On a Variation of Two-Part Tariff Pricing of Services: A Data-Driven Approach

On a Variation of Two-part Tariff Pricing of Services: A Data-Driven Approach

Georgia Perakis

MIT Sloan, Cambridge, MA 02139, georgiap@mit.edu

Charles Thraves

Departamento de Ingeniería Industrial, Universidad de Chile, cthraves@dii.uchile.cl

Problem definition: We present a data-driven pricing problem motivated from our collaboration with a satellite service provider. In particular, we study a variant of the two-part tariff pricing scheme. The firm offers a set of data plans consisting of a bundle of data at a fixed price plus additional data at a variable price. Most literature on two-part tariff problems focuses on models that assume full information. However, little attention has been devoted to this problem from a data-driven perspective without full information. One of the main challenges when working with data comes from missing data.

Methodology / Results: First we develop a new method to address the missing data problem, which comes from different sources: (i) the number of unobserved customers, (ii) customers' willingness to pay (WTP), and (iii) demand from unobserved customers. We introduce an iteration procedure to maximize the likelihood by combining the EM algorithm with a gradient ascent method. We also formulate the pricing optimization problem as a dynamic program (DP) using a discretized set of prices. From applying SAA, the DP obtains a solution within 3.8% of the optimal solution of the sampled instances, on average, and within 18% with 95% confidence from the optimal solution of the exact problem. By extending the DP formulation, we show it is better to optimize on prices rather than bundles, obtaining revenues close to optimizing jointly on both.

Managerial Implications: The sensitivity analysis of the problem parameters is key for decision-makers to understand the risks of their pricing decisions. Indeed, assuming a higher variability of customers' WTP induces higher revenue risks. In addition, revenues are barely (highly) sensitive to the customers' assumed WTP variability when considering a high (low) number of unobserved customers. Finally, we extend the model to incorporate price dependent consumption.

Key words: Two-part Tariff, Revenue Management, Missing Data, Pricing

1. Introduction

1.1. Motivation

Two-part tariff is a pricing strategy widely used in the supply chain and pricing literature. With this pricing strategy, the seller charges a fee for using a service and a price for each unit of the service consumed. In this work, we focus on a variant of this pricing structure that differs in two ways. First, the activation fee comes bundled with a fixed amount of consumption units. Second, the seller can offer multiple buying options under this pricing mechanism. The motivation to address this pricing strategy comes from our collaboration with one of the world's biggest satellite service providers, with revenues over $1 billion in 2018.

Due to an increasing demand for satellite data, it is critical for the company to determine the best way to update their prices for the data plans they will offer in the years to come. In addition, since there is almost no marginal cost for each unit of data consumed (downloaded), the pricing problem is crucial for the company's finances. In the satellite context, the firm charges a fee (called *fixed price*), which includes a certain amount of data (called *bundle data*), and a *variable price* is charged for those units consumed above the bundle data. Due to the high heterogeneity in customers' demand, the firm does not offer a single price but rather a variety of price contracts (called *plans*). The company needs to decide for each plan the particular fixed and variable prices. That is, having observed the demand from current customers, the firm would like to re-set the plans' prices in order to maximize its expected revenues. In contrast to prices in industries such as online sales, these plan prices cannot be changed frequently, as contracts with customers are established over a long time period. Price experimentation, thus, is not a feasible option. Therefore, the pricing decisions must be made with extreme care since these new prices will be responsible for the company's revenues in the years to come.

An important aspect of this problem relates to the estimation of unobserved data. As is the case in many revenue management problems, there are different sources of unobserved (or missing) data related to this problem that must be incorporated in order to determine a new pricing strategy. The challenge is not only the price optimization problem itself but also the issue of missing data. Missing data includes information that is not observed by the service provider, such as customers' willingness to pay and demand of potential customers who are currently not subscribed to any of the plans offered, but who might subscribe if prices were lower. (Throughout the paper, we will use the word *customers* to refer to *actual* and *potential customers*. In case we exclusively want to refer to customers who are actually purchasing a plan, we will refer to them as *actual* (or *current*, or *observed*) *customers*.) Once we integrate these missing data, we will be able to formulate the associated price optimization problem.

For the price optimization problem, we introduce a DP formulation. This formulation is further extended to address the joint optimization in prices and bundles, as with bundles alone. Thus, it is in the interest of this work to study how much additional revenue can be captured if the seller is allowed to optimize over both prices and bundles, as opposed to just one or the other. Given the stochastic nature of the problem, we use Sample Average Approximation (SAA) to solve several instances of the problem.

One of the goals in this paper is to determine near-optimal prices that attain revenues that outperform the revenues from the current plans set by the firm. In addition, it is important that the suggested prices are robust with respect to misspecifications in the model parameters. Hence, it is important to understand the impact in revenues of over- or underestimating the problem parameters (e.g., the share of unobserved customers or information on their willingness to pay) in order to understand the risks of the possible outcomes resulting from pricing decisions in this context.

1.2. Contributions

As mentioned in the previous subsection, this paper explores two main topics: estimation with missing data, and price optimization. We explore the proposed approach in collaboration with one of the biggest satellite service providers in the world. In addition, the proposed methodology can be applied to other industries that face a similar problem structure, such as energy, amusement parks, and credit card clubs. In these examples, consumers face different price options, which can be framed as a variation of the two-part tariff introduced in this work. For example, in an amusement park, a consumer can either (i) pay a fee for each visit, (ii) pay for a bundle that includes 5 visits at a certain price where additional visits are charged separately, or (iii) pay for a season pass that has no visit limit. The main contributions of this paper are summarized below.

- **Estimation in the Presence of Missing Data:** We propose a modeling framework that captures the main aspects of the problem, and a methodology that allows us to estimate the parameters of the problem distributions by combining the EM algorithm with a gradient ascent method in order to maximize the log-likelihood function. In addition, we show in the EM algorithm (E and M steps) how to compute the relevant quantities in closed form expressions.
- **DP formulation and solution for price, bundle, and price-bundle optimization:** We propose a DP for the pricing optimization problem, which can be solved in a polynomial number of steps using a price grid. In addition, we provide a second DP that computes an upper bound of this pricing DP. Furthermore, we present a DP method to solve the bundle as well as the price and bundle optimization problems. We show the impact in terms of revenues for each alternative for the instances solved and also for expected revenue.
- **Robustness of optimal prices:** We study how robust the optimal prices are in the cases in which we wrongly assume the model parameters. In particular, we show how the revenues are affected when the number of unobserved customers is under- or overestimated for different variability scenarios of customers' valuation. The analysis of the interplay between these two factors (i.e., the number of unobserved customers and the customers' WTP variability) is key to understand the revenue risks resulting from making the wrong assumptions.

