Unbiased Objective Estimation in Predictive Optimization

Shinji Ito [1] **Akihiro Yabe** [1] **Ryohei Fujimaki** [1]

Abstract

For data-driven decision-making, one promising approach, called *predictive optimization*, is to solve maximization problems i n which the objective function to be maximized is estimated from data. Predictive optimization, however, suffers from the problem of a calculated optimal solution's being evaluated too optimistically, i.e., the value of the objective function is *overestimated*. This paper investigates such optimistic bias and presents two methods for correcting it. The first, which is analogous to cross-validation, successfully corrects the optimistic bias but results in *underestimation* of the true value. Our second method employs resampling techniques to avoid both overestimation and underestimation. We show that the second method, referred to as the parameter perturbation method, achieves asymptotically unbiased estimation. Empirical results for both artificial and real-world datasets demonstrate that our proposed approach successfully corrects the optimistic bias.

1. Introduction

Data-driven decision-making has become the subject of increased interest and been used in a number of practical applications. One of the most promising approaches is mathematical programming based on predictive models generated by machine learning. Recent advances in machine learning have made it easier to create accurate predictive models, and resulting predictions have been used to build mathematical programming problems (we refer to such approaches as *predictive optimization*). Predictive optimization is employed in applications for which *frequent trial-and-error process are not practical*, such as water distribution optimization (Draper et al., 2003), energy generation planning (Baos et al., 2011), retail price optimization (Johnson et al., 2016; Ito & Fujimaki, 2016), supply chain management (Thomas et al., 1996; Jung et al., 2004; Bertsimas & Thiele, 2004), and portfolio optimization (Markowitz, 1952; Chan et al., 1999; Konno & Yamazaki, 1991). Another important use for data-driven decision-making is in reinforcement learning (Kaelbling et al., 1996; Sutton & Barto, 2013). Here it is employed in situations mainly in which frequent trial-and-error operations are possible, except for batch reinforcement learning (Lange et al., 2012). The focus of this paper is on the first approach, i.e., predictive optimization.

In many practical applications of predictive optimization, it is essential to estimate the quality of the computed strategy because executing a strategy is often costly and risky. For example, predictive price optimization has been used to estimate revenue functions through regressions of demand as functions of product prices, and then, to optimize pricing strategies by maximizing estimated revenue functions (Johnson et al., 2016; Ito & Fujimaki, 2016; 2017; Yabe et al., 2017). In practice, users need to assess the return for the computed "optimal" strategy before changing prices, in order to prevent unforeseen heavy losses. In a situation in which costs for trial-and-error processes are unrealistically high, a key challenge in predictive optimization is how to assess the quality (or expected return) of the "optimal" solution by means of an estimated objective function.

Predictive optimization consists of two steps: estimation and optimization. In the estimation step, we construct an *estimated* objective function $f(z, \hat{\theta})$ for the *true* objective function $f(z, \theta^*)$, where θ is a parameter of f, and z is a decision variable corresponding to the strategy to be optimized. In the optimization step, we compute the *estimated* optimal strategy $\hat{z} = \arg\max_{z \in Z} f(z, \hat{\theta})$, where Z is the domain of z. Because it would be expensive to observe $f(\hat{z}, \theta^*)$ (i.e., to perform $\hat{z}$ in a real environment), we usually estimate it by $f(\hat{z}, \hat{\theta})$, which we call *simple evaluation*, in order to assess the quality of $\hat{z}$.

It has been empirically seen, however, that this simple evaluation tends to be too optimistic. For example, in the contexts of algorithmic investment and portfolio optimization, it has been reported (Michaud, 1989; Chapados, 2011; Harvey & Liu, 2015) that $f(\hat{z}, \hat{\theta})$ is much better than the acutual return. Michaud (Michaud, 1989) argued that this bias ap-

[1]NEC Corporation. Correspondence to: Shinji Ito <s-ito@me.jp.nec.com>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

pears because the mean-variance optimizers act as "error maximizers", i.e., optimizers tend to choose solutions containing large errors. According to (Harvey & Liu, 2015), a common practice in evaluating trading strategies is simple heuristics that discount the estimated objective to 50%, i.e., consider $0.5f(\hat{z}, \hat{\theta})$ to be an estimator of $f(\hat{z}, \theta^*)$. Heuristics referred to as portfolio resampling techniques (Michaud, 1998; Scherer, 2002) have been studied for nearly 20 years but have not yet to be theoretically justified. A few recent studies (Bailey & Marcos, 2016; Bailey et al., 2014; Harvey & Liu, 2015) have statistically analyzed and proposed algorithms to mitigate the bias issue, but their algorithms are restricted to particular applications (e.g., algorithmic investment) and, as far as we know, there exists no principled algorithm for an unbiased estimator of $f(z, \theta^*)$ in general predictive optimization problems.

The goal of this study is to address this optimistic bias issue, and to propose methods for unbiased estimation of true objective values. Our key contributions are summarized as follows.

First, we prove that the estimated optimal value $f(\hat{z}, \hat{\theta})$ is *biased* even if the estimated objective function $f(z, \hat{\theta})$ is an *unbiased* estimator of the true objective function $f(z, \theta^*)$. Further, we correlate the bias issue to overfitting in machine learning, which yields a valuable insight into bias correction methods.

Second, we propose two algorithms for estimating the value of true objective functions under mild assumptions. The first algorithm is based on a procedure similar to cross-validation and has been inspired by the analogy between our problem and overfitting in supervised learning. This algorithm corrects the optimistic bias, but suffers from *pessimistic bias*, i.e., the estimated value is biased in a direction suggesting a poorer result, similar to that which occurs in cross-validation. The magnitude of this pessimistic bias tends to be larger than that of cross-validation, and hence, it is not negligible in many cases. To mitigate this issue, we propose another algorithm, which we refer to as a *parameter perturbation method*. This algorithm employs a resampling technique and is theoretically proven here to achieve asymptotically unbiased estimation.

Our experimental results show that the proposed algorithms are able to estimate the value of a true objective function more accurately than a state-of-the-art hold-out validation technique commonly used in algorithmic investment (Bailey & Marcos, 2016; Bailey et al., 2014). In a simulation experiment with real-world retail datasets for price optimization, we have observed that our evaluation algorithms estimate a 17% increase in the gross profit, which seems to be more realistic and convincing than the value estimated without bias correction.

The remainder of this paper is structured as follows. In Section 2, we introduce the framework of the combination of machine learning and mathematical optimization in examples of usage. We also show that such a framework suffers from bias w.r.t. optimal values. Section 4 gives solutions to this problem and theoretical guarantees for them. In Section 5, the empirical performance of our algorithms is demonstrated.

