Prescriptive price optimization using optimal regression trees

Shunnosuke Ikeda[1], Naoki Nishimura[1], Noriyoshi Sukegawa[2] and Yuichi Takano[2*]

[1*]Product Development Management Office, Recruit Co., Ltd., 1-9-2 Marunouchi, Chiyoda-ku, 100-6640, Tokyo, Japan.
[2]Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba-shi, 305-8573, Ibaraki, Japan.

*Corresponding author(s). E-mail(s): ytakano@sk.tsukuba.ac.jp;

Abstract

This paper focuses on prescriptive price optimization, which derives the optimal pricing strategy that maximizes future revenue or profit by using demand forecasting models for multiple products. Prescriptive price optimization requires accurate demand forecasting models because the accuracy of these models has a direct impact on pricing strategies aimed at increasing revenue or profit. However, existing methods have not been able to fully exploit them due to computational limitations. Therefore, the purpose of this paper is to establish a new prescriptive price optimization model that utilizes highly accurate demand forecasting models and can be solved exactly in realistic time. The prescriptive price optimization problem can be formulated as a mixed integer nonlinear optimization (MINLO) problem by using optimal regression trees as the demand forecasting model, which have a generalization performance as high as that of gradient boosting trees without losing interpretability. Although the MINLO problem is hard to solve, we reformulate it as a mixed integer linear optimization (MILO) problem by using exact linearization. This allows the problem to be solved exactly by an optimization solver. The effectiveness of our method is evaluated through simulation experiments by comparing with existing methods. Our simulation results demonstrate that our method improves the accuracy of demand forecasting and pricing strategies compared to existing methods.

Keywords: price optimization, optimal regression trees, mixed-integer optimization

1 Introduction

The price of products and services plays an important role in determining consumer demand, and implementing appropriate pricing strategies can maximize a firm's revenues and profits. As such, the problem of price optimization represents a crucial decision-making task for firms and has been extensively studied as a central topic in the field of marketing. Indeed, it has been successfully applied in a wide range of industries (retail [1, 2], car rental [3, 4], hotels [5, 6], internet providers [7], passenger railways [8], cruise lines [9], electricity[10–12], etc.), including the airline industry. (see [13], [14, 15] for a more comprehensive survey on revenue management and price optimisation respectively). Recent advances in information and communication technology have made it possible to reflect consumer demand data in real time, thereby increasing the importance of pricing strategies in terms of their impact on revenues and profits. As a result, pricing optimization has become an increasingly important topic in various fields of study.

In general, the framework for price optimization comprises two stages: demand forecasting and optimal price determination. In the stage of demand forecasting, a regression model is employed to construct a demand forecasting model that reveals the relationship between prices and the demand for products, including price elasticity and cross-price elasticity. Subsequently, in the stage of optimal price determination, an optimization problem is solved to determine the price that maximizes a given utility function, such as total revenue or profit, based on the obtained demand forecasting models. Previous research on price optimization has primarily focused on the relationship between the price of a single product and its demand. However, recent advances in regression technology have made it possible to forecast demand for a large number of products, and data-driven demand forecasting that takes into account cross-price elasticities of multiple products has also attracted attention. This framework for finding optimal pricing strategies, particularly by integrating demand forecasting models and mathematical optimization for multiple products, is referred to as prescriptive price optimization. [16].

There have been several studies on prescriptive price optimization [16–20]. The methods used in these studies, which target fast fashion [17] and hotels [18, 19], are domain-specific and have strong restrictions in demand modeling capability, e.g., making it difficult to generalize them to accross industries. To address this, Ito et al. [16, 20] established a framework for prescriptive price optimization in a general problem setting. Specifically, they used parametric linear regression with feature transformations to obtain demand forecasting models for multiple products. Then, based on the regression models, they formulated the revenue maximization problem as a binary quadratic optimization (BQO) problem and proposed algorithms to solve it efficiently. However, it is recognized that the relationship between demand and price can take various shapes depending on changes in purchasing conditions, such season and competitive prices, so it is desirable to use a more expressive and flexible forecasting model.

In price optimization, accurately modeling the relationship between price and demand is critical to increasing revenues and profits. Therefore, more flexible and accurate regression models are needed. In recent years, gradient boosting tree-based methods, such as LightGBM [21] and XGBoost [22], have been shown to have high generalization performance on a variety of datasets. However, when applied to demand forecasting models, these gradient boosting tree-based regression models present difficulties in formulating them as solvable optimization problems, requiring an exhaustive search of all possible price combinations to determine the optimal price for multiple products. Ferreira et al. [23] employed regression trees with bagging as demand forecasting models to improve prediction accuracy, but at the expense of not considering cross-price elasticities in addressing this challenge.

In this paper, we propose a novel prescriptive price optimization model that can efficiently obtain an exact solution while utilizing a highly accurate regression model. Specifically, we formulate the prescriptive price optimization problem as a mixed-integer nonlinear optimization (MINLO) problem by using optimal regression trees [24] as the demand forecasting models. These models are expected to have high generalization performance, comparable to that of gradient boosting trees, while maintaining interpretability. However, this formulation is hard to solve due to its non-convex objective function. To address this, we transform the problem into an equivalent mixed integer linear optimization (MILO) problem that can be solved exactly by using optimization solvers.

The effectiveness of our method is evaluated through simulation experiments. We compare the performance of our method with that of Ito et al.'s model [16, 20], a general-purpose prescriptive price optimization method. The simulation results demostrate that our method outperforms in terms of demand forecasting accuracy and the effectiveness of the pricing strategy. The main contributions of this paper are the following.

- We formulate the prescriptive price optimization problem as a MINLO problem by using optimal regression trees as demand forecasting models. This allows the problem to be described as a mathematical optimization problem even when a highly accurate regression model is used, opening up new possibilities for determining pricing strategies.
- We reformulated the hard-to-solve MINLO problem as a MILO problem by using exact linearization, which can be solved efficiently by an optimization solver, allowing the exact solution to be obtained in realistic time.
- We conduct simulation experiments to verify the effectiveness of the proposed methods in terms of forecasting accuracy and pricing strategy. The results indicate that the proposed method is effective and provide valuable insights for future research.

2 Methods

In this section, we first provide a brief overview of the general formulation of prescriptive price optimization as presented in [16, 20]. We then present our proposed method for prescriptive price optimization using optimal regression trees.

Prescriptive price optimization

Prescriptive price optimization is a framework for simultaneously determining optimal prices for multiple products based on demand forecasting models for each product. It consists of two steps: constructing demand forecasting models for each product, and solving an optimization problem on the basis of the demand forecasting models.