1.3. Literature Review

Two-part tariffs were introduced more than a hundred years ago (Lewis (1941)). Numerous works have studied different aspects of this pricing strategy. Two-part tariffs were first adopted in the electricity industry in London at the end of the nineteenth century. This pricing strategy was introduced due to the motivation to raise prices during peak-demand periods, in which marginal costs were higher than those in regular demand periods. It was not possible at that time to charge a time-dependent price. Lewis (1941) provides more details on the origins of the applications of the two-part tariff pricing policy. In the following decades, other industry providers, such as telephone and gas, started to implement this pricing strategy. In a more recent work, Essegaier et al. (2002) state that this pricing mechanism has gained particular interest in industries that share some of the following characteristics: high-usage heterogeneity, low marginal costs, and capacity constraints. As a result, services such as credit card memberships, internet providers, and amusement parks have adopted this strategy to some extent. Essegaier et al. (2002) call these industries "access industries," as consumers pay for the privilege to access a facility but do not acquire any right to use the facility itself. One of the first works that captured the attention of researchers is that of Oi (1971). This work studied the case of two-part tariffs by Disneyland under a profit-maximization objective. The author concludes that the two-part tariff pricing structure acts as a discriminatory pricing mechanism enhancing the park's revenues. Similarly, Mitchell (1978) looks at the application of two-part tariffs by AT&T. Other authors, such as Fibich et al. (2017), have looked into the three-part tariff (which differs from the classical two-part tariff by charging an allowance at a fixed fee plus an entry fee) from an analytical perspective. From a data perspective, Lambrecht et al. (2007) study the impact of demand uncertainty under this pricing mechanism considering different consumer attributes, obtaining elasticities of choices and usage. Our work differs from this stream of research since we do not rely on customer data other than consumption; moreover, our focus is not only on the estimation but also on the pricing problem. Varian (1985) provides a more complete survey of different variations of the two-part tariff problem under a profit-maximization framework. In a more general setting, Iyengar and Gupta (2009) provide a survey of non-linear pricing, including two-part tariffs. The authors mention several applications and conclude that customer heterogeneity is key for such a pricing scheme. While some works have focused on the Pareto optimality of the problem (Feldstein (1972) and Ng and Weisser (1974)), others have focused on the welfare impact of two-part tariffs (Littlechild (1975) and Schmalensee (1981)). More recently, two-part tariff contracts have been studied in the supply chain context (see, e.g., Savaskan et al. (2004), Cachon and Lariviere (2005), Tsay et al. (1999), Bernstein and Marx (2005), and Oliveira et al. (2013)) or in collaborative production contracts between clients and vendors (see, e.g., Rahmani et al. (2017)). Other works have studied two-part tariffs with time-sensitive consumers in a queueing system (see, e.g., Afeche et al. (2019)). Most of these works assume a model in which the parameters are fully known and observable. This paper differs in the following three aspects: (i) the access fee gives the right to consume a given amount of the service, (ii) the firm can offer multiple plan options of the service, and (iii) we study the problem from a data-driven perspective where a firm does not have any additional demand information other than the demand of its current customers. In particular, using a model based on observable data usually entails non-observable elements, namely missing data, that must be taken into consideration.

The problem of missing data has been widely studied during the last sixty years in many different settings. Kaplan and Meier (1958) introduce a framework to analytically compute a distribution for a survivor function of censored data. Several other works have also studied nonparametric likelihood estimation problems, including Turnbull (1974) and Laird (1978). The EM algorithm was formally introduced (Dempster et al. (1977)) as a general way to compute a maximum likelihood estimate from incomplete data when the direct maximization problem is hard to solve. Until today, this algorithm has been widely used in the literature to address missing data problems through parametric and nonparametric models. This method has been used to estimate information that might not be observed by the firm directly, such as lost sales, product substitution, and reservation prices. Vulcano et al. (2012) use the EM algorithm to estimate lost sales and substitutions from observed sales transaction data using a discrete choice model. Kök and Fisher (2007) have developed a demand and substitution estimation model using the EM algorithm for assortment optimization. Zhu et al. (2013) use limited demand distribution information in a newsvendor setting in order to find the optimal ordering quantity, while others have studied the value of full demand information (Gallego and Özer (2002)). Other works have used Bayesian methods to estimate missing data (see Musalem et al. (2010)). Van Ryzin and Vulcano (2015) have developed a nonparametric procedure for estimating the probability density functions of customers' ranking profiles for arriving customers that face an available set of purchasing options. They use an iterative method by expanding the set of ranking profiles to maximize the log-likehood. This framework is not applicable in our case, since in our setting there is no variability in the available set, which is always composed by all the offered plans. In addition, in our problem we have additional information besides customers' choices, such as downloads, which can be incorporated into the model to explain customers' purchase decisions. Other works have introduced data-driven inventory policies from censored demand data, such as Huh et al. (2011) and Huh and Rusmevichientong (2009). Chen et al. (2015) have studied the joint inventory and pricing problem with lost sales in which the firm learns from the data while optimizing in an exploration exploitation procedure. Applications involving missing data are not restricted to customers' demand, but also apply to product life cycles (see, e.g, Hu et al. (2018)), or to patients' survival time in health care settings (see, e.g., Dag et al. (2016)).

Another stream of literature has focused on the pricing optimization problem. Rusmevichientong et al. (2006) have develop approximation algorithms for a pricing problem for items when the customer's purchase decision is based on ranking. Unfortunately, ranking-based models require a high number of parameters to be estimated. Works by Kohli and Mahajan (1991) and Jedidi and Zhang (2002) estimate the reservation prices of customers from the attributes of the products using conjoint analysis. Schlereth et al. (2010) have developed a mixed-integer nonlinear programming formulation for a two-part tariff and study it from an analytical perspective, comparing the performance of different heuristics. Dobson and Kalish (1988) and Shioda et al. (2011) study a pricing optimization problem for multiple items in the context of customer segments that differ by size and willingness to pay. These papers formulate the firm's pricing problem as an MIP and develop heuristics to solve it. Like these authors, we assume consumers choose the product that delivers the highest non-negative surplus; however, in our paper the prices for the various plans are determined by a combination of fixed and variable prices, instead, of by a single parameter. In addition, we address the problem by first estimating the problem parameters and then solving the pricing optimization problem. In our paper, customers have a fixed demand and a willingness to pay for the service. As a result, customer purchasing decisions will be the outcome of their demand and willingness to pay and the plans' prices. As in other revenue management papers (see, for example, Yılmaz et al. (2017)), we are interested in testing the performance of our model when we assume the wrong values for the parameters (in the remainder of the paper, we will refer to this as misspecifying the model parameters).

1.4. Paper Structure

The remainder of the paper is structured as follows. Section 2 describes the model and assumptions. The methodology we introduce to address the missing data problem is presented in Section 3. Section 4 describes in detail the price optimization problem and the DP formulation to solve it, along with extensions of the latter that address the price and bundle optimization problem. Section 5 describes the performance of the DP with respect to the optimal solution and current prices, including a sensitivity analysis of the assumed parameters. Section 6 shows the results from the price-dependent consumption model. In Section 7, we discuss our conclusions as well as directions for future research. Proofs and additional material are relegated to the Appendices.

2. Model

Consider a single satellite service provider that offers S different plans. Each plan $s \in \mathcal{S} := \{1, \dots, S\}$ is defined by three variables: (i) a bundle of data, b_s megabytes, (ii) a fixed price f_s, and (iii) a variable price v_s; in other words, a customer who purchases plan s will pay the fixed price f_s plus the variable price v_s for each megabyte downloaded above the bundle data b_s. Wlog, we assume that

$b_1 < b_2 < \cdots < b_S$. This suggests that a customer who is subscribed to a plan $s \in \mathcal{S}$ and downloads d megabytes in a month would incur a payment of $f_s + v_s \times \max\{d - b_s, 0\}$.

The firm currently has a set of plans that it offers to its customers. These can be described as the set of vectors $\{(f_s^0, v_s^0, b_s)\}_{s=1}^S$. We will refer to these plans as *current plans*. As mentioned in the previous section, the idea is to find the optimal values for the fixed and variable prices while preserving the current number of plans and current bundle data for each plan. The reason for focusing mainly on the pricing decisions is because the part of the company with which we collaborated has decision-making authority only on pricing, and not on the plan structure itself. Despite the latter, we also formulate and solve the joint optimization problem on prices and bundle (see Sections 4 and 5).