2. Predictive Optimization

Suppose we have a sequence of training data $\mathbf{x} = (x_1, \ldots, x_N) \in X^N$, where N is the number of data instances. Each x_n is generated from a probabilistic model $\{p(x|\theta) : \theta \in \Theta\}$ parameterized by $\theta \in \Theta$. We further suppose having a set of objective functions $\{f(z, \theta) : \theta \in \Theta\}$ where $z \in Z$ is a decision variable that corresponds to strategies to be optimized. The goal of predictive optimization is to find $z^* \in \arg\max_{z \in Z} f(z, \theta^*)$, where θ^* is the true parameter. However, such a true parameter is unknown in practice, and therefore we estimate θ^* by $\hat{\theta}$ from $\mathbf{x}$, and compute the estimated optimal solution $\hat{z} \in \arg\max_{z \in Z} f(z, \hat{\theta})$ rather than z^*. This section discusses three examples of predictive optimization problems in order to provide a better picture of the process.

Example 1 (Coin-Tossing)**.** Suppose that we have a coin coming up heads with probability θ^* and tails with probability $1 - \theta^*$, where $\theta^* \in \Theta := [0, 1]$. Consider predicting heads or tails for this coin. If we predict the subsequent face correctly, we win \$1, and, otherwise, nothing. Predicting heads, then, will result in earning \$1 with probability θ^* and \$0 with probability $1 - \theta^*$, and hence, the expectation value of the earnings for predicting heads is $f(\text{'head'}, \theta^*) = 1 \cdot \theta^* + 0 \cdot (1 - \theta^*) = \theta^*$. Similarly, the expected earnings for predicting tails is $f(\text{'tail'}, \theta^*) = 1 - \theta^*$. If we knew the true parameter θ^*, we could maximize the expected earnings by choosing $z^* \in \arg\max_{z \in Z} f(z, \theta^*)$, where $Z = \{\text{'head'}, \text{'tail'}\}$ stands for a set of feasible strategies. Since we do not know the true parameter θ^*, however, we use, rather, past data $\mathbf{x} \in X^N := \{\text{'head'}, \text{'tail'}\}^N$ of N tossings, for estimating θ^*.

Table 1 illustrates how the optimistic bias occurs in predictive optimization. Suppose $\theta^* = 1/2$ (a) and that there are four cases of the observed pattern for three tossings (b). The estimators of θ^* might then be obtained as (c), using maximum likelihood estimation. On the basis of $\hat{\theta}$, the "best" strategies are estimated as (d), and the estimated and true optimal values are summarized in (e) and (f). It is worth noting that the expectation of (e) over four cases (bottom middle), which is $3/4$, is larger than the true expectation (bottom right), which is $1/2$ even if the $\hat{\theta}$ is unbiased, i.e., the expectation of $\hat{\theta}$ matches θ^* (bottom left).

Example 2 (Portfolio optimization (Markowitz, 1952))**.**

Table 1. Example of optimistic bias in coin-tossing.

	Case 1	Case 2	Case 3	Case 4
(a) θ^*	1/2	1/2	1/2	1/2
(b) $\mathbf{x}$	{HHH}	{HHT}	{HTT}	{TTT}
(c) $\hat{\theta}$	1	2/3	1/3	0
(d) $\hat{z}$	H	H	T	T
(e) $f(\hat{z}, \hat{\theta})$	1	2/3	2/3	1
(f) $f(\hat{z}, \theta^*)$	1/2	1/2	1/2	1/2

$E[\hat{\theta}]$	$E[f(\hat{z}, \hat{\theta})]$	$E[f(\hat{z}, \theta^*)]$
$1/2 = \theta^*$	$\mathbf{3/4}$	$1/2$

Suppose that there are d assets, and let R_j stand for the return on each component asset for $j \in \{1, \ldots, d\}$. Let $\mu^* = (\mu_1^*, \ldots, \mu_d^*)^\top \in \mathbb{R}^d$ be the expected return for each asset, i.e., $\mu_j^* = \mathbb{E}[R_j]$. Then the portfolio expressed as $R_z = \sum_{j=1}^d z_j R_j$, where $z_j \geq 0$ is the weighting of the j-th component asset and $z = (z_1, \ldots, z_d)^\top \in \mathbb{R}^d_{\geq 0}$, has expected return $\mathbb{E}[R_z] = \sum_{j=1}^d z_j \mu_j^* = \mu^{*\top} z$. Variance in the portfolio return can be expressed as $\mathrm{var}[R_z] = z^\top \Sigma^* z$, where Σ^* is the covariance matrix of $(R_1, \ldots, R_d)$. Denote $\theta^* = (\mu^*, \Sigma^*)$. Then, with a given risk tolerance $\lambda \geq 0$, the optimal portfolio is obtained as the solution of the following problem:

$$\begin{aligned} &\text{Maximize} \quad f(z, \theta^*) := \mu^{*\top} z - \lambda z^\top \Sigma^* z, \qquad (1) \\ &\text{subject to} \quad \sum_{j=1}^d z_j = 1, \quad z_j \geq 0 \quad (j = 1, \ldots, d). \end{aligned}$$

In practice, however, since θ^* is never available, we estimate it from historical data $\mathbf{x} = (x_1, \ldots, x_N)$, where $x_n \in \mathbb{R}^d$ is an observation of past returns for individual component assets (Qiu et al., 2015; Agarwal et al., 2006; Li & Hoi, 2012). Under the assumption that x_n follow the same distribution,[1] the estimators of μ^* and Σ^* are obtained by $\hat{\mu} = \frac{1}{N} \sum_{n=1}^N x_n$ and $\hat{\Sigma} = \frac{1}{N-1} \sum_{n=1}^N (x_n - \hat{\mu})(x_n - \hat{\mu})^\top$. We obtain the optimal solution by solving (1) with the replacement of μ^* and Σ^* by $\hat{\mu}$ and $\hat{\Sigma}$, respectively.

Example 3 (Predictive price optimization(Ito & Fujimaki, 2017; 2016))**.** Suppose we have d products whose prices are denoted by $z = (z_1, \ldots, z_d)$. Let us denote their sales quantities by $q^*(z) = (q_j^*(z))_{j=1}^d \in \mathbb{R}^d$, which are functions of the price z. The gross revenue function is then defined by $f(z, \theta^*) = q^*(z)^\top z$, and the true optimal solution is obtained by solving the following problem:

$$\text{Maximize} \quad q^*(z)^\top z \quad \text{subject to} \quad z \in Z, \qquad (2)$$

where $Z \subseteq \mathbb{R}^d$ is a pre-defined domain of prices (e.g., list price, 3%-off, 5%-off, and so on). However, we can never know the true demand-price relationship $q^*(z)$, and the predictive price optimization approximates $q^*(z)$ by the following regression functions:

$$q(z, \theta) = \sum_{k=1}^K \theta_k \psi_k(z) + \epsilon, \quad \epsilon \sim N(0, \Sigma), \qquad (3)$$

where $\{\psi_k : \mathbb{R}^d \to \mathbb{R}^d\}_{k=1}^K$ are fixed basis functions and $\{\theta_k\}_{k=1}^K \subseteq \mathbb{R}$ are regression coefficients. We estimate $\theta = (\theta_1, \ldots, \theta_K)$ as a standard regression problem and then solve (2) after replacing $q^*(z)$ by $q(z, \hat{\theta})$, where $\hat{\theta}$ is the estimator of θ^*.