First, we explain the construction of the demand forecasting models for each product. Let $\mathcal{D}$ and $\mathcal{M}$ denote the set of indices of external variables and the set of products, respectively. Demand q_m is expressed by a linear regression model with price $p_m \in \mathcal{M}$ and external variables $g_d \in \mathcal{D}$ as explanatory variables. Let $\mathcal{N}$ be the set of arbitrary transformation types, and introduce an arbitrary transformation $\phi_{mn}(\cdot)$ $(n \in \mathcal{N})$ for price p_m, the model can be written as follows:

$$q_m(\mathbf{p}) = \sum_{m' \in \mathcal{M}} \sum_{n \in \mathcal{N}} \beta^{(m)}_{m'n} \phi_{m'n}(p_{m'}) + \sum_{d \in \mathcal{D}} \beta^{(m)}_d g_d + \beta^{(m)}_0 \tag{1}$$

where $\beta^{(m)}_0, \beta^{(m)}_{m'n}, \beta^{(m)}_d$ are the intercept, the coefficient of $\phi_{m'n}(p_{m'})$ and the coefficient of g_d, respectively, which are the parameters to be estimated. We also define $\mathbf{p} := (p_m)^{\mathrm{T}}_{m \in \mathcal{M}}$.

Next, we describe the determination of the pricing strategy by using the demand forecasting models calculated above. Denoting the cost of each product as c_m, the total profit can be expressed as follows:

$$\sum_{m \in \mathcal{M}} (p_m - c_m) q_m(\mathbf{p}). \tag{2}$$

In this paper, we focus on the case of total revenue maximization, i.e., $\mathbf{c} = \mathbf{0}$ $(\mathbf{c} = (c_m)^{\mathrm{T}}_{m \in \mathcal{M}})$. For the case of total profit maximization $(\mathbf{c} \neq \mathbf{0})$, if we replace $p'_m = p_m - c_m$, the total profit can be expressed as $\mathbf{p'}^{\mathrm{T}} \mathbf{q}(\mathbf{p'})$, which has the same mathematical structure. Therefore, it does not lose its generality.

Assuming that the price p_m is chosen from a set of $|\mathcal{K}|$ price candidates $\{P_{mk} \mid k \in \mathcal{K}\}$, the price optimization problem to maximize total revenue is formulated as follows:

$$\underset{\mathbf{p},\mathbf{q},\mathbf{x}}{\text{maximize}} \quad \sum_{m \in \mathcal{M}} p_m q_m(\mathbf{p}) \tag{3}$$

$$\text{subject to} \quad p_m = P_{mk} x_{mk}, \quad {}^{\forall} m \in \mathcal{M}, {}^{\forall} k \in \mathcal{K} \tag{4}$$

$$\sum_{k \in \mathcal{K}} x_{mk} = 1, \quad {}^{\forall} m \in \mathcal{M} \tag{5}$$

$$x_{mk} \in \{0, 1\}, \quad {}^{\forall} m \in \mathcal{M}, {}^{\forall} k \in \mathcal{K} \tag{6}$$

where x_{mk} is a binary variable that takes the value 1 if the price p_m is the price candidate P_{mk} and 0 otherwise. Furthermore, by substituting Eq (1) for the objective function (3), for any function ϕ_{mn}, the equation $\phi_{mn}(p_m) = \sum_{k \in \mathcal{K}} \phi_{mn}(P_{mk}) x_{mk}$ holds, the above formulation can be rewritten as the BQO problem as follows:

$$\underset{\mathbf{x}}{\text{maximize}} \quad \sum_{m \in \mathcal{M}} \sum_{m' \in \mathcal{M}} \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} \sum_{k' \in \mathcal{K}} \gamma^{(m)}_{m'nkk'} x_{mk} x_{m'k'} + \sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}} \delta^{(m)}_k x_{mk} \tag{7}$$

$$\text{subject to} \quad \text{Eqs } (5), (6) \tag{8}$$

where $\gamma^{(m)}_{m'nkk'}$ and $\delta^{(m)}_k$ are defined by the following equations.

$$\gamma^{(m)}_{m'nkk'} = \beta^{(m)}_{m'n} P_{mk} \phi_{m'n}(P_{m'k'}) \tag{9}$$

$$\delta^{(m)}_k = P_{mk} \left(\beta^{(m)}_0 + \sum_{d \in \mathcal{D}} \beta^{(m)}_d g_d \right) \tag{10}$$

To efficiently solve this BQO problem, Ito et al. proposed an algorithm using SDP relaxation and a proximity gradient algorithm based on network flow optimization in [16] and [20], respectively. BQO is also a challenging problem to solve because the objective function contains a product of variables term and is nonconvex. However, it can be equivalently transformed into a binary linear optimization (BLO) problem by variable transformation, and thus can be solved exactly by an optimization solver up to a certain scale (see [16] for more details).

In price optimization, accurate forecasting of demand is critical to maximizing revenues and profits, so the use of more expressive forecasting models is desirable. While linear regression models with feature transformations are used to forecast demand in the above prescriptive price optimization, the use of more expressive forecasting models could potentially lead to further improvements in revenues and profits.

2.1 Our Prescriptive Optimization Framework

In this paper, we propose a prescriptive price optimization model that can be solved efficiently in realistic time while utilizing a highly accurate demand forecasting model. Specifically, we formulate the prescriptive price optimization problem as a MINLO problem by using optimal regression trees for demand forecasting. This allows the problem to be described as a mathematical optimization problem while ensuring high forecasting accuracy. Since the MINLO problem is difficult to solve in its original form, we use techniques to reformulate it into a MILO problem, which is easier to solve.

2.1.1 Demand Forecasting Model

As for our demand forecasting, we begin with a brief review of optimal regression trees used for demand forecasting, and then describe their specific forecasting structure.

Optimal regression trees are a method for the regression task of optimal decision trees, proposed in 2018 by Dunn et al.[24]. Optimal decision trees are a type of decision trees used for regression and classification tasks, and are a method for constructing globally optimal decision trees that also consider future branches, as opposed to CART (Classification and Regression Trees)[25], which greedily branches top-down. This idea has been considered since the early 1970s, but with recent advances in optimization algorithms and improvements in computing power, research has been particularly active since the publication of OCT [26] and DTIP [27], which use a mixed integer optimization approach. The advantage of optimal decision trees is that they achieve high prediction accuracy without losing interpretability because only one tree is constructed. In particular, prediction accuracy is reported to be higher than that of random forests [28] and comparable to that of gradient boosting decision trees. For a more detailed review of optimal decision trees, see [29, 30].