The current plans give rise to two types of customers: (i) observed customers and (ii) unobserved customers. *Observed customers* correspond to actual clients, i.e., those who are subscribed to one of the current plans. *Unobserved customers* are potential customers who are not subscribed to any of the current plans, although they might, if prices were lower. We denote by n the number of observed customers and by m the number of unobserved customers. As expected, we know the number of observed customers, but not the number of unobserved customers. Also, we assume that the number of customers does not vary from one month to another, and neither do the customers' downloads. This simplification will allow us to study the problem in a monthly horizon. From the company data, we notice that customers' demand across months does not differ significantly. Denote $\mathcal{I}_o$ and $\mathcal{I}_u$ as the set of observed and unobserved customers, respectively, with cardinalities n and m, and let $\mathcal{I} := \mathcal{I}_o \cup \mathcal{I}_u$ be the set of all customers. Each customer $i \in \mathcal{I}$ has a download value d_i and a willingness to pay w_i; both quantities are known by them. The willingness to pay is the maximum amount a consumer would pay for purchasing a plan. (We assume each customer has a fixed willingness to pay for her particular consumption level, which is independent of the price. Despite the fact that each customer's consumption is invariant with respect to the price, the aggregate demand of customers is price dependent. This will be more clear once the model is fully introduced in the next paragraphs.) Hence, a consumer is represented by the pair of her download data and willingness to pay. Customers are assumed to be rational, i.e., they choose the cheapest plan according to their level of download; however, they will only purchase this plan if the price of this cheapest plan does not exceed their willingness to pay. Let us define the function of *minimum payment* as $l(x, \mathbf{f}, \mathbf{v}) : \mathbb{R} \times \mathbb{R}^S \times \mathbb{R}^S \to \mathbb{R}$ so that, for each download amount x, the minimum payment to be incurred is given by $l(x, \mathbf{f}, \mathbf{v}) = \min_{s \in \mathcal{S}} f_s + v_s \times \max\{x - b_s, 0\}$. As a result, a customer $i \in \mathcal{I}$ will purchase plan s if and only if $w_i \geq l(d_i, \mathbf{f}, \mathbf{v})$ and $l(d_i, \mathbf{f}, \mathbf{v}) = f_s + v_s \times \max\{d_i - b_s, 0\}$. In case the customer is indifferent between two or more plans, we assume he/she chooses randomly among them. Customers subscribe to at most one plan. Figure 1 depicts the customers and plans in the price and download space. Customers who are willing to pay more than the price of any of the plans will purchase the cheapest plan, i.e., the customer in the top-left of Figure 1 will purchase plan $t-1$, while the two customers in the top-middle and top-right will purchase plan t. On the contrary, customers who are below the plans' price curve are those whose willingness to pay is lower than the lowest-priced plan offered. As a result, these customers will not purchase any plan.

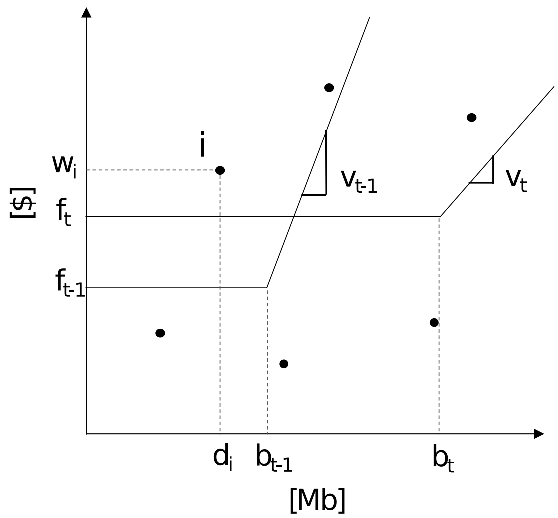

Figure 1 **Each customer is represented by a pair of download (x-axis) and willingness to pay (y-axis), while plans are represented by the piecewise linear functions. The top-left customer purchases plan $t-1$, the top-right and the top-middle customers purchase plan t, and the customers below the plans' price curve do not purchase.**

In setting the prices of the plans, there are important business rules that must be respected. First, a plan with a larger bundle data should have a higher fixed price than a plan with a smaller bundle data. Then, fixed prices are increasing in the index, thus $f_{s-1} \leq f_s$. The idea is that a plan that includes more consumption at a fixed price must have a higher price. Second, as opposed to fixed prices, variable prices should decrease with the bundle data that the plan offers, thus $v_{s-1} \geq v_s$. The rationale behind this rule is to ensure that plans with larger bundle data charge a cheaper rate per megabyte consumed. The third business rule is that the plan with the last index should be an "all you can eat" plan, i.e., bundle data is infinite for this plan ($b_S = \infty$). Finally, the bundle data of each plan must be attractive for some range of downloads. More specifically, each plan $s \in \mathcal{S}$ must have a range of at least c_s megabytes under which customers strictly prefer to purchase plan s versus the other plans. The idea behind this requirement is that the bundle data of each plan (megabytes that are essentially offered at a fixed price) should be a "good deal" for some consumers. Otherwise, the plan could be dominated (in the sense that there is another plan that is cheaper) for every download, making its existence pointless. The left panel of Figure 2 depicts the case in which the business rule is satisfied for plan t. In contrast, the right panel of Figure 2 illustrates a case in which the business rule is not met, resulting in plan t being irrelevant. Naturally, $c_s \in [0, b_s - b_{s-1}]$ for all $s \in \{2, \ldots, S-1\}$ and $c_1 = b_1$ since the plan with a smaller bundle data is the cheapest plan within the region of downloads in $[0, b_1]$. c_S is irrelevant due to $b_S = \infty$, and therefore is set to zero. We consider $c_s = (b_s - b_{s-1})/2$ for $s \in \{2, \ldots, S-1\}$ in this particular application. The latter business rule comes from the particular company that faced this problem. Note that the rule might lead to suboptimal solutions. (This is discussed in Section 4.)

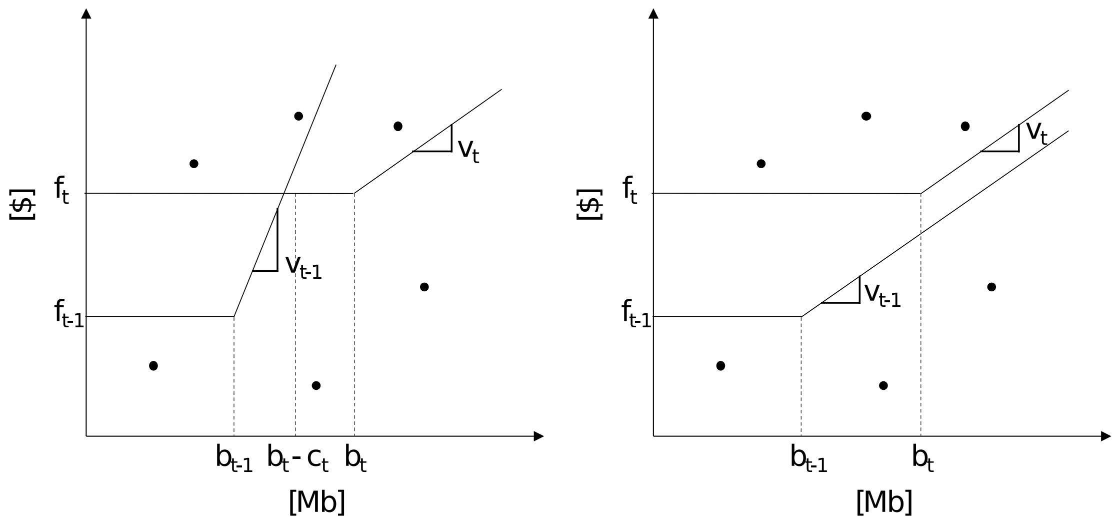

Figure 2 **Left: Case where the bundle data of plan t is attractive to some customers within a range of downloads. Right: Plan t bundle data at fixed price f_t is dominated by plan $t-1$; therefore, plan t's bundle data does not attract any consumers.**

The firm's problem is to find the optimal plan of prices that maximizes the expected revenue. Ideally, if we had full information on the pool of customers (their downloads and willingness to pay), we would directly formulate the pricing optimization problem. Unfortunately, this is not the case. For example, we know the consumption of current customers and the price they pay for this. However, we do not know their willingness to pay for this consumption. Thus, the only information that is captured by the seller are the payments and downloads of the observed customers. As a result, we cannot go directly into the optimization problem of the firm without first addressing the missing data. The next section provides the details on the methodology proposed to overcome this issue.