3. Optimistic Bias in the Optimal Value

3.1. Existence of Optimistic Bias

This section formally proves the existence of optimistic bias in estimated optimal values. In the above examples, the objective functions $f(z, \theta)$ w.r.t. θ were affine functions and $\hat{\theta}$ were unbiased estimators of θ^*. Hence, the constructed objective function $f(z, \hat{\theta})$ was an unbiased estimator of the true objective function $f(z, \theta^*)$, i.e., it holds that

$$\mathbb{E}_{\mathbf{x}}[f(z, \hat{\theta})] = \mathbb{E}_{\mathbf{x}}[f(z, \theta^*)], \quad z \in \mathcal{Z}. \qquad (4)$$

From this equation, one might expect that $\mathbb{E}_{\mathbf{x}}[f(\hat{z}, \hat{\theta})]$ and $f(\hat{z}, \hat{\theta})$ would be reasonable estimators of $\mathbb{E}_{\mathbf{x}}[f(\hat{z}, \theta^*)]$ and $f(\hat{z}, \theta^*)$, respectively. However, the following proposition contradicts this intuition.

Proposition 1 (Optimistic Bias)**.** *Suppose* (4) *holds. For* $\hat{z} \in \arg\max_{z \in \mathcal{Z}} f(z, \hat{\theta})$ *and* $z^* \in \arg\max_{z \in \mathcal{Z}} f(z, \theta^*)$*, it holds that*

$$\mathbb{E}_{\mathbf{x}}[f(\hat{z}, \hat{\theta})] \geq f(z^*, \theta^*) \geq \mathbb{E}_{\mathbf{x}}[f(\hat{z}, \theta^*)]. \qquad (5)$$

The right inequality is strict if $\hat{z}$ *is suboptimal w.r.t. the true objective function* $f(z, \theta^*)$ *with non-zero probability.*

Proof. By taking the expectation of both sides of $f(\hat{z}, \hat{\theta}) \geq f(z^*, \hat{\theta})$, we obtain the left inequality of (5) as follows:

$$\mathbb{E}_{\mathbf{x}}[f(\hat{z}, \hat{\theta})] \geq \mathbb{E}_{\mathbf{x}}[f(z^*, \hat{\theta})] = f(z^*, \theta^*),$$

where the equality comes from (4). Similarly, the right inequality of (5) comes from $f(z^*, \theta^*) \geq f(\hat{z}, \theta^*)$. Further, if $\hat{z} \notin \arg\max_{z \in \mathcal{Z}} f(z, \theta^*)$ holds with non-zero probability, then $f(z^*, \theta^*) > f(\hat{z}, \theta^*)$ holds with non-zero probability and $f(z^*, \theta^*) \geq f(\hat{z}, \theta^*)$ always holds, which implies $f(z^*, \theta^*) > \mathbb{E}[f(\hat{z}, \theta^*)]$. □

This proposition implies that the estimated optimal value $f(\hat{z}, \hat{\theta})$ is not an unbiased estimator of $f(\hat{z}, \theta^*)$ even if the estimated objective function $f(z, \hat{\theta})$ is an unbiased estimator of the true objective function $f(z, \theta^*)$. This optimistic bias

[1] This condition can easily be relaxed.

has been empirically learned in the context of portfolio optimization (Michaud, 1989). Recently, (Harvey & Liu, 2015; Harvey et al., 2016) have proposed bias correction methods based on statistical tests, though their methods are applicable only to cases in which the objective function is the Sharpe ratio. Other recent studies (Bailey & Marcos, 2016; Bailey et al., 2014) have also focused on the Sharpe ratio and proposed a hold-out validation method. Although their methods apply to general predictive optimization problems, they have not been proven to obtain unbiased estimators. Note that a similar inequality has been discovered in the context of stochastic programs,[2] one that corresponds to the left inequality of (5). For the special case in which Z is a finite set, the same inequality as (5) has been shown in the context of decision analysis (Smith & Winkler, 2006).

3.2. Connection to Empirical Risk Minimization

This subsection discusses the connection of the optimistic bias issue to overfitting in machine learning, which connection has led to the ideas underlying our proposed algorithms. In supervised machine learning, we choose the prediction rule $\hat{h}$ from a hypothesis space $\mathcal{H}$ by minimizing the empirical error, i.e., we let $\hat{h} \in \arg\min_{h\in\mathcal{H}} \frac{1}{n}\sum_{n=1}^{N} \ell(h, x_n)$, where x_n is the observed data generated from a distribution $\mathcal{D}$ and ℓ is a loss function. The empirical error $\frac{1}{N}\sum_{n=1}^{N} \ell(h, x_n)$ is an unbiased estimator of the generalization error $\ell_{\mathcal{D}}(h) := \mathbb{E}_{x\sim\mathcal{D}}[\ell(h, x)]$ for arbitrary fixed prediction rule h, i.e., it holds that $\mathbb{E}_{x_n\sim\mathcal{D}}[\frac{1}{N}\sum_{n=1}^{N} \ell(h, x_n)] = \ell_{\mathcal{D}}(h)$ for any fixed h. Despite this equation, the empirical error $\frac{1}{N}\sum_{n=1}^{N} \ell(\hat{h}, x_n)$ for the computed parameter $\hat{h}$ is smaller than the generalization error $\ell_{\mathcal{D}}(\hat{h})$ in most cases, because $\hat{h}$ *overfits* the observed samples, as is well known (Vapnik, 2013). The analogy between the optimistic bias in our setting and the overfitting issue in machine learning suggests the reuse of datasets for estimation of their objective functions and evaluation of objective values.

A comparison between empirical risk minimization (ERM) and our prediction-based optimization is summarized in Table 2. As is shown in the Table, our problem concerning bias in predictive optimization has a structure similar to that of the problem of overfitting in empirical risk minimization. Typical methods for estimating generalization error in machine learning would be cross-validation and such asymptotic bias correction as AIC (Akaike, 1973). This paper follows the concept of cross-validation in the context of predictive optimization and, in the following section, proposes a more accurate algorithm.