Next, we describe the predictive structure of the optimal regression trees. If the maximum depth of the tree is D, at most $|\mathcal{T}| = 2^{(D+1)} - 1$ nodes are created, and each node is represented as $t \in \{1, \ldots, |\mathcal{T}|\}$. For example, Figure 1 shows the optimal regression trees when the maximum depth is set to $D = 2$. Here, we denote the set of branch nodes and the set of leaf nodes by $\mathcal{T}_B = \{1, \cdots, \lfloor |\mathcal{T}|/2 \rfloor\}$ and $\mathcal{T}_L = \{\lfloor |\mathcal{T}|/2 \rfloor + 1, \cdots, |\mathcal{T}|\}$, respectively, so that $\mathcal{T}_B = \{1, 2, 3\}, \mathcal{T}_L = \{4, 5, 6, 7\}$ in the case of Figure 1. Let $\mathcal{I}$ denote a set of data instances, and then each instance $i \in \mathcal{I}$ starts at root node 1 at the top of the tree and branches according to the branching conditions at the branch nodes until it reaches a leaf node. Additionally, let $\mathcal{X}$ be a set of explanatory variables and let $\mathbf{x}_i \in \mathbb{R}^{|\mathcal{X}|}$ represent the explanatory variables for instance i. We consider a discrete decision variable $\boldsymbol{a}_t \in \mathbb{R}^{|\mathcal{X}|}$ which takes the values $0, 1$ and indicates the selection of a variable from the set $\mathcal{X}$ as the branching condition at the branch node t. The branching threshold of the explanatory variable chosen by $\boldsymbol{a}_t$ is denoted by the continuous decision variable $b_t \in \mathbb{R}$. Then, instances satisfying $\boldsymbol{a}_t^{\mathrm{T}} \mathbf{x}_i < b_t$ proceed to the bottom-left node, otherwise to the bottom-right node. Each leaf node t is assigned a regression model $\boldsymbol{\beta}_t^{\mathrm{T}} \mathbf{x}_i + \beta_{0t}$ with regression coefficient $\boldsymbol{\beta}_t \in \mathbb{R}^{|\mathcal{X}|}$ with intercept termβ_{0t}, and the prediction for each instance is computed by the regression model of the leaf nodes reached. For more details on the formulation of optimal regression trees, see [24].

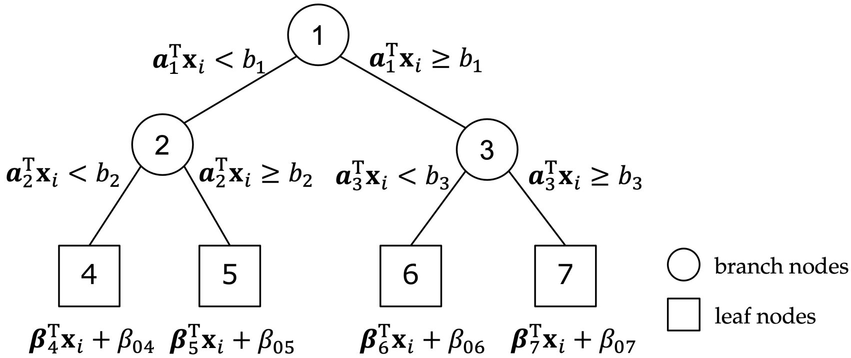

Fig. 1 An example of optimal regression trees for depth $D = 2$.

In our framework, we use optimal regression trees as demand forecasting models for each product m to predict demand q_m. However, due to the inclusion of 0-1 variables in the formulation of the optimal regression trees, the optimization solver can only solve it exactly for a medium-sized problem size. To address this issue, Dunn et al. proposed heuristics in [24] to solve each subproblem via local search, which is computationally efficient even for large datasets.

2.1.2 Formulations of Price Optimization

Based on the optimal regression trees for demand forecasting, we present a mathematical formulation of the price optimization problem. Specifically, we first present the MINLO formulation in its original form, followed by the MILO formulation using an exact linearization technique.

As the optimal regression trees are constructed prior to addressing the optimization problem, the variables that determine the structure of the optimal regression trees can be regarded as fixed constants. Symbols for the optimal regression trees are specific to each product, and are denoted by a superscript (m). For example, the optimal regression trees that forecast the demand q_m is denoted as $\mathrm{ORT}^{(m)}$.

MINLO Formulation

First, we naturally formulate the price optimization problem using optimal regression trees. Since the objective function is total revenue maximization, it can be expressed as the product of the price and demand of the product as follows:

$$\text{maximize} \sum_{m \in \mathcal{M}} p_m q_m. \tag{11}$$

As in the general price optimization problem, Eqs (4)–(6) are also considered to specify that the price p_m is chosen only from among the price candidates.

We now describe constraints derived from the optimal regression trees. The demand q_m is uniquely determined by $\mathrm{ORT}^{(m)}$ with the price of all products $\mathbf{p}$ and the external variables $\mathbf{g} := (g_d)^{\mathrm{T}}_{d\in\mathcal{D}}$ as inputs (explanatory variables). That is, the demand q_m satisfies all the branching conditions of $\mathrm{ORT}^{(m)}$ and reaches a single leaf node. These constraints can be described using big-M as follows:

$$\sum_{m'\in\mathcal{M}}\sum_{n\in\mathcal{N}} a^{(m)}_{m'nt'}\phi_{m'n}(p_{m'}) + \sum_{d\in\mathcal{D}^{(m)}} a^{(m)}_{dt'} g_d \geq b^{(m)}_{t'} - M_a(1 - z_{mt}), \quad {}^\forall t' \in \mathcal{R}^{(m)}_t, {}^\forall t \in \mathcal{T}^{(m)}_L, {}^\forall m \in \mathcal{M} \tag{12}$$

$$\sum_{m'\in\mathcal{M}}\sum_{n\in\mathcal{N}} a^{(m)}_{m'nt'}\phi_{m'n}(p_{m'}) + \sum_{d\in\mathcal{D}^{(m)}} a^{(m)}_{dt'} g_d + \epsilon^{(m)} \leq b^{(m)}_{t'} + M_a(1 - z_{mt}), \quad {}^\forall t' \in \mathcal{L}^{(m)}_t, {}^\forall t \in \mathcal{T}^{(m)}_L, {}^\forall m \in \mathcal{M} \tag{13}$$

$$\sum_{t\in\mathcal{T}^{(m)}_L} z_{mt} = 1, \quad {}^\forall m \in \mathcal{M} \tag{14}$$

$$z_{mt} \in \{0, 1\}, \quad {}^\forall t \in \mathcal{T}^{(m)}_L, {}^\forall m \in \mathcal{M} \tag{15}$$

where M_a is a large positive constant, i.e. big-M, and $\mathcal{R}^{(m)}_t$ is the set of ancestors of node t whose right branch has been followed on the path from the root node to node t, and similarly $\mathcal{L}^{(m)}_t$ is the set of left branch ancestors. Eq (12) means that at branch node t of $\mathrm{ORT}^{(m)}$, if the input satisfies the branching condition, it branches to the left. Conversely, equation (13) states that if the branching condition is not satisfied, it branches to the right. Since the optimization solver does not support exact inequalities, a small positive $\epsilon^{(m)}$ is added to the left side to avoid using them. In addition, Eq (14) specifies that the demand q_m reaches one of the leaf nodes. Finally, Eq (15) defines the variable z_{mt} as a 0-1 variable.