3. Missing Data and Estimation Procedure

As mentioned in Section 1.1, there are several sources of missing data, which can be summarized as follows: (i) the number of unobserved customers (m), (ii) the download demand of unobserved customers (d_i for $i \in \mathcal{I}_u$), and (iii) the willingness to pay of all customers (w_i for $i \in \mathcal{I}$).

3.1. Number of Unobserved Customers

The number of unobserved customers, who are currently not purchasing any plan but might purchase if prices were lower, is unfortunately unidentifiable in the model. Therefore, we consider scenarios that specify different numbers of these customers. This will allow us to analyze how optimized prices and revenues vary in the different cases. Let us define $\nu := m/n$ as the ratio between unobserved and observed customers. The values considered for the unobserved customers are $m = \nu n$ for $\nu \in \{1/3, 2/3, 1, 4/3, 5/3\}$. These scenarios were provided to us from the specific company with whom we collaborated and are based on estimates of the market size of the industries they serve and the available information from their competitors.

3.2. Downloads

We assume the downloads (D) of all customers, i.e., before knowing if they are observed or unobserved, follow a mixture of log-normal distributions with parameters denoted by $\tau := \{(\mu_j, \sigma_j, \eta_j)\}_{j=1}^{J}$. We choose this particular distribution as it seems a good fit for the downloads of observed customers (see Figure 3). Naturally, the parameters of the mixture distribution of downloads of all customers (observed and unobserved) might differ from the mixture fitted on the observed customers' downloads (shown in Figure 3). The procedure to estimate the values of the former mixture parameters is described in Section 3.4. In order to choose the number, J, of log-normal distributions in the mixture, we investigate the following four statistics pertaining to the downloads of observed customers: (i) BIC, (ii) log-likelihood, (iii) Cramer-Von Mises, and (iv) Kolmogorov-Smirnov; (ii), (iii), and (iv) are computed using cross-validation. Table 1 shows these statistics for different numbers of mixtures. The best BIC is attained with six mixtures, while for the other statistics the best value is achieved for nine mixtures. However, note that the improvement of statistics (ii), (iii), and (iv) is quite significant from one to six mixtures; yet it is much smaller from six to more mixtures. As a result, we chose to stay with six mixtures ($J = 6$).

J	BIC	Log-Likelihood*	Cramer-Von Mises*	Kolmogorov-Smirnov*
1	−363473.39	−18173.09	0.0038022	0.10831
2	−355807.81	−17788.46	0.0001638	0.02655
3	−355039.63	−17767.12	0.0001805	0.02722
4	−354497.98	−17720.38	0.0000614	0.01739
5	−354059.44	−17701.30	0.0000600	0.01754
6	−354037.52	−17694.80	0.0000567	0.01696
7	−354057.48	−17693.99	0.0000569	0.01693
8	−354069.63	−17693.06	0.0000561	0.01671
9	−354098.82	−17692.92	0.0000560	0.01647
10	−354124.64	−17693.45	0.0000562	0.01653

Table 1 **Statistics of log-normal mixtures over observed customers' downloads.**
***Computed using 10-fold cross-validation.**

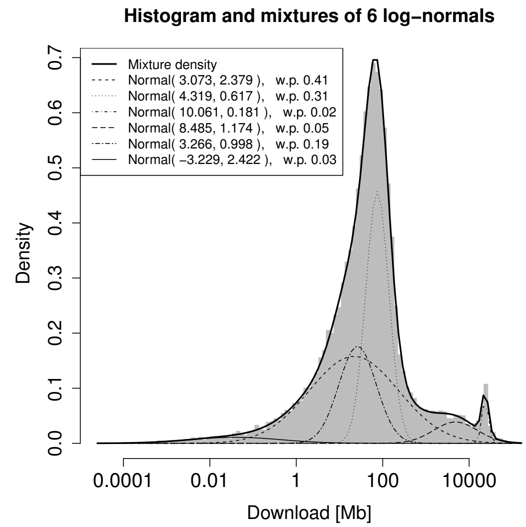

Figure 3 Histogram of densities of download values for observed customers and the fitted mixture with six log-normals.

3.3. Willingness to Pay

Customers' willingness to pay (W) is modeled through a parametric probability distribution. The parametric assumption has been widely used in operations management and the marketing literature (see e.g., Alberini (1995), and Miller et al. (2011)). One of the advantages of using a parametric distribution is its flexibility in relating its parameters with other variables of the problem, in our case the customers' WTP with their downloads. In the following, we describe in more detail this relation. As expected, the WTP distribution differs for different levels of downloads. Consequently, we assume that the distribution of customers' willingness to pay depends upon the download amount. Let $g : \Theta \times \mathbb{R}_+ \to \mathbb{R}^t$ be a function that maps the parameters to estimate $\theta \in \Theta \subset \mathbb{R}^u$ and a download d into the space of the distribution parameters $\xi \in \mathbb{R}^t$. For example, $W|D \sim \mathcal{N}(\theta_1 + \theta_2 D, (\theta_3 + \theta_4 D)^2)$, where $\xi = g(\theta, d) \in \mathbb{R}^2$ is the mean and standard deviation, and $\theta \in \mathbb{R}^4$ are the parameters that need to be estimated. We tried other distributions and found very similar results in terms of log-likelihood; therefore, we decided to stay with the Normal distribution (more details of this can be found in Section 3.4 and Appendix G). In order to ensure that the model is identifiable, we enforce the relation between the mean and standard deviation to be equal to a particular coefficient of variation, κ. Thus, the function g takes the form $g(\theta, d) = (\gamma(\theta, d), \kappa \cdot \gamma(\theta, d))$, where γ is a function for which the output is the mean of the normal distribution. Similar to the case of ν, we consider different scenarios on the coefficient of variation (κ) of the customers' willingness to pay, so we can analyze the outcomes in each setting. For convergence purposes of the parametric estimation (see more in Section 3.4), we assume $\gamma(\theta, d)$ is continuously differentiable on θ. In addition, we state the following assumption to ensure identifiability of the model:

Assumption 1. *If $\theta^1 \neq \theta^2$, then there exists $d_0 > 0$ such that $\gamma(\theta^1, d_0) \neq \gamma(\theta^2, d_0)$.*

We use a download and WTP distribution for all customers *a priori* knowing what type of customer they are (observed or unobserved). However, once conditioning either on the event of purchasing (for observed customers) or not purchasing (for unobserved customers), posterior distributions can be obtained for each customer type. Thus, downloads and WTP posterior distributions for each customer type will not be the same. The reason for not using separate distributions from the beginning (for observed and unobserved customers), is that the former approach (the one used in the paper) has less distributions to fit, and also we can leverage from the observed download values to estimate the general (*a priori*) distribution of downloads.