[2] In stochastic programs, the objective is a random function, and it has been shown in, e.g., (Mak et al., 1999), that *the expectation of the minimum* of the objective is a lower bound of *the minimum of the expectation* of the objective.

Table 2. Correspondence of empirical risk minimization and predictive optimization

	ERM	Optimization
Decision variable	Predictor h	Strategy z
True objective	$\mathbb{E}_{x\sim\mathcal{D}}[\ell(h, x)]$	$f(z, \theta^*)$
Estimated objective	$\frac{1}{N}\sum_{n=1}^{N} \ell(h, x_n)$	$f(z, \hat{\theta})$

4. Bias Correction Algorithms

Our goal is to construct unbiased estimators for the value $f(\hat{z}, \theta^*)$ of the true objective function, i.e., to construct $\rho : X^n \to \mathbb{R}$ such that $\mathbb{E}_{\mathbf{x}}[\rho(\mathbf{x})] = \mathbb{E}_{\mathbf{x}}[f(\hat{z}, \theta^*)]$, where $\hat{z} \in \arg\max_{z\in Z} f(z, \hat{\theta})$ is the computed strategy. We assume the following conditions.

Assumption 2. *(i)* $f(z, \theta)$ *is affine in* θ*, i.e.,* $\exists a : Z \to \mathbb{R}^d$, $\exists b : Z \to \mathbb{R}$, $f(z, \theta) = \theta^\top a(z) + b(z)$.

(ii) The optimal solution $z(\theta) \in \arg\max_{z\in Z} f(z, \theta)$ *is uniquely determined for almost all* θ.

(iii) One of the following holds: (iii.a) Z *is a finite set, or (iii.b)* Z *is a compact subset of* $\mathbb{R}^d$*, and* $z \mapsto (a(z), b(z))$ *is a continuous injective function.*

(iv) $\hat{\theta}$ *is an unbiased estimator of* θ^**, i.e., we have* $\mathbb{E}_{\mathbf{x}}[\hat{\theta}] = \theta^*$.

The assumptions (i)-(iii) are conditions on mathematical programming problems, and such typical ones as (mixed-integer) linear/quadratic/semidefinite programming problems satisfy these conditions. Assumption (iv) is a condition on the machine learning algorithm for estimating the objective function in the optimization problem, and we can employ any standard unbiased estimation algorithm. Note that the examples in Section 3 satisfy all these assumptions. We assume (i) and (iv) in Section 4.1, and assume (i)-(iv) in Section 4.2.

4.1. Cross-Validation Method

As noted in Section 3.2, our problem is closely related to the problem of estimating generalization error. Inspired by the cross-validation method, one of the most popular methods for estimating generalization error in machine learning, we propose a cross-validation method for estimating the value of the true objective function in predictive optimization. In the context of algorithmic investment, a similar method, referred to as the hold-out method is mentioned in (Bailey et al., 2014). The method discussed below is essentially an extension of the hold-out method for general predictive optimization problems.

One of the reasons that the value $f(\hat{z}, \hat{\theta})$ contains biases is that $\hat{z}$ and $\hat{\theta}$ are dependent random variables. Indeed,

Algorithm 1 k-fold cross-validation

Input: data $\mathbf{x} \in X^N$, the number $K \geq 2$ of partition
Divide data $\mathbf{x}$ into K parts $\mathbf{x}_1, \ldots, \mathbf{x}_K$.
for $k = 1$ **to** K **do**
  Compute $\hat{\theta}_k, \tilde{\theta}_k$ from $\mathbf{x}_k, \mathbf{x}_{-k}$ respectively, where we define $\mathbf{x}_{-k}$ to be all samples in $\mathbf{x}$ except for $\mathbf{x}_k$, and compute $\tilde{z}_k \in \arg\max_{z \in Z} f(z, \tilde{\theta}_k)$.
end for
Output $\rho_{CV}(\mathbf{x}) := \frac{1}{K} \sum_{k=1}^{K} f(\tilde{z}_k, \hat{\theta}_k)$.

if $\hat{z}$ and $\hat{\theta}$ are independent, $\mathbb{E}_{\mathbf{x}}[f(\hat{z}, \hat{\theta})] = \mathbb{E}_{\mathbf{x}}[f(\hat{z}, \theta^*)]$ straightforwardly holds from assumptions (i) and (iv). The main idea of the cross-validation method (as with the standard cross-validation in machine learning) is to divide the data $\mathbf{x} \in X^N$ into two parts $\mathbf{x}_1 \in X^{N_1}, \mathbf{x}_2 \in X^{N_2}$, where $N_1 + N_2 = N$. Note that each element in $\mathbf{x}_1$ and $\mathbf{x}_2$ follows $p(x, \theta^*)$ independently, and, hence, $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent random variables. Let us denote the estimators based on $\mathbf{x}_1$ and $\mathbf{x}_2$ by $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively. Also, the optimal strategy on each estimator is denoted by $\hat{z}_i := \arg\max_{z \in Z} f(z, \hat{\theta}_i)$ for $i = 1, 2$. Then $\hat{z}_1$ and $\hat{\theta}_2$ are independent (the opposite also holds), and we have $\mathbb{E}_{\mathbf{x}}[f(\hat{z}_1, \hat{\theta}_2)] = \mathbb{E}_{\mathbf{x}_1}[f(\hat{z}_1, \mathbb{E}_{\mathbf{x}_2}[\hat{\theta}_2])] = \mathbb{E}_{\mathbf{x}_1}[f(\hat{z}_1, \theta^*)]$. Further, if N_1 is sufficiently close to N, $\mathbb{E}_{\mathbf{x}_1}[f(\tilde{z}_1, \theta^*)]$ is close to $\mathbb{E}_{\mathbf{x}}[f(\hat{z}, \theta^*)]$. This idea can be extended to k-fold cross-validation, in which we divide data $\mathbf{x} \in \mathbb{R}^N$ into K parts $\mathbf{x}_1, \ldots, \mathbf{x}_K \in \mathbb{R}^{N'}$, where $KN' = N$. We compute $\tilde{z}_k$ from $\{\mathbf{x}_1, \ldots, \mathbf{x}_K\} \setminus \{\mathbf{x}_k\}$, and compute $\hat{\theta}_k$ from $\mathbf{x}_k$. Then the value $\rho_{CV}(\mathbf{x}) := \frac{1}{K} \sum_{k=1}^{K} f(\tilde{z}_k, \hat{\theta}_k)$ satisfies

$$\mathbb{E}_{\mathbf{x}}[\rho_{CV}(\mathbf{x})] = \mathbb{E}_{\mathbf{x}'}[f(\tilde{z}, \theta^*)], \tag{6}$$

where $\tilde{z}$ stands for the strategy computed from $(K-1)N'$ samples, under assumptions (i) and (iv).