Finally, we add the following inequalities using big-M to apply the regression model at the reached leaf nodes.

$$q_m - \left(\sum_{m'\in\mathcal{M}}\sum_{n\in\mathcal{N}} \beta^{(m)}_{m'nt}\phi_{m'n}(p_{m'}) + \sum_{d\in\mathcal{D}^{(m)}} \beta^{(m)}_{dt} g_d\right) \geq -M^{(m)}_q(1 - z_{mt}), \quad {}^\forall t \in \mathcal{T}^{(m)}_L, {}^\forall m \in \mathcal{M} \tag{16}$$

$$q_m - \left(\sum_{m'\in\mathcal{M}} \beta^{(m)}_{m'nt}\phi_{m'n}(p_{m'}) + \sum_{d\in\mathcal{D}^{(m)}} \beta^{(m)}_{dt} g_d\right) \leq M^{(m)}_q(1 - z_{mt}), \quad {}^\forall t \in \mathcal{T}^{(m)}_L, {}^\forall m \in \mathcal{M} \tag{17}$$

where M_q is a large positive constant. Thanks to these constraints, only one regression model is assigned to the demand q_m when the leaf node t is reached, i.e., when $z_{mt} = 1$, and the following equation holds.

$$q_m = \sum_{m' \in \mathcal{M}} \sum_{n \in \mathcal{N}} \beta_{m'nt}^{(m)} \phi_{m'n}(p_{m'}) + \sum_{d \in \mathcal{D}^{(m)}} \beta_{dt}^{(m)} g_d \tag{18}$$

To summarize, the prescriptive pricing problem using optimal regression trees can be written as a MINLO problem as follows:

$$\begin{aligned} \underset{\mathbf{p},\mathbf{q},\mathbf{x},\mathbf{z}}{\text{maximize}} \quad & \sum_{m \in \mathcal{M}} p_m q_m \\ \text{subject to} \quad & \text{Eqs (4)–(6), (12)–(17)} \end{aligned} \tag{19}$$

where we define $\mathbf{q} := (q_m)^{\mathrm{T}}_{m \in \mathcal{M}}$ and $\mathbf{x} := (x_{mk})^{\mathrm{T}}_{m \in \mathcal{M}, k \in \mathcal{K}}, \mathbf{z} := (z_{mt})^{\mathrm{T}}_{m \in \mathcal{M}, t \in \mathcal{T}_L^{(m)}}$. This proposed formulation is consistent with the general prescriptive price optimization model, especially when the tree depth is set to $D = 0$, as it results in the presence of a single leaf node (root node) and the application of a single linear regression model to the demand variable q_m. Therefore, our formulation can be considered as an extension of the existing one.

3 Experimental results and discussion

In this section, we describe simulation experiments to evaluate the effectiveness of this method.

3.1 Experimental design

We outline the procedures for simulation experiments to assess the effectiveness of our proposed method. In our simulation experiments, we compare the performance of our proposed method to that of the existing method outlined in [16] using a synthetic dataset. To ensure an accurate comparison, the formulations of the existing method are compared using the BLO problem, for which exact solutions can be derived.

The synthetic dataset is generated from two types of generative models: the first is the linear regression model used for forecasting in the existing method, and the second is the optimal regression trees.

First, we describe the former generative model. Following the previous studies[16, 20], the demand q_m was generated from the following regression model.

$$q_m = \sum_{m \in \mathcal{M}} \beta_{m'}^{(m)*} p_{m'} + \beta_0^{(m)*} + \varepsilon, \quad {}^{\forall} m \in \mathcal{M} \tag{20}$$

where let $\varepsilon \sim N(0, \sigma^2)$ and σ^2 be the variance of ε. The noise level is defined as $\delta := \sqrt{\sigma^2 / E[q_m^2]}$, which is the ratio of the amount of noise added to the value of the generated data. The true coefficients $\beta_{m'}^{(m)*}$ are generated by Gaussian random numbers, $\beta_{m'}^{(m)*} \sim N(-1,1)(m = m')$, $\beta_{m'}^{(m)*} \sim N(1,1)(m \neq m')$. Additionally, the intercept $\beta_0^{(m)*}$ is generated from a uniform random number generated from the range [100, 200]. The price P_m is uniformly sampled from the price candidates $\{0.8, 0.85, 0.9, 0.95, 1\}$ ($|\mathcal{K}| = 5$).

The latter generative model is described next. Since the optimal regression trees assign a linear regression model to each leaf node, the true coefficients for that part of the tree are generated in the same way as in the generative model described above. Thus, the demand q_m at each leaf node is as follows.

$$q_m = \sum_{m \in \mathcal{M}} \beta_{m't}^{(m)*} p_{m'} + \beta_{0t}^{(m)*} + \varepsilon, \quad {}^{\forall} m \in \mathcal{M}, {}^{\forall} t \in \mathcal{T}_L \tag{21}$$

The optimal regression trees differ from the first linear regression model in that there is a conditional branch to reach a leaf node, and the parameters for the conditional branch must also be generated. Specifically, it is necessary to determine $a_{m't}^{(m)}$ and $b_t^{(m)}$ for the next branching condition.

$$\sum_{m' \in \mathcal{M}} a_{m't}^{(m)} p_{m'} < b_t^{(m)}, \quad {}^{\forall} m \in \mathcal{M}, {}^{\forall} t \in \mathcal{T}_B \tag{22}$$

$$\sum_{m' \in \mathcal{M}} a_{m't}^{(m)} p_{m'} \geq b_t^{(m)}, \quad {}^{\forall} m \in \mathcal{M}, {}^{\forall} t \in \mathcal{T}_B \tag{23}$$

The coefficient $a_{m't}^{(m)}$ is used to indicate the variable used for branching. The coefficient corresponding to a random sample from the set of variables $\{p_{m'} | m' \in \mathcal{M}\}$ is set to 1, while the remaining coefficients are set to 0. As for the coefficient $b_t^{(m)}$, which represents the threshold of the branching condition, it is randomly sampled from the set of possible values of the variable chosen at random when determining $a_{m't}^{(m)}$, with the exception of the minimum and maximum values. In this case, the possible price candidates are $0.8, 0.85, 0.9, 0.95, 1$, thus the coefficient $b_t^{(m)}$ is a randomly chosen value from the set $0.85, 0.9, 0.95$. The minimum and maximum values are excluded because if they are chosen as thresholds, there will be no data satisfying only one of the branching conditions, and the branching will be meaningless.