In Section 3.4, we explain the procedure used to estimate the distribution parameters τ and θ, and the selection of the parametric function $\gamma(\theta, d)$.

3.4. Estimation of the Parametric Model

In what follows, we describe the procedure used to estimate the parameters (τ and θ) of the probability distribution functions. Recall that the two distributions we are trying to estimate are the download distribution of all customers and the willingness to pay.

Denote D_i as the random variable of the customer i download (note that we already know that $D_i = d_i$ for $i \in \mathcal{I}_o$) and W_i her willingness to pay. To simplify the notation, let $l^0(x) := l(x, \mathbf{f^0}, \mathbf{v^0})$ be the minimum payment function with the current prices $(\mathbf{f^0}, \mathbf{v^0})$ for a given level of download x. Then, the log-likelihood can be written as follows:

$$\begin{aligned} ll(\tau, \theta) &= \sum_{i \in \mathcal{I}_o} \ln\left(\mathbb{P}\left(W_i \geq l^0(D_i) | D_i = d_i\right) f_D(d_i)\right) + \sum_{i \in \mathcal{I}_u} \ln\left(\mathbb{E}_{D_i}\left[\mathbb{P}\left(W_i < l^0(D_i)\right)\right]\right) \\ &= \sum_{i \in \mathcal{I}_o} \ln\left(1 - \Phi\left(l^0(d_i) | \gamma(\theta, d_i)\right)\right) + \sum_{i \in \mathcal{I}_o} \ln\left(\sum_{j=1}^{J} \phi(\ln(d_i) | \mu_j, \sigma_j)\eta_j\right) - \sum_{i \in \mathcal{I}_o} \ln(d_i) \\ &\quad + m \ln\left(\int_0^\infty \Phi\left(l^0(x) | \gamma(\theta, x)\right) \sum_{j=1}^{J} \frac{\phi(\ln(x) | \mu_j, \sigma_j)\eta_j}{x}\, dx\right), \end{aligned} \tag{1}$$

where f_D is the PDF of the downloads, and Φ (ϕ) represents the CDF (PDF) of a Gaussian distribution. In order to reduce the notation, we omitted writing the standard deviation ($\kappa\gamma(\theta, d)$) in the willingness to pay PDF and CDF. Before discussing the likelihood maximization, let us state a proposition that addresses the identifiability of the model.

Proposition 1. *Under Assumption 1, the model is identifiable besides from the relabeling of the mixture indexes.*

Proof. See Appendix A. □

Unfortunately, Expression (1) is hard to optimize directly on τ and θ. However, we can optimize separately on τ and θ, leading to two optimization problems that are more tractable. More precisely, the optimization problem on τ can be addressed using the EM algorithm, while the optimization on θ can be solved using a gradient ascent method (since the derivative with respect to θ can be computed easily).

Optimization on τ: A key idea of the EM algorithm is to lower bound the log-likelihood function. Denote Z_i as the random variable that represents the mixture where the download of observed customer i belongs, and Z as the analogous random variable for an unobserved customer. Unlike observed customers, unobserved customers are indistinguishable; therefore, we define a single representative random variable for them. Also consider $p_i(j)$ as the probability that an observed customer i belongs to mixture j, and $p(j,x)$ as the probability density function that an unobserved customer belongs to mixture j and downloads x. Let C_1 and C_2 be terms that absorb the quantities independent of τ, which, for now, we are not interested in (see Appendix B to see the exact expression of these terms). Then the log-likelihood, as a function of τ, can be expressed as follows:

$$
\begin{aligned}
ll(\tau) &= \sum_{i\in\mathcal{I}_o} \ln\left(\sum_{j=1}^{J}\phi(\ln(d_i)|\mu_j,\sigma_j)\eta_j\right) + C_1 + m\ln\left(\int_0^\infty \Phi\left(l^0(x)|\gamma(\theta,x)\right)\sum_{j=1}^{J}\frac{\phi(\ln(x)|\mu_j,\sigma_j)\eta_j}{x}dx\right)\\
&= \sum_{i\in\mathcal{I}_o} \ln\left(\sum_{j=1}^{J}p_i(j)\frac{\phi(\ln(d_i)|\mu_j,\sigma_j)\eta_j}{p_i(j)}\right) + C_1\\
&\quad + m\ln\left(\int_0^\infty \sum_{j=1}^{J}p(j,x)\Phi\left(l^0(x)|\gamma(\theta,x)\right)\frac{\phi(\ln(x)|\mu_j,\sigma_j)\eta_j}{p(j,x)x}dx\right)\\
&\geq \sum_{i\in\mathcal{I}_o}\sum_{j=1}^{J}p_i(j)\ln\left(\frac{\phi(\ln(d_i)|\mu_j,\sigma_j)\eta_j}{p_i(j)}\right) + C_1\\
&\quad + m\int_0^\infty \sum_{j=1}^{J}p(j,x)\ln\left(\Phi\left(l^0(x)|\gamma(\theta,x)\right)\frac{\phi(\ln(x)|\mu_j,\sigma_j)\eta_j}{p(j,x)x}\right)dx\\
&= \sum_{i\in\mathcal{I}_o}\sum_{j=1}^{J}p_i(j)\ln\left(\phi(\ln(d_i)|\mu_j,\sigma_j)\eta_j\right) + C_1 + C_2 + m\int_0^\infty\sum_{j=1}^{J}p(j,x)\ln\left(\phi(\ln(x)|\mu_j,\sigma_j)\eta_j\right)dx. \quad (2)
\end{aligned}
$$

The above inequality follows from Jensen's inequality.

- **E-Step:** We compute the expressions for $p_i(j)$ and $p(j,x)$ so that they match the posterior probability of the unobserved data, given the observed and assumed information for a fixed parameter $\tau^{(k)} = \{(\mu_j^{(k)},\sigma_j^{(k)},\eta_j^{(k)})\}_{j=1}^{J}$, where k represents a particular iteration.

$$
p_i(j) = \frac{\phi(\ln(d_i)|\mu_j^{(k)},\sigma_j^{(k)})\eta_j^{(k)}}{\sum_{h=1}^{J}\phi(\ln(d_i)|\mu_h^{(k)},\sigma_h^{(k)})\eta_h^{(k)}}, \quad (3)
$$

$$
p(j,x) = \frac{\Phi(l^0(x)|\gamma(\theta,x))\phi(\ln(x)|\mu_j^{(k)},\sigma_j^{(k)})\eta_j^{(k)}x^{-1}}{\int_0^\infty\sum_{h=1}^{J}\Phi(l^0(y)|\gamma(\theta,y))\phi(\ln(y)|\mu_h^{(k)},\sigma_h^{(k)})\eta_h^{(k)}y^{-1}dy}. \quad (4)
$$

Details are provided in Appendix C.