A major drawback to Algorithm 1 is that it can only estimate the objective value attained by $N - N'$ samples, as is shown in (6), even though the value attained by all N samples is desired. In machine learning, to mitigate this gap, a leave-one-out method (i.e., setting $N' = 1$) can be used. In predictive optimization, however, the number N' of hold-out samples needs to be large enough to compute another estimator, $\hat{\theta}_k$, which limits the accuracy of the estimation of $f(\hat{z}, \theta^*)$. The accuracy of Algorithm 1 is considered in Sec. 5 in an empirical evaluation.

4.2. Parameter perturbation method

This subsection proposes another algorithm that addresses the drawbacks of Algorithm 1. Denote the error in the estimated parameter by $\delta := \hat{\theta} - \theta^*$. The error δ depends on the training data $\mathbf{x}$ and can be regarded as a random variable when $\mathbf{x}$ is considered to be a random variable. For $\gamma \geq 0$, let us first define $\eta(\gamma)$ as follows:

$$\eta(\gamma) = \mathbb{E}_{\delta}[f(z(\theta^* + \gamma\delta), \theta^*)],$$

where $z(\theta) := \arg\max_{z \in Z} f(z, \theta)$. Since $\hat{z} = z(\hat{\theta}) = z(\theta^* + \delta)$, we have $\eta(1) = \mathbb{E}[f(\hat{z}, \theta^*)]$. Hence, our goal, unbiased estimation of $f(\hat{z}, \theta^*)$, is equivalent to unbiased estimation of $\eta(1)$. Let us next define $\phi(\gamma)$ as follows:

$$\phi(\gamma) = \mathbb{E}_{\delta}[f(z(\theta^* + \gamma\delta), \theta^* + \gamma\delta)]. \tag{7}$$

Note that we have $\phi(1) = \mathbb{E}[f(\hat{z}, \hat{\theta})]$. Further, $\phi(\gamma)$ and $\eta(\gamma)$ satisfy $\phi(0) = \eta(0) = f(z^*, \theta^*)$ and $\phi(\gamma) \geq f(z^*, \theta^*) \geq \eta(\gamma)$ for all $\gamma \geq 0$, which can be proved in a way similar to that of the proof of Proposition 1.

The following proposition plays a key role in our second algorithm.

Proposition 3. *Suppose that assumptions (i)-(iv) hold. For all $\gamma > 0$, $\phi(\gamma)$ is differentiable, and its derivative $\phi'(\gamma)$ satisfies*

$$\eta(\gamma) = \phi(\gamma) - \gamma\phi'(\gamma). \tag{8}$$

The proof of this proposition is summarized in the supplementary material.

Let us explain this proposition using Figure 1, which is based on the simulation experiment for portfolio optimization used in Section 5 and shows how the values of ϕ and η behave for some $\gamma \geq 0$. The tangent to $\phi(\gamma)$ at $\gamma = \gamma_0$ (the blue broken-line) has a y-intercept (the red broken-line) equal to the value of $\eta(\gamma_0)$, for all $\gamma_0 > 0$. From this relationship, the derivative $\phi'(1)$ of $\phi(\gamma)$ at $\gamma = 1$ satisfies $\phi'(1) = \phi(1) - \eta(1) = \mathbb{E}[f(\hat{z}, \hat{\theta})] - \mathbb{E}[f(\hat{z}, \theta^*)]$, i.e., the value of $\phi'(1)$ is equal to the value of the bias in our predictive optimization problem.

Our problem is now to obtain an unbiased estimator ζ of $\phi'(1)$ that will give us an unbiased estimator of $f(\hat{z}, \theta^*)$, i.e. $\rho = f(\hat{z}, \hat{\theta}) - \zeta$. From the definition of the derivative, the value of $\phi'(1)$ can be approximated by $(\phi(1+h) - \phi(1))/h$ for small h. Further, from the definition of ϕ, the estimated optimal value $f(\hat{z}, \hat{\theta})$ is an unbiased estimator of $\phi(1)$. Also, the value of $\phi(1+h) = \mathbb{E}[\max_{z \in Z} f(z, \theta^* + (1+h)\delta)]$ is the expectation of the optimal value for the objective function with a parameter having an "enhanced" error. If we get samples $\hat{\theta}_h$ following the distribution of $\theta^* + (1+h)\delta$, we can develop an estimator of $\phi(1+h)$, and accordingly, we can estimate $\eta(1) = \mathbb{E}[f(\hat{z}, \theta^*)]$.

Suppose that $\hat{\theta}_h^{(1)}, \ldots, \hat{\theta}_h^{(s)}$ follows the distribution of $\theta^* + (1+h)\delta$, and define

$$\rho_h := \frac{1+h}{h} \max_{z \in Z} f(z, \hat{\theta}) - \frac{1}{hs} \sum_{j=1}^{s} \max_{z \in Z} f(z, \hat{\theta}_h^{(j)}). \tag{9}$$

The value ρ_h, then, has the following property.

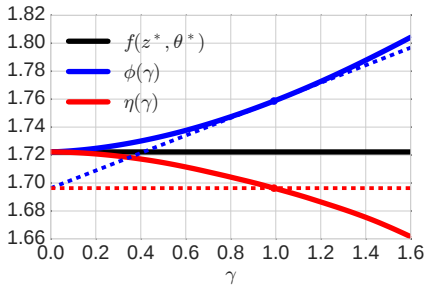

Figure 1. Comparison among $\phi(\gamma)$, $\eta(\gamma)$ and $f(z^*, \theta^*)$. The blue broken-line is the tangent to $\phi(\gamma)$ at $\gamma = 1$, and the red broken-line represents its y-intercept.

Algorithm 2 Parameter perturbation method

Input: data $\mathbf{x} \in X^n$, parameters $h > 0$, $s \in \{1, 2, \ldots\}$
Compute $\hat{\theta}$ from $\mathbf{x}$ and set $\hat{v}_0 = \max_{z \in \mathcal{Z}} f(z, \hat{\theta})$.
Generate $\{\hat{\theta}_h^{(j)}\}_{j=1}^s$ by (i) for asymptotic normal estimators or (ii) for M-estimators.

(i) Set $\hat{\theta}_h^{(j)}$ to be the estimator computed from $N/(1+h)^2$ samples randomly chosen from $\mathbf{x}$ without replacement.

(ii) Generate $\hat{\delta}_j$ by (10), and set $\hat{\theta}_h^{(j)} = \hat{\theta} + \hat{\delta}_j$.

for $j = 1$ **to** s **do**
  Set $\hat{v}_j = \max_{z \in \mathcal{Z}} f(z, \hat{\theta}_h^{(j)})$.
end for
Output $\rho_h := \frac{1+h}{h}\hat{v}_0 - \frac{1}{hs}\sum_{j=1}^s \hat{v}_j$.