We explain the indicators and symbols used to evaluate pricing strategies in simulation experiments. Let $\mathbf{w} \in \mathcal{W}$ denote the decision variable, and let $f^*(\mathbf{w})$ and $\hat{f}(\mathbf{w})$ represent the objective functions calculated using the true parameters and those estimated by the regression model, respectively. The corresponding optimal solutions can be expressed as follows:

$$\mathbf{w}^* = \underset{\mathbf{w} \in \mathcal{W}}{\operatorname{argmax}} \ f^*(\mathbf{w}) \tag{24}$$

$$\hat{\mathbf{w}} = \underset{\mathbf{w} \in \mathcal{W}}{\operatorname{argmax}} \ \hat{f}(\mathbf{w}). \tag{25}$$

Using the previously defined symbols, the performance index (PI) of the pricing strategy, calculated on the basis of the estimated parameters, can be represented as follows:

$$\mathrm{PI} = \frac{f^*(\hat{\mathbf{w}})}{f^*(\mathbf{w}^*)} \ (\leq 1). \tag{26}$$

The estimation accuracy index (EI) of the revenue estimate can be represented as follows. Note that a value greater than 1 indicates an overestimation of revenue, while a value less than 1 indicates an underestimation.

$$\mathrm{EI} = \frac{\hat{f}(\hat{\mathbf{w}})}{f^*(\mathbf{w}^*)} \tag{27}$$

We evaluate the performance of the method by considering the aforementioned indicator, specifically due to its direct impact on actual revenue.

The optimal regression trees utilized in this method were constructed by using interpretableai (https://docs.interpretable.ai/stable/), a library published by interpretable AI. The hyperparameters were set as shown in Table 1. The max_depth, which represents the depth of the tree, was calculated by

Table 1 Hyperparameter settings for optimal regression trees.

Hyperparameter	Value
cp	0.0
regression_lambda	0.0
normalize_y	False
normalize_x	False
regression_weighted_betas	True
regression_sparsity	"all"
regression_features	"all"

splitting the generated data into training and validation data in a ratio of 7:3 and adopting the one with the smallest error with respect to the validation data. All other hyperparameters were used with their default settings. A re-training was also performed by combining the training data with the validation data. We used GUROBI Optimizer 9.5.1 (https://www.gurobi.com/), a state-of-the-art commercial solver for mathematical optimization, to solve the MILO and BLO problems. All experiments were performed on a machine equipped with Intel Core i7 @ 3GHz, 32GB RAM, and all computational results are averaged over 10 simulation experiments.

3.2 Evaluation of Results

Figure 2 and 3 are the results of calculating the ratio of optimal values for number of data when linear regression and optimal regression trees are set up for the generative model, respectively. Here, ORT and LR refer to this method and the existing method, respectively. In Figure 2, both our method and the existing one show that both PI and EI asymptotically approach 1 as the number of data increases, which is consistent with the optimal solution. When the number of data is small, the method overestimates the objective function, and its performance is slightly inferior to the existing method. This is due to the fact that when tuning the depth of the tree, a small number of data is susceptible to noise, and the demand forecasting model is overfitted by selecting a deeper tree. However, as the number of data increases, the effect of noise becomes less pronounced, and the results are comparable to those of the existing method using linear regression models. In addition, as shown in Figure 3, as the number of data points increases, both PI and EI for our method converge to 1, and an optimal solution is derived in most cases. In contrast, for the existing method, both values remain relatively constant around 0.95 even as the number of data increases. This is due to the fact that the original generative model has a complex structure, while the existing method employs linear regression, which has limited expressiveness. This results in an inability to estimate demand well, leading not only to an underestimation of the objective function, but also to poor performance overall.

Figures 4 and 5 show the results of the prediction accuracy when linear regression and optimal regression trees are used as the generative model, respectively. In this experiment, the Mean Absolute Percentage Error (MAPE) was used to evaluate the prediction accuracy. This is because it allows for easy comparative evaluations, even when the number of demands vary among products. Figures 4 and 5 present the average MAPE for all products. By comparing Figures 2, 3, and 4, 5, it can be observed that the performance of the pricing strategy improves as the accuracy of the prediction on the test data increases. Therefore, it can be confirmed that with approximately 1,000 data points, our method performs as well or better than the existing method for both generative models.

Figures 6 and 7 show the results of calculating the ratio of the optimal values for each noise level when linear regression and optimal regression trees are used as the generative model, respectively. In Figure 6, both PI and EI are relatively similar between our method and the existing one from 0 to about 0.4, at which point a value close to the optimal solution can be derived. As the noise level increases beyond 0.6, our method slightly overestimates the revenue, and the performance of the method also decreases. However, it is worth noting that a scenario where the error exceeds 60% of the mean is not a significant concern, as the demand forecasting in such a situation would be unreliable and its practical application would be unrealistic. In Figure 7, it can be observed that the PI of both methods deteriorates as the noise level increases. However, our method shows better performance at all noise levels. Additionally, it can be seen that the overestimation is more sensitive to the noise level for our method, indicating that the objective function is overestimated as the noise level increases. This could be attributed to the fact that the more complex model is more sensitive to noise and the prediction model overfits to the noise.

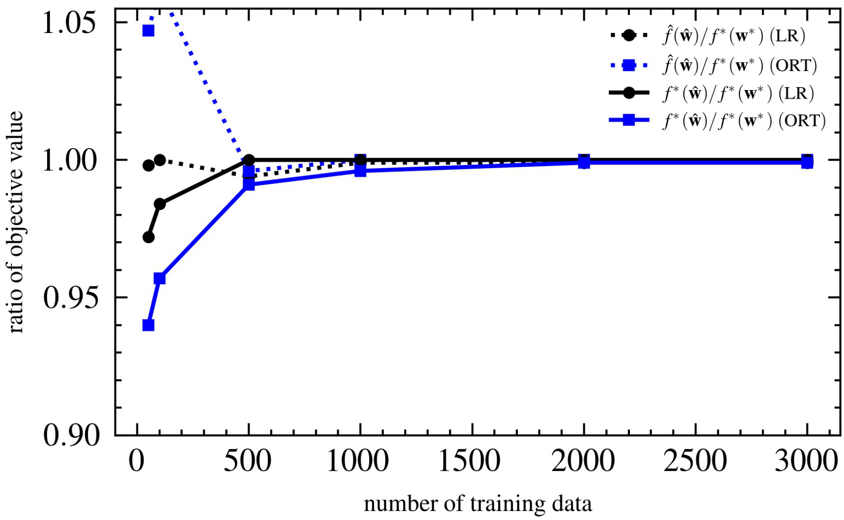

Fig. 2 Comparison of ratios of optimal values for number of data (generative model: linear regression).
$(\delta = 0.2, |\mathcal{M}| = 5, |\mathcal{K}| = 5)$

