 2.00000001) \times 2.00000001\big] \Big\} \\ = &\max_{z \in \{0,1\}} \Big\{ z \times 1 + (1-z) \times \big[0.5 \times 0 + 0.5 \times 2.00000001\big] \Big\}. \end{aligned} \tag{13}$$

The optimal full-information solution allocates the single resource to product 2 ($z = 0$), yielding an optimal expected profit of 1.000000005.

However, as we cannot ascertain the actual distribution of $\tilde{y}$, we apply the PO framework. In the context of the PO framework and observing the objective function (11), we find that the single resource will be allocated to product 1 with varying probabilities depending on the data size, which is analyzed in the following.

Let n denote the number of historical data samples available. We note that according to the strong law of large numbers, we can achieve the optimal full-information solution with infinite data. Hence, we have the following property: for a contextual stochastic optimization problem with a linear objective, the prediction of the uncertain parameter does not necessarily lead to a strictly monotonic decrease in decision quality as data size grows; therefore, it is not guaranteed that decision quality will worsen with each additional data point.

In fact, our subsequent analysis reveals that in the linear case, the PO framework's prediction can lead to a non-monotonic decrease in decision quality as data size increases.

1. If $n = 1$, there is a 50% probability of allocating the resource to product 1, yielding an expected profit of 1.0000000025. In detail, if the target value in the observed sample is 0, the predicted unit profit for product 2 will be 0, and the probability of allocating the resource to product 1 rises to 100%. Conversely, if the target value is

2.000000001, the predicted unit profit for product 2 is 2.000000001, and the probability of allocating the resource to product 1 drops to 0. Thus, for $n = 1$, the probability of allocating the resource to product 1 is $1 \times 0.5 + 0 \times 0.5 = 0.5$, and the expected profit is $0.5 \times 1 + 0.5 \times (0.5 \times 0 + 0.5 \times 2.00000001) = 1.0000000025$.

2. If $n = 2$, if both observed target values are 0, the predicted unit profit for product 2 remains at 0, and the probability of allocating the resource to product 1 remains at 100%. If one target value is 0 and the other is 2.000000001, the predicted unit profit for product 2 averages out to 1, and the probability of allocating the resource to product 1 drops to 0. If both observed target values are 2.000000001, the predicted unit profit for product 2 is 2.000000001, and the probability of allocating the resource to product 1 stays at 0. Therefore, when $n = 2$, the probability of allocating the resource to product 1 is $1 \times 0.25 + 0 \times 0.5 + 0 \times 0.25 = 0.25$, the expected profit is $0.25 \times 1 + 0.75 \times (0.5 \times 0 + 0.5 \times 2.00000001) = 1.00000000375$.
3. By analogy, if $n = 3$, the probability of allocating the resource to product 1 is 50%, yielding an expected profit of 1.0000000025.
4. If $n = 4$, the probability of allocating the resource to product 1 is $5/16$, yielding an expected profit of 1.0000000034375.
5. If $n = 5$, the probability of allocating the resources to product 1 is 50%, yielding an expected profit of 1.0000000025.

This pattern of fluctuation continues as the data size increases, showing a non-monotonic relationship between the decision quality and the data size. We further conduct a simulation analysis on the decision quality across data sizes in this example, where the final results are shown in Figure 2. In this simulation, different data sizes are tested to see how the decision quality changes with an increasing data size. Surprisingly, the results suggest a non-monotonic relationship between the decision quality and the data size. In other words, contrary to the common expectation that more data would generally improve the decision quality, the decision quality in the PO framework can actually decrease as the data size increases. This counterintuitive result highlights a potential pitfall of the PO framework.