Proposition 4. *Under assumptions (i)-(iv), the value ρ_h defined by (9) is an asymptotically unbiased estimator of $f(\hat{z}, \theta^*)$, i.e., it holds that $\lim_{h \to 0} \mathbb{E}[\rho_h] = \mathbb{E}[f(\hat{z}, \theta^*)]$.*

Proof. From the definition of ρ_h and $\phi(\gamma)$, we have $\mathbb{E}[\rho_h] = \rho(1) - \frac{\phi(1+h)-\phi(1)}{h}$. Hence, we have $\lim_{h \to 0} \mathbb{E}[\rho_h] = \phi(1) - \phi'(1)$. From Proposition 3, this value is equal to $\eta(1) = \mathbb{E}[f(\hat{z}, \theta^*)]$. □

The remaining problem is how to obtain samples $\hat{\theta}_h$, with enhanced errors, from the distribution of $\theta^* + (1+h)\delta$. If $\hat{\theta}$ is an asymptotically normal estimator of θ^*, its distribution can be approximated by the normal distribution $\mathcal{N}(\theta^*, \frac{1}{N}\Sigma^*)$, where Σ^* is a constant matrix not dependent on N. Further, when we compute an estimator $\hat{\theta}_h$ from $N/(1+h)^2$ data, the distribution of $\hat{\theta}_h$ can be approximated by $\mathcal{N}(\theta^*, \frac{(1+h)^2}{N}\Sigma^*)$. This is an approximation of the distribution of $\theta^* + (1+h)\delta$. This procedure for generating $\hat{\theta}_h$ is used in (i) of Algorithm 2.

If $\hat{\theta}$ is an M-estimator, an asymptotically normal estimator commonly used in machine learning, we can eliminate repetitive computation in (i) of Algorithm 2. For M-estimators, $\hat{\Sigma}$ is given in a closed form, as described in (van der Vaart, 1998), such that $\mathcal{N}(0, \frac{1}{N}\hat{\Sigma})$ approximates the error distribution of the estimator. Once we have computed $\hat{\Sigma}$, we generate samples from an approximated distribution of $\theta^* + (1+h)\delta$, by adding $\hat{\delta}$ to $\hat{\theta}$, which is obtained by

$$\hat{\delta} \sim \mathcal{N}(0, \frac{(1+h)^2 - 1}{N}\hat{\Sigma}). \tag{10}$$

We can, in fact, confirm that the distribution of $\hat{\theta} + \hat{\delta}$ approximates that of $\theta^* + (1+h)\delta$ by applying the normal approximation to $\hat{\theta} - \theta^* = \delta$. From the normal approximation $\delta \sim \mathcal{N}(0, \frac{1}{N}\hat{\Sigma})$, we obtain $\theta^* + (1+h)\delta \sim \mathcal{N}(\theta^*, \frac{(1+h)^2}{N}\hat{\Sigma})$ and $\hat{\theta} + \hat{\delta} \sim \mathcal{N}(\theta^* + 0, \frac{1}{N}\hat{\Sigma} + \frac{(1+h)^2-1}{N}\hat{\Sigma}) = \mathcal{N}(\theta^*, \frac{(1+h)^2}{N}\hat{\Sigma})$. This procedure corresponds to (ii) in Algorithm 2.

5. Experiments

We have compared our Algorithm 1 and Algorithm 2 with the hold-out method (Bailey & Marcos, 2016; Bailey et al., 2014) and the portfolio resampling method (Scherer, 2002) by means of the simulation models of the examples in Section 2. We used GUROBI Optimizer 6.0.4[3] for portfolio optimization, and the algorithm in (Ito & Fujimaki, 2016) for price optimization.

5.1. Predictive Portfolio Optimization

The portfolio optimization problem described in Example 2 of Section 2 was constructed with $\theta^* = (\mu^*, \Sigma^*)$ defined by $\mu^* = \mathbf{1} + \epsilon$ and $\Sigma^* = X^\top X$, where $\epsilon \in \mathbb{R}^d$ were generated by $N(0, I)$ and each entry of $X \in \mathbb{R}^{D \times D}$ was drawn from $\mathcal{N}(0, D^{-1})$. We generated datasets $\{\mathbf{x}_n\}_{n=1}^N$ following $\mathcal{N}(\mu^*, \Sigma^*)$, from which we computed $\hat{\theta}$, as in Example 2, and solved the optimization problem (1) with θ^* replaced by $\hat{\theta}$, to obtain $\hat{z}$. We chose $D = 50$, $N = 20$, and $\lambda = 1.0$ for our simulation experiments. When using the portfolio resampling method, we computed $\bar{z}$ by means of 10 bootstrap resamplings and outputted $f(\bar{z}, \hat{\theta}) \leq f(\hat{z}, \hat{\theta})$. For details regarding portfolio resampling, see, e.g., (Scherer, 2002). For the hold-out validation, we first divided N data into N' and $N - N'$, then computed $\hat{z}_1$ from the former N' data and estimated $\hat{\theta}_2$ from the letter $N - N'$ data, and then calculated $f(\hat{z}_1, \hat{\theta}_2)$.

Accuracy Comparisons Figure 2 shows the means and the standard deviations of computed values of $f(z^*, \theta^*)$, $f(\hat{z}, \hat{\theta})$ and $f(\hat{z}, \theta^*)$ for 400 randomly-initialized datasets. We have observed that:

- $f(\hat{z}, \hat{\theta})$ was much larger than $f(\hat{z}, \theta^*)$, which is consistent with Proposition 1.
- The hold-out method performed much worse than our

[3] http://www.gurobi.com/

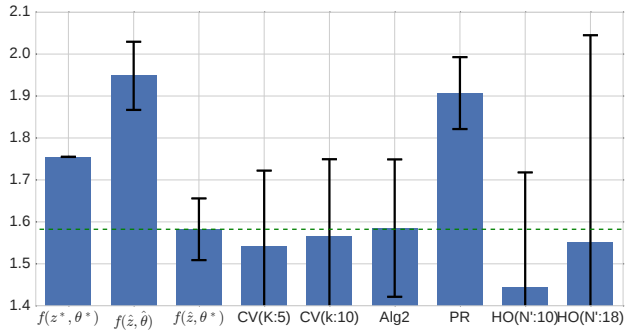

Figure 2. Values of the objective function and estimated values of $f(\hat{z}, \theta^*)$ with Algorithms 1, 2, and the hold-out validation. CV, PR and HO stand for Algorithm 1, portfolio resampling, and Hold-out validation, respectively. Blue bars and error bars represent means and standard deviations, respectively.

CV and perturbation methods, though its performance improved with an increasing N'. Also, the variance in the proposed methods was much smaller. Note that we could not set N' to be larger than $N' = 18$ since the estimation of $\hat{\theta}_1$ and $\hat{\theta}_2$ would fail.