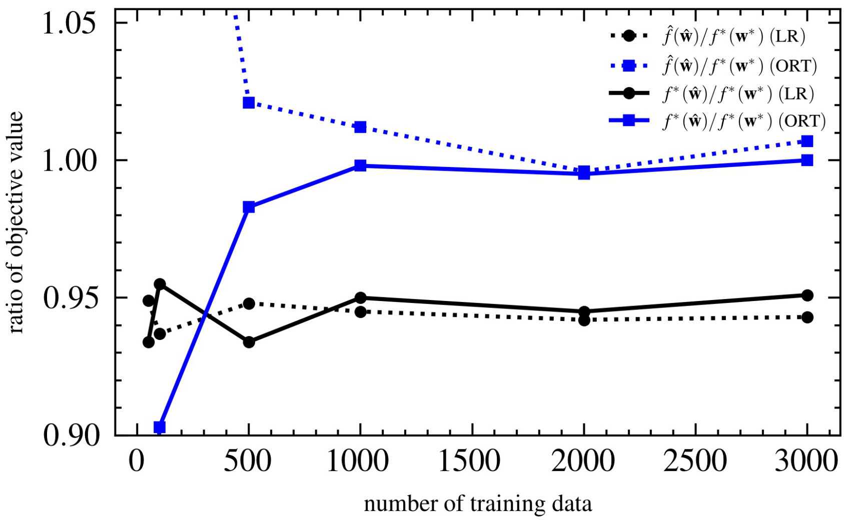

Fig. 3 Comparison of ratios of optimal values for number of data (generative model: optimal regression trees).
$(\delta = 0.2, |\mathcal{M}| = 5, |\mathcal{K}| = 5, D = 2)$

These results confirm that our method performs as well as or better than the existing method for both generative models within the practical range.

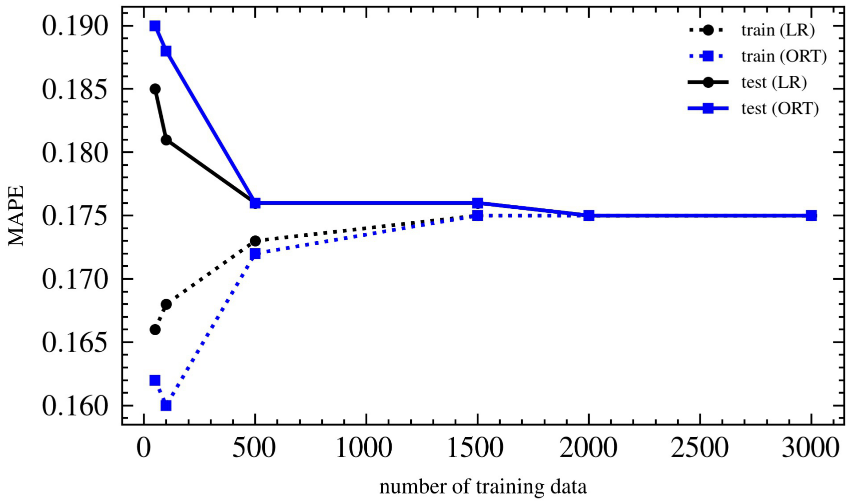

Fig. 4 Comparison of prediction accuracy for number of data (generative model: linear regression).
($\delta = 0.2, |\mathcal{M}| = 5, |\mathcal{K}| = 5$, # of test data: 3000)

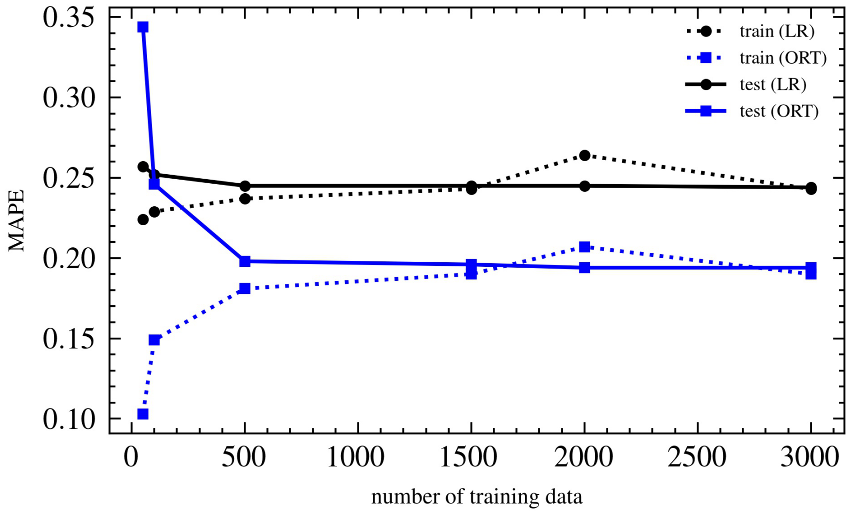

Fig. 5 Comparison of prediction accuracy for number of data (generative model: optimal regression trees).
($\delta = 0.2, |\mathcal{M}| = 5, |\mathcal{K}| = 5, D = 2$, # of test data: 3000)

Figures 8 and 9 show the results of calculating the ratio of optimal values for each number of products when linear regression and optimal regression trees are used as the generative model, respectively. In Figure 8, it can be observed that both PI and EI are close to 1 for both our method and the

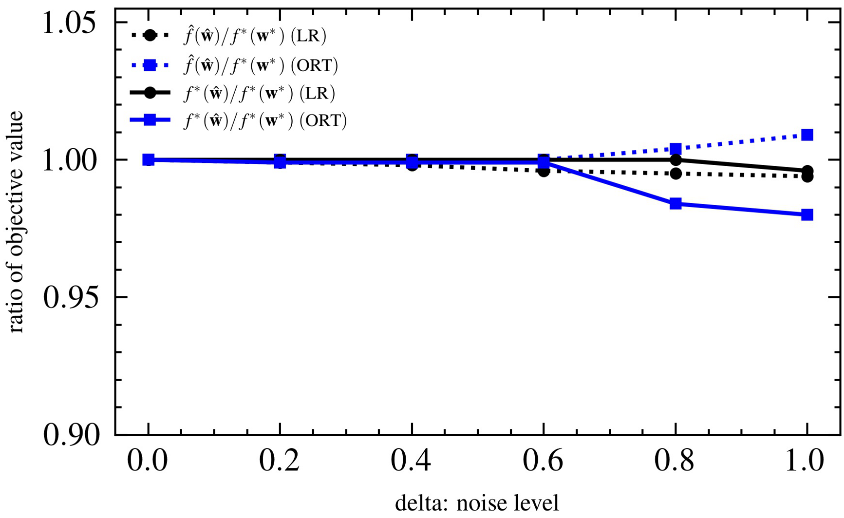

Fig. 6 Comparison of ratios of optimal values for noise level (generative model: linear regression).
($|\mathcal{M}| = 5, |\mathcal{K}| = 5$, # of training data: 3000)