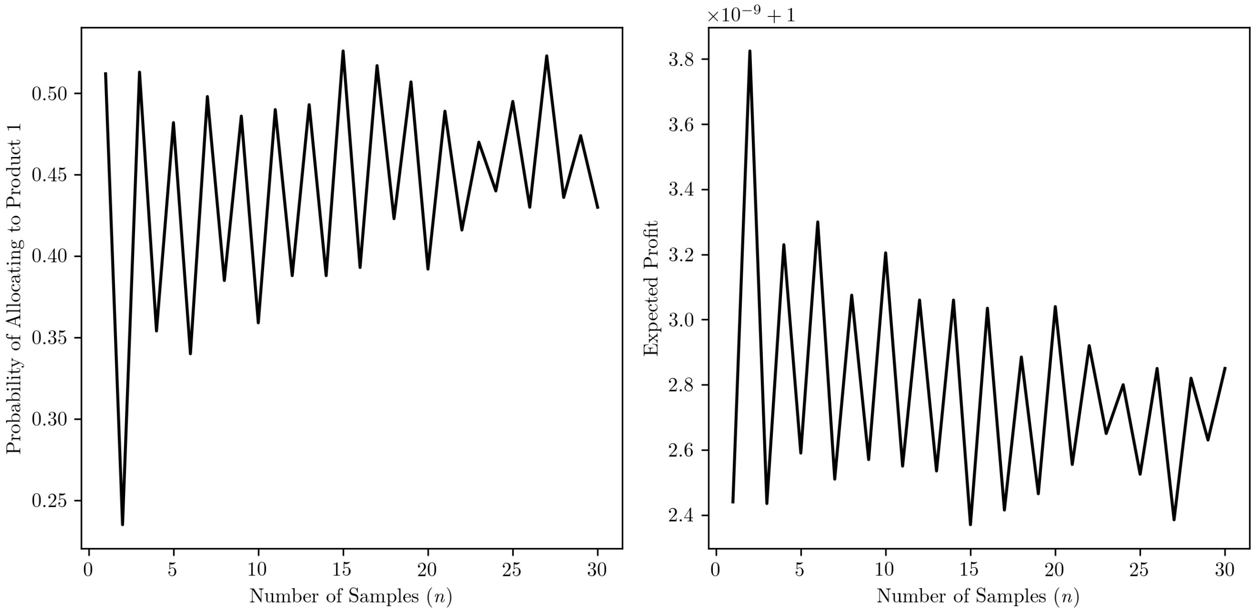

Figure 2. The decision quality across data sizes for the linear case.

5. Conclusions and Future Research Directions

This paper focuses on the question of whether larger data sizes inevitably lead to better decision making within the PO framework. Our examination of a non-linear contextual stochastic optimization problem and a linear contextual stochastic optimization problem reveals a nuanced relationship between the decision quality and the data size. We discover that the quality of decisions made under the PO framework does not always improve with an increase in data size, and in fact, may worsen, particularly in non-linear scenarios. Furthermore, in linear contexts, our research exposes a non-monotonic relationship between the decision quality and the data size.

These findings hold significant implications for decision-making practitioners. While larger datasets are generally assumed to lead to improved decisions, our results underscore that in the context of the PO framework, this assumption can be misleading. This can lead to suboptimal or even detrimental decision-making outcomes, particularly in contexts involving non-linear optimization problems.

Despite these implications, our study is not without limitations. Our exploration predominantly centers around a non-linear and a linear contextual stochastic optimization problem. While these cases provide valuable insights, there may be other contexts or problem types where the relationship between data size and decision quality varies. Consequently, the applicability of our findings might be restricted to similar problem contexts. Furthermore, while we have shown that non-monotonic relationships can occur in the PO framework, the specific conditions under which this happens need more investigation.

However, these findings and limitations open avenues for future research. Future work could seek to extend our approach to more complex problems, uncover the precise conditions triggering non-monotonicity, and devise strategies to handle or mitigate this phenomenon. In doing so, we can broaden our understanding of the intricacies of the PO framework and provide robust guidance for its application in diverse decision-making scenarios.

By highlighting a previously underexplored aspect of the PO framework, our study contributes to the ongoing discourse on optimizing decision-making processes in an era of abundant data. Ultimately, the aim of this research is to enhance our understanding of the PO framework's complexities and provide practical guidance for leveraging this tool effectively, especially in contexts where data size is a key factor.

Author Contributions: Conceptualization, S.W.; methodology, S.W. and X.T.; software, X.T.; validation, S.W. and X.T.; formal analysis, S.W. and X.T.; investigation, S.W. and X.T.; writing—original draft preparation, S.W. and X.T.; writing—review and editing, S.W. and X.T.; visualization, X.T.; supervision, S.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PO	predict-then-optimize
ML	machine learning
SAA	sample average approximation
wSAA	weighted sample average approximation
MSE	mean squared error
*k*NN	*k*-nearest neighbor