- The portfolio resampling method computed slightly less optimistic value than $f(\hat{z}, \hat{\theta})$, but a large amount of optimistic bias remained.
- The perturbation method corrected bias better than the CV method w.r.t. both bias and variance. Indeed, it almost perfectly corrected the optimistic bias in expectation. Note that $K = 10$ was the largest possible value because at least two samples are necessary for estimating the covariance matrix. This means that the value of CV ($K = 10$) achieved the minimum bias for the CV method.
- The CV method and the hold-out method produced conservative estimates. The pessimistic bias in the CV method came from the difference between $\hat{z} \in \arg\max_{z \in Z} f(z, \hat{\theta})$ and $\tilde{z}$ in (6).

Note that $E[f(\hat{z}, \theta^*)]$ was poorer than $E[f(z^*, \theta^*)]$, where the former was the best objective value achieved with the available finite training samples. This negative difference is unavoidable with our bias correction, which appears to raise an interesting open challenge w.r.t. the combination of our bias correction with robust optimization (Bertsimas et al., 2011), i.e., the former mitigates the optimistic bias, and the later mitigates uncertainty in objective functions.

Sensitivity of the Perturbation Method We investigated the sensitivity of the perturbation method w.r.t. $h > 0$, which is the important trade-off parameter in bias and variance. We applied it to 100 different randomly-initialized datasets, for which we set $h = 0.05, 0.10, \ldots, 0.50$. Because s is not sensitive, we fixed it to $s = 10$. Figure 3 demonstrates the changes in bias and variance (top figure) and RMSE against $f(\hat{z}, \theta^*)$, over h. As the value

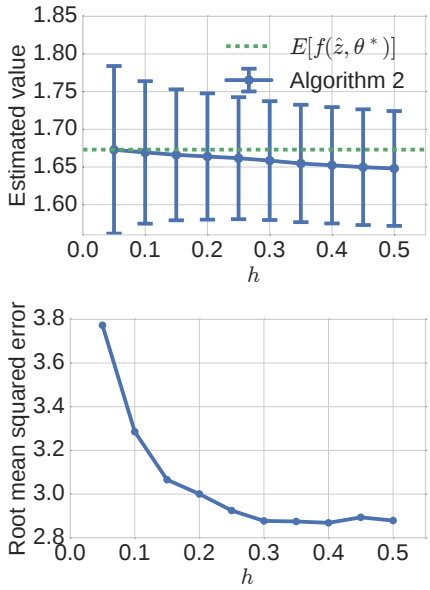

Figure 3. Bias, variance (top), and RMSE (bottom) values over h obtained with the perturbation method. The error bars in the top figure represent the standard derivations.

of h increased, the bias increased though the variance decreased (top figure), as was implied in Proposition 4, and this resulted in significantly larger RMSE values with smaller values of h. This observation indicates that an appropriate balance between bias and variance must be determined, and that a variance-sensitive measure such as RMSE can be used as a guide to determine the trade-off.

5.2. Predictive Price Optimization

We applied our algorithms to the predictive price optimization discussed as Example 3 in Section 2. As reported in (Ito & Fujimaki, 2017), the optimal value in this problem contains optimistic bias, which is consistent with Proposition 1. Unlike in the portfolio optimization, the parameter $\hat{\theta}$ is estimated by regression techniques, and the set of feasible strategies Z is discrete.

Simulation Experiment In this experiment, we investigated the effect of the optimistic bias and our bias correction over parameter dimensionality, i.e., the number of products d. We generated the same simulation data as in (Ito & Fujimaki, 2017). The sales quantity q_i of the i-th product was generated from the regression model $q_i = \alpha_i + \sum_{j=1}^{d} \beta_{ij} p_j$, where α_i and β_{ij} were generated by uniform distributions, where $\alpha_i \in [d, 3d]$, $\beta_{ij} \in [0, 2]$ for $i \neq j$, and $\beta_{ii} \in [-2d, -d]$. The feasible region Z was defined by $Z = \{0.6, 0.7, \ldots, 1.0\}^d$. We chose $N = 500$ for our experiments.

Figure 4 shows the change in the objective values normalized by the ideal objective value $f(z^*, \theta^*)$ over the number

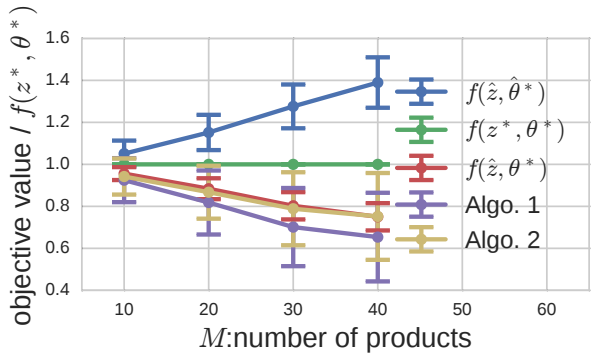

Figure 4. Bias and variance over parameter dimensionality. The horizontal axis represents objective values normalized by the ideal objective value.

of products d. For Algorithm 1 (CV method), we chose $K = 2$ so that the hold-out samples would be sufficient to estimate parameters $\{\alpha_i\}$ and $\{\beta_{ij}\}$. We observed that:

- $f(\hat{z}, \theta^*)$ degraded against $f(z^*, \theta^*)$ with increasing d because the estimation error in machine learning increased.
- The optimistic bias, $f(\hat{z}, \hat{\theta}) - f(\hat{z}, \theta^*)$, rapidly increased because $f(\hat{z}, \hat{\theta}) - f(z^*, \theta^*)$ also increased in addition to the increase in $f(z^*, \theta^*) - f(\hat{z}, \theta^*)$.
- The CV method suffered from pessimistic bias, which increased as d increased.
- The perturbation method corrected the bias accurately even if the parameter dimensionality, i.e., d, increased.

These results confirm the robustness of our proposed method over parameter dimensionality and also its general applicability to a wide range of problems (the portfolio optimization in Section 5.1 is continuous and convex while the price optimization in this section is discrete and non-convex).

Real-World Retail Dataset The real-world retail dataset used in (Ito & Fujimaki, 2017; 2016) contains sales information for a middle-size supermarket located in Tokyo.[4] Using this information, we selected 50 regularly-sold beer products. The data range was approximately the three years from 2012/01 to 2014/11. We used the first 35 months (1063 samples) for training regression models and simulated the best price strategy for the next day 2014/12/1. We estimated parameters in regression models, using the least squares method. The other settings were same as in (Ito & Fujimaki, 2016).