- **M-Step:** Setting the FOC to zero with respect to μ_j, σ_j, and η_j on the Lagrangian of Equation (2) (namely Equation (2) plus the term $\lambda\left(\sum_{j=1}^{J}\eta_j - 1\right)$ due to the constraint $\sum_{j=1}^{J}\eta_j = 1$), leads to:

$$\mu_j = \frac{\sum_{i\in\mathcal{I}_o} p_i(j)\ln(d_i) + m\int_0^\infty p(j,x)\ln(x)\,dx}{\sum_{i\in\mathcal{I}_o} p_i(j) + m\int_0^\infty p(j,x)dx} \tag{5}$$

$$\sigma_j^2 = \frac{\sum_{i\in\mathcal{I}_o} p_i(j)(\ln(d_i)-\mu_j)^2 + m\int_0^\infty p(j,x)(\ln(x)-\mu_j)^2\,dx}{\sum_{i\in\mathcal{I}_o} p_i(j) + m\int_0^\infty p(j,x)dx} \tag{6}$$

$$\eta_j = \frac{\sum_{i\in\mathcal{I}_o} p_i(j) + m\int_0^\infty p(j,x)dx}{n+m}. \tag{7}$$

Optimization on θ: We consider the gradient ascent method with respect to θ while keeping τ fixed. Taking the FOC of Equation (1) with respect to θ_i gives rise to:

$$\begin{aligned}\frac{\partial}{\partial\theta_i} ll(\tau,\theta) = &\sum_{i\in\mathcal{I}_o} \frac{\phi(l^0(d_i)|\gamma(\theta,d_i))}{1-\Phi(l^0(d_i)|\gamma(\theta,d_i))}\frac{l^0(d_i)}{\gamma(\theta,d_i)}\frac{\partial\gamma(\theta,d_i)}{\partial\theta_i} \\ &+ \frac{m\int_0^\infty \phi(l^0(x)|\gamma(\theta,x))\frac{l^0(x)}{\gamma(\theta,x)}\sum_{j=1}^{J}\phi(\ln(x)|\mu_j,\sigma_j)\eta_j\frac{\partial\gamma(\theta,x)}{\partial\theta_i}x^{-1}dx}{\int_0^\infty \Phi\left(l^0(x)|\gamma(\theta,x)\right)\sum_{j=1}^{J}\phi(\ln(x)|\mu_j,\sigma_j)\eta_j x^{-1}dx}.\end{aligned} \tag{8}$$

The full log-likelihood maximization algorithm is summarized below:

Algorithm 3.1: Estimate parameters of the probability distributions of customers' download and willingness to pay

Input : ν, κ, $\gamma(\theta,d):\mathbb{R}^u\times\mathbb{R}_+\to\mathbb{R}_+$, ϵ, ϵ_τ, ϵ_θ
Output: $\tau=\{(\mu_j,\sigma_j,\eta_j)\}_{j=1}^J$, θ

Initialize starting values of $\theta^{(0)}$ and $\tau^{(0)}$. $h=1$, $ll^{(0)}=-\infty$, $ll^{(h)}=ll(\tau^{(0)},\theta^{(0)})$;
while $ll^{(h)}-ll^{(h-1)}>\epsilon$ **do**
  /* Optimization on τ */
  $ll_\tau^{(0)}=-\infty$, $ll_\tau^{(1)}=ll^{(h)}$, $\theta=\theta^{(k-1)}$, $k=1$;
  while $ll_\tau^{(k)}-ll_\tau^{(k-1)}>\epsilon_\tau$ **do**
    /* E-Step */
    $k=k+1$;
    Compute $p_i(j)$ from Equation (3), for all $i\in\{1,\ldots,n\}$;
    Use $p(j,x)$ from Equation (4) to compute $\int_0^\infty p(j,x)dx$, $\int_0^\infty p(j,x)\ln(x)dx$, and $\int_0^\infty p(j,x)(\ln(x)-\mu_j)^2dx$, for all $j\in\{1,\ldots,J\}$;
    /* M-Step */
    Compute $\tau^{(k)}$ ($\mu_j^{(k)}$, $\sigma_j^{(k)}$, and $\eta_j^{(k)}$) from Equations (5), (6), and (7);
    $ll_\tau^{(k)}=ll(\tau^{(k)},\theta)$;
  /* Optimization on θ */
  $ll_\theta^{(0)}=-\infty$, $ll_\theta^{(1)}=ll_\tau^{(k)}$, $\tau=\tau^{(k)}$, $k=1$;
  while $ll_\theta^{(k)}-ll_\theta^{(k-1)}>\epsilon_\theta$ **do**
    $k=k+1$;
    $\theta^{(k)}=\theta^{(k-1)}+\beta\cdot\nabla_\theta ll(\tau,\theta^{(k-1)})$;
    $ll_\theta^{(k)}=ll(\tau,\theta^{(k)})$;
  $h=h+1$;
  $ll^{(h)}=ll_\theta^{(k)}$;

A discussion regarding the computation of the integrals in Algorithm 3.1 is provided in Appendix D. We now consider the convergence of Algorithm 3.1. First, we consider the optimization problem with respect to τ. Let $Q(\tau|\tau^{(k)})$ be defined as in Equation (2), where $p_i(j)$ and $p(j,x)$ are the expressions of Equations (3) and (4). The next proposition states the convergence result for the EM algorithm.

PROPOSITION 2. *All the limit points of any instance* $\{\tau^{(k)}\}$ *of the EM algorithm are stationary points of the log-likelihood function (as a function of* τ*), and* $ll(\tau^{(k)},\theta)$ *converges monotonically to* $ll^*(\theta) = ll(\tau^*,\theta)$.

Proof. The results follow from the continuity of $Q(\tau|\tau^{(k)})$ on τ and $\tau^{(k)}$, and using Theorem 2 of Wu (1983). □

Note that the EM algorithm might converge to a local maximum. Nonetheless, this complexity is beyond the EM algorithm itself and might still arise regardless of the optimization machinery in use. The next proposition states the convergence of Algorithm 3.1.

PROPOSITION 3. *All the limit points of any instance of Algorithm 3.1 converge to stationary points of the log-likelihood function in Equation* (1).

Proof. From Algorithm 3.1, the optimization on θ converges to a stationary point. This follows since $ll(\tau,\theta)$ is a continuously differentiable function of θ. Putting this together with Proposition 2 leads to the desired result. □

The last thing that needs to be specified is the mean WTP function $\gamma(\theta,d)$. We consider different functional forms of $\gamma(\theta,d)$ (see first column of Table 2), all of which are exponential functions of polynomial expressions of downloads. The reason for using the exponential function is twofold: (i) this results in a log-log relationship between customers' valuations and consumption, in particular log-normal WTP, which has been used in previous works (see Liu et al. (2000) and Cameron and James (1987)), and (ii) the shape of the current prices for different levels of downloads seems to follow a more log-log relation rather than linear-linear, linear-log, or log-linear (see Appendix E for details). We consider polynomial functions because of their ease to manipulate (for example compute their derivatives), as well as their flexibility to adopt increasing/decreasing and concave/convex shapes. For each functional form (γ) considered, we run Algorithm 3.1 to estimate the distribution parameters τ and θ on a training set and to evaluate the log-likelihood attained in the testing set. Training and testing sets are selected via cross-validation. This procedure is done for every combination of $\nu \in \{1/3, 2/3, 1, 4/3, 5/3\}$ and $\kappa \in \{1/4, 1/3, 1/2\}$. Table 2 shows the average out-of-sample log-likelihood for $\kappa = 1/3$. The analogous tables for $\kappa \in \{1/4, 1/2\}$ are shown in Appendix F. We observe from Table 2 that the parametric function that consistently achieves the best performance across all values of ν is $\gamma(\theta,d)=\theta_1+\theta_2\ln(d+\theta_4)+\theta_3\ln(d+\theta_4)^2$. Therefore, in order to generate the missing data, we select this functional form. We also try other distributions for the WTP. More specifically, we consider the Uniform distribution with the same standard deviation values considered for the Normal case, where the mean has the same functional forms as the ones shown in the first column of Table 2 (see details in Appendix G). The results in terms of log-likelihood are similar under both distributions, although there is slightly better performance under the Normal distribution. Thus, we decided to stay with the Normal distribution.