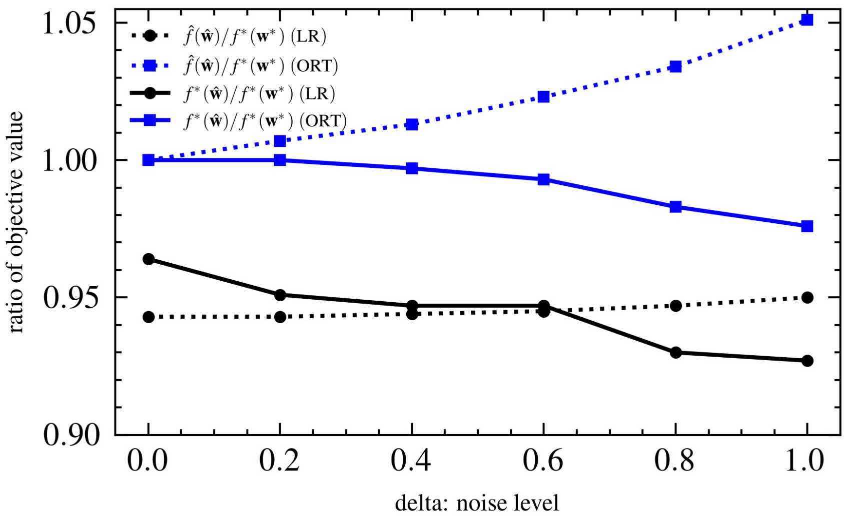

Fig. 7 Comparison of ratios of optimal values for noise level (generative model: optimal regression trees).
($|\mathcal{M}| = 5, |\mathcal{K}| = 5, D = 2$, # of training data: 3000)

existing one, even as the number of products increases. This indicates that our method is able to derive the same optimal pricing strategy as the existing one, even when the number of products increases. In Figure 9, it can be observed that our method consistently outperforms the existing one in terms of performance, regardless of the number of products. Additionally, it can be seen that the objective function tends to be overestimated by our method and underestimated by the existing one, even when the number of products increases. Therefore, it was found that the trend remained consistent even as the number of products increased, and our method performed as well or better than the existing one.

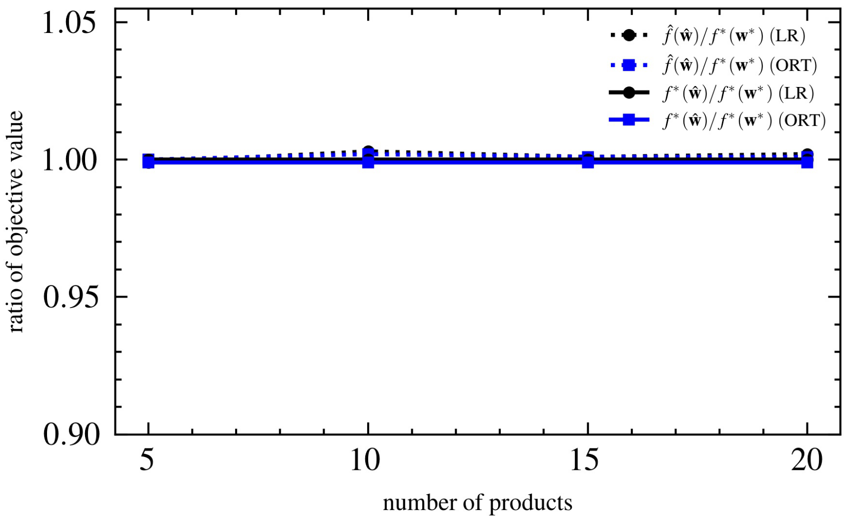

Fig. 8 Comparison of ratios of optimal values for number of product (generative model: linear regression).
($\delta = 0.2, |\mathcal{K}| = 5$, # of training data: 3000)

Finally, we provide some considerations for computation time. Figure 10 illustrates the results of computation time for each tree depth and number of products for our method, which is the result when solving the optimization problem using the true generative model. As previously mentioned, the plot shows how the computation time changes when the tree depth is varied from 0 to 5 and the number of products is 5, 10, 15, or 20. Note that when the tree depth is zero ($D = 0$), the computation time is the same as that of the existing method. In general, the computation time increases as both the tree depth and the number of products increase. This is mainly due to the increasing number of 0-1 discrete variables, but for tree depths of up to 3, even 20 products can be computed within a few seconds. For tree depths of 4 and 5, the computation time increases significantly, especially when the number of products is 20, but it can still be computed within a few minutes. Therefore, it can be used in practice without significant issues when the number of products is up to around 20.

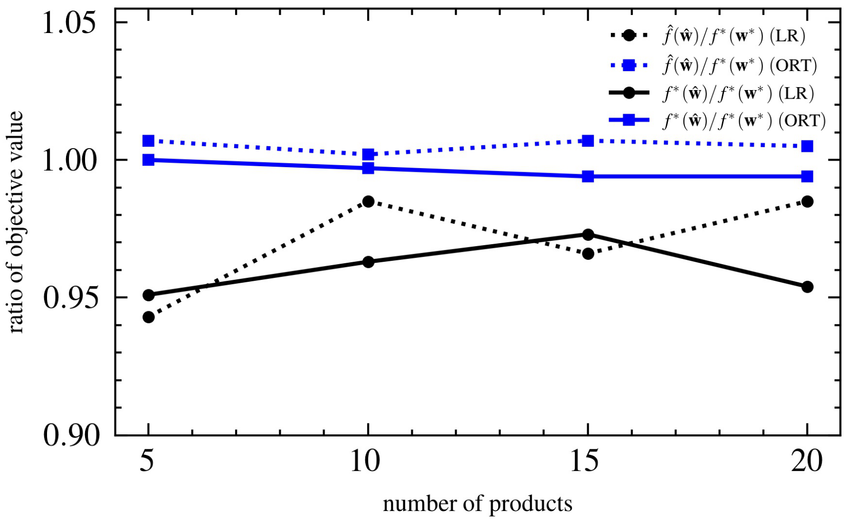

Fig. 9 Comparison of ratios of optimal values for number of product (generative model: optimal regression trees).
($\delta = 0.2, |\mathcal{K}| = 5, D = 2$, # of training data: 3000)

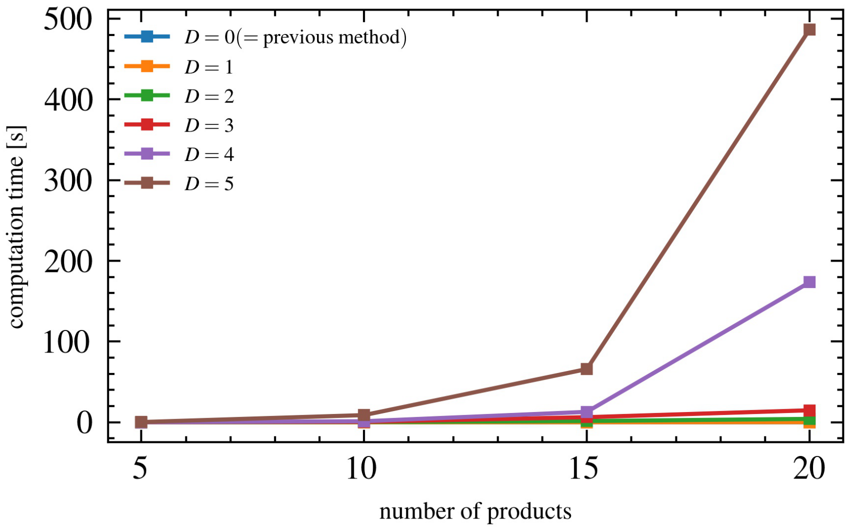

Fig. 10 Computaton time for depth of tree and number of products.