The actual (non-optimized) gross profit in the past data was $106,348$ JPY, while the estimated optimal value $f(\hat{z}, \hat{\theta})$ was $490,502$ JPY, which represents an approximately 361% increase in gross profit, but this value was obviously unrealistically huge and unreliable (price changes alone could not increase a profit 4.6 by times!). The *bias-corrected* optimal gross profit with the perturbation method at $h = 0.1$ and $s = 100$ was $124,477$ JPY, which represents an approximately 17% increase in the gross profit. Although we were unable to confirm the validity of this value since this experiment was conducted on past historical data, intuitively speaking, a 17% increase in gross profit seems much more realistic than one of 361%, and considering the facts noted in the simulation studies, our result would surely seem more convincing to domain users. One of important remaining issues in real applications is the estimation of the confidence region. As noted above, we can never learn the value of $f(\hat{z}, \theta^*)$ without performing $\hat{z}$, but the user has to make a decision as to whether to perform it or not without knowing the value. In such a case, it would be helpful to provide a confidence region w.r.t. the *bias-corrected* optimal value, which is available with neither the CV method nor the perturbation method.

[4] The data were provided by KSP-SP Co., LTD, http://www.ksp-sp.com.

6. Conclusion

In this paper, we have focused on the framework of a combination of mathematical optimization and machine learning with which we solve an optimization problem whose objective is formulated with the aid of predictive models or estimators. We have demonstrated that such a framework suffers from a kind of bias w.r.t. optimal values because of overfitting of the solution to the constructed objective function. We have proposed a solution to this bias problem by means of developed methods that are guaranteed to compute an asymptotically unbiased estimator of the value of the true objective function. Empirical results have demonstrated that the proposed approach results in successful estimates of the value of the true objective function.

A major open question remaining in this work is how to evaluate and reduce variance in the estimators of objective functions. The variance in estimators, i.e., uncertainty in estimation, is essential information for decision makers in many situations, and reducing variance in the estimator would help them make better decisions.

References

Agarwal, A., Hazan, E., Kale, S., and Schapire, R. E. Algorithms for portfolio management based on the Newton method. *Proceedings of the 23rd international conference on Machine learning - ICML '06*, pp. 9–16, 2006.

Akaike, H. Information theory and an extension of the maximum likelihood principle. In *International Symposium on Information Theory*, pp. 267–281, 1973.

Bailey, D. H. and Marcos, L. Stock portfolio design and

backtest overfitting. *SSRN Working Paper*, pp. 1–14, 2016.

Bailey, D. H., Borwein, J. M., López de Prado, M., and Zhu, Q. J. Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance. *Notices of the AMS*, 61(5):458–471, 2014.

Baos, R., Manzano-Agugliaro, F., Montoya, F., Gil, C., Alcayde, A., and Gmez, J. Optimization methods applied to renewable and sustainable energy: A review. *Renewable and Sustainable Energy Reviews*, 15(4):1753 – 1766, 2011.

Bertsimas, D. and Thiele, A. A robust optimization approach to supply chain management. *Integer programming and combinatorial optimization*, 1(11):145–156, 2004.

Bertsimas, D., Brown, D. B., and Caramanis, C. Theory and Applications of Robust Optimization. *SIAM Review*, 53 (3):464–501, 2011.

Chan, L. K. C., Karceski, J., and Lakonishok, J. On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies*, 12(5):937–974, 1999.

Chapados, N. *Portfolio choice problems: An introductory survey of single and multiperiod models*. Springer Science & Business Media, 2011.

Draper, A. J., Jenkins, M. W., Kirby, K. W., Lund, J. R., and Howitt, R. E. Economic-Engineering Optimization for California Water Management. *Journal of water resources planning and management*, 129(June):155–164, 2003.

Harvey, C. R. and Liu, Y. Backtesting. *The Journal of Portfolio Management*, 42(1):13–28, 2015.

Harvey, C. R., Liu, Y., and Zhu, H. ...and the Cross-Section of Expected Returns. *Review of Financial Studies*, 29(1): 5–68, 2016.

Ito, S. and Fujimaki, R. Large-scale price optimization via network flow. In *Advances in Neural Information Processing Systems*, pp. 3855–3863, 2016.

Ito, S. and Fujimaki, R. Optimization beyond prediction: Prescriptive price optimization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1833–1841, 2017.

Johnson, K., Hong, B., Lee, A., and Simchi-levi, D. Analytics for an Online Retailer : Demand Forecasting and Price Optimization. *Manufacturing & Service Operations Management*, 18(1):69–88, 2016.

Jung, J. Y., Blau, G., Pekny, J. F., Reklaitis, G. V., and Eversdyk, D. A simulation based optimization approach to supply chain management under demand uncertainty. *Computers and Chemical Engineering*, 28(10):2087–2106, 2004.

Kaelbling, L. P., Littman, M. L., and Moore, A. W. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

Konno, H. and Yamazaki, H. Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. *Management science*, 37(5):519–531, 1991.

Lange, S., Gabel, T., and Riedmiller, M. Batch reinforcement learning. In *Reinforcement learning*, pp. 45–73. Springer, 2012.

Li, B. and Hoi, S. C. H. On-Line Portfolio Selection with Moving Average Reversion. *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pp. 273–280, 2012.

Mak, W. K., Morton, D. P., and Wood, R. K. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1): 47–56, 1999.

Markowitz, H. Portfolio Selection. *The Journal of Finance*, 7(1):77–91, 1952.

Michaud, R. Efficient asset management: a practical guide to stock portfolio management and asset allocation. *Financial Management Association, Survey and Synthesis Series. HBS Press, Boston, MA*, 1998.

Michaud, R. O. The markowitz optimization enigma: Is optimized optimal? *ICFA Continuing Education Series*, 1989(4):43–54, 1989.

Qiu, H., Han, F., Liu, H., and Caffo, B. Robust portfolio optimization. In *Advances in Neural Information Processing Systems*, pp. 46–54, 2015.

Scherer, B. Portfolio resampling: Review and critique. *Financial Analysts Journal*, pp. 98–109, 2002.

Smith, J. E. and Winkler, R. L. The optimizers curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, 52(3):311–322, 2006.

Sutton, R. S. and Barto, A. G. [Draft-2] Reinforcement learning : an introduction. *Neural Networks IEEE Transactions on*, 9(5):1054, 2013.

Thomas, D. J., Thomas, D. J., Griffin, P. M., and Griffin, P. M. Coordinated supply chain management. *European Journal of Operational Research*, 94(1):1–15, 1996.

van der Vaart, A. W. Asymptotic Statistics. *Asymptotic Statistics*, 3:443, 1998.

Vapnik, V. *The nature of statistical learning theory*. Springer science & business media, 2013.

Yabe, A., Ito, S., and Fujimaki, R. Robust quadratic programming for price optimization. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, 2017.