$\ln(\gamma(\theta,d))$	$\nu=1/3$	$\nu=2/3$	$\nu=1$	$\nu=4/3$	$\nu=5/3$
$\theta_1+\theta_2\ln(d)$	−39910.11	−42160.70	−43764.99	−45022.97	−46063.32
$\theta_1+\theta_2\ln(d+1)$	−39902.77	−42147.69	−43749.03	−45003.76	−46039.15
$\theta_1+\theta_2\ln(d+\theta_3)$	−39902.27	−42148.16	−43749.45	−45003.74	−46038.59
$\theta_1+\theta_2\ln(d)+\theta_3\ln(d)^2$	−39900.25	−42148.80	−43752.46	−45008.20	−46044.13
$\theta_1+\theta_2\ln(d+1)+\theta_3\ln(d+1)^2$	−39900.19	−42144.75	−43745.08	−44997.56	−46029.96
$\theta_1+\theta_2\ln(d+\theta_4)+\theta_3\ln(d+\theta_4)^2$	−39900.15	−42143.14	−43742.10	−44994.71	−46027.27
$\theta_1+\theta_2\ln(d+1)+\theta_3\ln(d+1)^{0.5}$	−39901.95	−42146.11	−43747.10	−45002.16	−46036.45
$\theta_1+\theta_2\ln(d+\theta_4)+\theta_3\ln(d+\theta_4)^{0.5}$	−39901.04	−42145.13	−43744.91	−44998.37	−46033.02

Table 2 Average log-likelihood among 5-fold cross-validation training sets for $\kappa=1/3$ for different parametric functions of the mean willingness to pay after running Algorithm 3.1.

3.5. Generation of Missing Data

Once we run Algorithm 3.1, we end up with estimates of the parameters of the distribution of customers' downloads and willingness to pay, τ and θ, respectively. In order to incorporate uncertainty in the estimated parameters, we bootstrapped over a total of $1,000$ scenarios obtaining a different set of estimated parameters for each of them. (For ease of notation, we do not include the index of the particular bootstrap scenario in the estimated parameters in the description of the sampling of data described below.). Let N be the total number of customers for which we want to sample data. N represents the size of the sampled instance; it can take any arbitrary positive integer value. Each customer is considered to be an observed one with probability $1/(1+\nu)$ and an unobserved one with probability $\nu/(1+\nu)$. More details are given at the beginning of Section 4. The generation of missing data, which is used as an input for the pricing optimization problem, is summarized in the following steps (see Appendix H for details):

For the observed customers

I Sample their downloads according to their empirical distribution obtained from the bootstrap method.

II Sample their willingness to pay from the distribution with PDF

$$f_W(w|W_i\geq l^0(d_i)) = \frac{\mathbb{1}_{\{w\geq l^0(d_i)\}}\phi(w|\gamma(\theta,d_i))}{1-\Phi(l^0(d_i)|\gamma(\theta,d_i))}.$$

For the unobserved customers

III Sample their downloads from the distribution with PDF

$$f_D(x|W<l^0(D)) = \frac{\Phi\left(l^0(x)|\gamma(\theta,x)\right)\sum_{j=1}^{J}\phi(\ln(x)|\mu_j,\sigma_j)\eta_j x^{-1}}{\int_0^\infty \Phi\left(l^0(y)|\gamma(\theta,y)\right)\sum_{j=1}^{J}\phi(\ln(y)|\mu_j,\sigma_j)\eta_j y^{-1}dy}.$$

IV Sample their willingness to pay from the distribution with PDF

$$f_W(w|W_i<l^0(d_i)) = \frac{\mathbb{1}_{\{w<l^0(d_i)\}}\phi(w|\gamma(\theta,d_i))}{\Phi(l^0(d_i)|\gamma(\theta,d_i))}.$$

Note that the download values of the unobserved customers are known information at this point, as these were sampled in Step III. In addition, it is worth pointing out that the size of the sampled instances in terms of the number of customers can be of arbitrary size N. If we want to generate an instance of real size, i.e., $N=n+m$, then we can modify Step I by simply inputting the observed download values instead of sampling from the empirical distribution. Also, note that on Step I it would also be correct to sample observed customers' downloads according to the conditioned random variable $D|W\geq l(D)$ (analogous to Step III).

4. Optimization Model

At this point, we have introduced a methodology that allows us to generate the missing information. As a result, we can sample instances of the problem and use SAA to find the optimal prices. Each instance consists of a set of N customers and a tuple of download and willingness to pay associated with each of them. The objective function, as a function of the prices $(\mathbf{f},\mathbf{v})\in\mathbb{R}^{2S}$, can be expressed as the sum of the expected revenues from observed and unobserved customers, namely:

$$\begin{aligned}&\sum_{i\in\mathcal{I}_o} l(d_i,\mathbf{f},\mathbf{v})\mathbb{E}\left[\mathbb{1}_{\{W_i\geq l(d_i,\mathbf{f},\mathbf{v})\}}|W_i\geq l^0(d_i)\right] + \sum_{i\in\mathcal{I}_u}\mathbb{E}\left[l(D_i,\mathbf{f},\mathbf{v})\mathbb{1}_{\{W_i\geq l(D_i,\mathbf{f},\mathbf{v})\}}|W_i<l^0(D_i)\right]\\ &= n(1+\nu)\left(\mathbb{E}\left[\frac{1}{1+\nu}l(D^e,\mathbf{f},\mathbf{v})\mathbb{1}_{\{W\geq l(D^e,\mathbf{f},\mathbf{v})\}}|W\geq l^0(D^e)\right]\right.\\ &\quad\left.+\mathbb{E}\left[\frac{\nu}{1+\nu}l(D,\mathbf{f},\mathbf{v})\mathbb{1}_{\{W\geq l(D,\mathbf{f},\mathbf{v})\}}|W<l^0(D)\right]\right), \end{aligned} \tag{9}$$

where D^e is a random variable that follows the empirical distribution of the observed customers' downloads. Let $\overline{D}$ be a random variable equal to D^e w.p. $1/(1+\nu)$ and $D|W<l^0(D)$ (a download of an unobserved customer) w.p. $\nu/(1+\nu)$ (see Step III of Section 3.5). Then, the optimization problem can be stated as the maximization of the expected revenue obtained from a representative consumer that solves

$$\max_{\mathbf{f},\mathbf{v}\in\Omega}\ \mathbb{E}\left[l(\overline{D},\mathbf{f},\mathbf{v})\mathbb{1}_{\{W\geq l(\overline{D},\mathbf{f},\mathbf{v})\}}\right], \tag{10}$$

where $\Omega:=\{(\mathbf{f},\mathbf{v})\in\mathbb{R}^{2S}|f_s\leq f_{s+1},v_s\geq v_{s+1}\ \forall s\in\{1,\ldots,S-1\},f_1\geq 0,v_S=0,v_s\geq(f_{s+1}-f_s)/(b_{s+1}-c_{s+1}-b_s)\ \forall s\in\{1,\ldots,S-2\}\}$ is the prices' feasible (polyhedral) set. Note that the objective function in Equation (10) represents the expected revenue obtained from a representative consumer. The idea is to use SAA to solve instances of the problem by first sampling the customer type (observed or unobserved), followed by the sampling of her download and her willingness to pay. More specifically, each customer is an observed (unobserved) one w.p. $1/(1+\nu)$ $(\nu/(1+\nu))$. Depending on the case, her download and willingness to pay are sampled according to either Steps I and II, or III and IV of Section 3.5. Let N be the number of customers to be considered in the instance to solve and let $\mathcal{I}_N \coloneqq \{1,\dots,N\}$; then, the optimization problem over the sample instance is

$$\max_{\mathbf{f},\mathbf{v}\in\Omega} \frac{1}{N}\sum_{i\in\mathcal{I}_N} l(\overline{D}_i,\mathbf{f},\mathbf{v})\mathbb{1}_{\{W_i\geq l(\overline{D}_i,\mathbf{f},\mathbf{v})\}}. \tag{11}$$

The expectation and sample average of Optimization Problems (10) and (11), respectively, can also consider arbitrary distributions over the problem parameters, such as ν and κ. More on this is presented in the numerical computations. In the next subsection, we present a dynamic programming formulation to solve the optimization problem.