4 Conclusion

This paper addressed prescriptive price optimization, which finds a pricing strategy that maximizes revenue and profit based on demand forecasting models for multiple products. Accurately modeling of the relationship between price and demand is a critical factor in increasing revenue and profit, and therefore more flexible and expressive regression models should be used for demand forecasting. For this reason, we employed optimal regression trees as demand forecasting models, which have high generalization performance without losing interpretability. Then, we formulated prescriptive price optimization as a MINLO problem. However, since the MINLO problem is difficult to solve, we reformulated it as a MILO problem, which can be solved exactly by an optimization solver.

The effectiveness of our method was evaluated through simulation experiments, by comparing its performance to that of the existing one while varying the number of data, noise level, and number of products. The results of the simulation experiments confirmed that our method can construct a pricing strategy that is comparable to, or superior to the existing one, within a practical range. The computation time of the optimization problem was also examined and it was noted that the computation time increases significantly with the tree's depth, starting from around 20 products.

A future direction for this work is to incorporate uncertainty in the formulation. For example, an extension of the formulation to robust optimization, taking into account the estimation error of the regression coefficients, can be considered. Studies such as [31, 32] provide useful insights in this regard. Another area for further study is to improve the computational efficiency. Our method can be applied without significant issues for a moderate number of products, such as around 20, but for a larger number of products in the order of several hundred, the development of efficient heuristics is necessary. Additionally, we plan to apply our method to real-world services and evaluate its effectiveness using real data.

References

[1] Bitran, G.R., Mondschein, S.V.: Periodic pricing of seasonal products in retailing. Management Science **43**(1), 64–79 (1997)

[2] Subrahmanyan, S., Shoemaker, R.: Developing optimal pricing and inventory policies for retailers who face uncertain demand. Journal of Retailing **72**(1), 7–30 (1996)

[3] Carroll, W.J., Grimes, R.C.: Evolutionary change in product management: Experiences in the car rental industry. Interfaces **25**(5), 84–104 (1995)

[4] Geraghty, M.K., Johnson, E.: Revenue management saves national car rental. Interfaces **27**(1), 107–127 (1997)

[5] Bitran, G.R., Mondschein, S.V.: An application of yield management to the hotel industry considering multiple day stays. Operations Research **43**(3), 427–443 (1995)

[6] Bitran, G.R., Gilbert, S.M.: Managing hotel reservations with uncertain arrivals. Operations Research **44**(1), 35–49 (1996)

[7] Nair, S.K., Bapna, R.: An application of yield management for internet service providers. Naval Research Logistics (NRL) **48**(5), 348–362 (2001)

[8] Ciancimino, A., Inzerillo, G., Lucidi, S., Palagi, L.: A mathematical programming approach for the solution of the railway yield management problem. Transportation science **33**(2), 168–181 (1999)

[9] Ladany, S.P., Arbel, A.: Optimal cruise-liner passenger cabin pricing policy. European journal of operational research **55**(2), 136–147 (1991)

[10] Schweppe, F.C., Caramanis, M.C., Tabors, R.D., Bohn, R.E.: Spot Pricing of Electricity. Springer, ??? (2013)

[11] Smith, S.A.: A linear programming model for real-time pricing of electric power service. Operations Research, 470–483 (1993)

[12] Oren, S.S., Smith, S.A.: Service Opportunities for Electric Utilities: Creating Differentiated Products: Creating Differentiated Products;[papers Presented at a Symposium Held at the University of California at Berkeley, Sept. 12-14, 1990] vol. 13. Springer, ??? (1993)

[13] Klein, R., Koch, S., Steinhardt, C., Strauss, A.K.: A review of revenue management: Recent generalizations and advances in industry applications. European journal of operational research **284**(2), 397–412 (2020)

[14] Kunz, T.P., Crone, S.F.: Demand models for the static retail price optimization problem-a revenue management perspective. In: 4th Student Conference on Operational Research (2014). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik

[15] Bitran, G., Caldentey, R.: An overview of pricing models for revenue management. Manufacturing & Service Operations Management **5**(3), 203–229 (2003)

[16] Ito, S., 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)

[17] Caro, F., Gallien, J.: Clearance pricing optimization for a fast-fashion retailer. Operations Research **60**(6), 1404–1422 (2012)

[18] Koushik, D., Higbie, J.A., Eister, C.: Retail price optimization at intercontinental hotels group. Interfaces **42**(1), 45–57 (2012)

[19] Lee, S.: Study of Demand Models and Price Optimization Performance. Georgia Institute of Technology, ??? (2011)

[20] Ito, S., Fujimaki, R.: Large-scale price optimization via network flow. Advances in Neural Information Processing Systems **29** (2016)

[21] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y.: Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems **30** (2017)

[22] Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)

[23] Ferreira, K.J., Lee, B.H.A., Simchi-Levi, D.: Analytics for an online retailer: Demand forecasting and price optimization. Manufacturing & service operations management **18**(1), 69–88 (2016)

[24] Dunn, J.W.: Optimal trees for prediction and prescription. PhD thesis, Massachusetts Institute of Technology (2018)

[25] Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and Regression Trees. Routledge, ??? (2017)

[26] Bertsimas, D., Dunn, J.: Optimal classification trees. Machine Learning **106**(7), 1039–1082 (2017)

[27] Verwer, S., Zhang, Y.: Learning decision trees with flexible constraints and objectives using integer optimization. In: International Conference on AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems, pp. 94–103 (2017). Springer

[28] Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)

[29] Costa, V.G., Pedreira, C.E.: Recent advances in decision trees: an updated survey. Artificial Intelligence Review, 1–36 (2022)

[30] Carrizosa, E., Molero-Río, C., Romero Morales, D.: Mathematical optimization in classification and regression trees. Top **29**(1), 5–33 (2021)

[31] Hamzeei, M., Lim, A., Xu, J.: Robust price optimization of multiple products under interval uncertainties. Journal of Revenue and Pricing Management **21**(4), 442–454 (2022)

[32] Yabe, A., Ito, S., Fujimaki, R.: Robust quadratic programming for price

optimization. In: IJCAI, pp. 4648–4654 (2017)