References

1. Ren, L.; Zhu, B.; Xu, Z. Robust consumer preference analysis with a social network. *Inf. Sci.* **2021**, *566*, 379–400. [CrossRef]
2. Ren, P.; Zhu, B.; Ren, L.; Ding, N. Online choice decision support for consumers: Data-driven analytic hierarchy process based on reviews and feedback. *J. Oper. Res. Soc.* **2022**, 1–14.. [CrossRef]
3. Tian, X.; Yan, R.; Wang, S.; Liu, Y.; Zhen, L. Tutorial on prescriptive analytics for logistics: What to predict and how to predict. *Electron. Res. Arch.* **2023**, *31*, 2265–2285. [CrossRef]
4. Martyn, K.; Kadziński, M. Deep preference learning for multiple criteria decision analysis. *Eur. J. Oper. Res.* **2023**, *305*, 781–805. [CrossRef]

5. Bertsimas, D.; Koduri, N. Data-driven optimization: A Reproducing Kernel Hilbert Space approach. *Oper. Res.* **2021**, *70*, 454–471. [CrossRef]
6. Birge, J.; Louveaux, F. *Introduction to Stochatic Programming*; Springer: New York, NY, USA, 2011.
7. Ben-Tal, A.; ELGhaoui, L.; Nemirovski, A. *Robust Programming*; Princeton University Press Princeton: Princeton, NI, USA, 2009.
8. Bertsimas, D.; Brown, D.; Caramanis, C. Theory and applications of robust optimization. *SIAM Rev.* **2011**, *53*, 464–501. [CrossRef]
9. Qi, M.; Shen, Z. Integrating prediction/estimation and optimization with applications in operations management. In *Tutorials in Operations Research: Emerging and Impactful Topics in Operations*; INFORMS: Catonsville, MD, USA, 2022; pp. 36–58.
10. Kleywegt, A.; Shapiro, A.; Homem-de, Mello, T. The sample average approximation for stochastic discrete optimization. *SIAM J. Optim.* **2002**, *12*, 479–502. [CrossRef]
11. Bertsimas, D.; Gupta, V.; Kallus, N. Data-driven robust optimization. *Math. Program.* **2018**, *167*, 235–292. [CrossRef]
12. Delage, E.; Ye, Y. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **2010**, *58*, 595–612. [CrossRef]
13. Besbes, O.; Mouchtaki, O. How big should your data really be? Data-driven newsvendor: Learning one sample at a time. *Manag. Sci.* 2023, *in press*. [CrossRef]
14. Ban, G.; Rudin, C. The big data newsvendor: Practical insights from machine learning. *Oper. Res.* **2019**, *67*, 90–108. [CrossRef]
15. Kallus, N.; Mao, X. Stochastic optimization forests. *Manag. Sci.* **2023**, *69*, 1975–1994. [CrossRef]
16. Elmachtoub, A.; Grigas, P. Smart "predict, then optimize". *Manag. Sci.* **2022**, *68*, 9–26. [CrossRef]
17. Notz, P.; Pibernik, R. Prescriptive analytics for flexible capacity management. *Manag. Sci.* **2022**, *68*, 1756–1775. [CrossRef]
18. Bertsimas, D.; Kallus, N. From predictive to prescriptive analytics. *Manag. Sci.* **2020**, *66*, 1025–1044. [CrossRef]
19. Tian, X.; Yan, R.; Liu, Y.; Wang, S. A smart predict-then-optimize method for targeted and cost-effective maritime transportation. *Transp. Res. Part B-Methodol.* **2023**, *172*, 32–52. [CrossRef]
20. Chan, T.; Shen, Z.; Siddiq, A. Robust defibrillator deployment under cardiac arrest location uncertainty via row-and-column generation. *Oper. Res.* **2018**, *66*, 358–379. [CrossRef]
21. El Balghiti, O.; Elmachtoub, A.N.; Grigas, P.; Tewari, A. Generalization bounds in the predict-then-optimize framework. *Math. Oper. Res.* 2023, *in press*.
22. Levi, R.; Perakis, G.; Uichanco, J. The data-driven newsvendor problem: New bounds and insights. *Oper. Res.* **2015**, *63*, 1294–1306. [CrossRef]
23. Cheung, W.C.; Simchi-Levi, D. Sampling-based approximation schemes for capacitated stochastic inventory control models. *Math. Oper. Res.* **2019**, *44*, 668–692. [CrossRef]
24. Wang, S.; Tian, X. A deficiency of the weighted sample average approximation (wSAA) framework: Unveiling the gap between data-driven policies and oracles. *Appl. Sci.* **2023**, *13*, 8355. [CrossRef]