4.1. Dynamic Programming (DP) for Price Optimization

The price optimization problem can be formulated as a dynamic program, DP, where each stage corresponds to the pricing decision of some plan. This DP is denoted as DPP. More specifically, consider the set of stages $\{1,\dots,S\}$ so that at stage t we decide the action $a^t = (a_1^t, a_2^t) = (f_t, v_{t-1})$ composed of the fixed price of plan t and the variable price of plan $t-1$ for $t>1$; if $t=1$, then we only decide on the former. In order to satisfy the monotonic conditions of the fixed and variable prices (the former are increasing in the plan index while the latter are decreasing), we need to keep track of the previous plan prices. Therefore, we consider the state $s^t = (s_1^t, s_2^t) = (f_{t-1}, v_{t-2})$ to be the fixed price of plan $t-1$ and the variable price of plan $t-2$ for $t>2$; if $t=2$, then the state is given by $s^2 = (f_1, \overline{v})$, where $\overline{v}$ is the maximum variable price (more details of this are given in the remainder of the paragraph). For $t=1$ the unique state is simply given by $s^1 = (0,\overline{v})$. Note that the state space is driven by the action space of previous stages, as the recursion can be simply stated as $s^{t+1}(a^t) = a^t$. For tractability purposes, we discretized the set of actions (and therefore the set of states) at each stage. In particular, given a state s_t at stage t, the action space of the stage is defined as $\mathcal{A}_t(s^t) = \{(f,v) | f\in[s_1^t,\overline{f}_t], v\in[0,s_2^t], (f-s_1^t)\leq v(b_t-c_t-b_{t-1}), \exists k,j\in\mathbb{Z}_+ \text{ s.t. } f=k\varepsilon_f, v=j\varepsilon_v\}$, where $\varepsilon_f>0$ and $\varepsilon_v>0$ denote the "grid" refinement for fixed and variable prices, respectively, and $\overline{f}_t$ is a cap on the fixed prices of the action space. The inequality on the action set definition simply enforces the last business rule discussed in Section 2. Let us denote $e_{is} \coloneqq \max\{d_i - b_s, 0\}$ as the download amount of customer i that exceeds the bundle data of plan s. The revenue function at each stage $t>1$ as a function of the state s^t and action a^t can be defined as

$$V_t(s^t,a^t) = \sum_{i\in\mathcal{N}_t} \min\{s_1^t + a_2^t e_{i,t-1}, a_1^t\}\mathbb{1}_{\{W_i\geq \min\{s_1^t+a_2^t e_{i,t-1},a_1^t\}\}} + V_{t+1}^*(a^t), \tag{12}$$

where $\mathcal{N}_t$ is the set of customers whose downloads are in the interval $[b_{t-1}, b_t)$. If $t=1$, the revenue function is $V_1(s^1, a^1) = a_2^1 \sum_{i\in\mathcal{N}_1} \mathbb{1}_{\{W_i \geq a_1^t\}} + V_2^*(a^t)$, where $\mathcal{N}_1 = [0, b_1]$. Finally, the revenue-to-go can be expressed as

$$V_t^*(s^t) = \max_{a^t\in\mathcal{A}(s^t)} V_t(s^t, a^t), \tag{13}$$

where the boundary condition is $V_{S+1}^* = 0$. Thus, the dynamic program is solved from stages $t=S$ to $t=1$, obtaining the optimal value as $z^{DPP} = V_1^*((0,\overline{v}))$.

It is worth noting that the DPP has a polynomial number of stages and actions, and therefore it needs a polynomial number of computations at each stage; indeed, the number of states (and actions) is $O(\varepsilon_f^{-1}\varepsilon_v^{-1})$. We considered other solution methods besides DP, such as an MIP formulation. However, the MIP run time for an instance is of the order of days (more than 5 days to get a 2% optimality gap). In addition, these methods need to be applied to multiple instances due the fact that we are using an SAA approach.

Due to considering a discrete set of prices instead of a continuous one (Ω), DPP will return an approximate solution, i.e., a lower bound of Problem (11). We provide an Upper Bound Dynamic Program (UBDP) that provides an upper bound z^{UBDP} of the optimal objective function of Optimization Problem (11) and therefore of the DPP as well.

Proposition 4. *An upper bound of Optimization Problem* (11) *and DPP is attained by the dynamic program in which states are the same as the DPP, yet the revenue function is changed to*

$$V_t'(s^t, a^t) = \sum_{i\in\mathcal{N}_t} \min\{s_1^t + a_2^t e_{i,t-1}, a_1^t\} \mathbb{1}_{\{W_i \geq \min\{(s_1^t-\varepsilon_f)^+ + (a_2^t-\varepsilon_v)^+ e_{i,t-1}, (a_1^t-\varepsilon_f)^+\}\}} + V_{t+1}'^*(a^t), \tag{14}$$

for $t>1$, *and* $V_1'(s^1, a^1) = a_2^1 \sum_{i\in\mathcal{N}_1} \mathbb{1}_{\{W_i \geq (a_1^t-\varepsilon_f)^+\}} + V_2^*(a^t)$. *The revenue-to-go is* $V_t'^*(s^t) = \max_{a^t\in\mathcal{A}(s^t)} V_t'(s^t, a^t)$

Proof. See Appendix I. $\square$

4.2. Price and Bundle Optimization

So far, we have only focused on the optimal prices while keeping fixed the plans' bundles and the number of plans. In this section, we provide a methodology for jointly optimizing fixed and variable prices, and the bundle that each plan should include. We build on the DPP formulation given in Section 4.1 by including the plans' bundles into the states and actions at each stage. The new dynamic programming formulation is referred as DPPB. Let us denote $\mathcal{B}$ as the feasible set of values for the plans' bundle data. We assume $|\mathcal{B}| < \infty$. The reason to consider a discrete set of bundle values is for the sake of keeping the state space finite. The stages are the same as in the DPP formulation, namely, $\{1,\ldots,S\}$. At each stage t, the state is the triplet consisting of the bundle of plan $t-1$, the fixed price of plan $t-1$, and the variable price of plan $t-2$, thus $s^t = (b_{t-1}, f_{t-1}, v_{t-2})$. Similarly, the action space at stage t consists of the triplet of the bundle of plan t, the fixed price of the plan t, and the variable price of plan $t-1$. Then, the action space at stage t can be written as $\mathcal{A}_t = \{(b,f,s)|b \in \mathcal{B}, b > s_1^1, f \in [s_1^t, \overline{f}_t], v \in [0, s_2^t], f = k\varepsilon_f, v = j\varepsilon_v, k, j \in \mathbb{Z}_+, (f - s_1^t) \le v(b_t - c_t - b_{t-1})\}$. The recursion function is $s^{t+1}(a^t) = a^t$. The revenue function is defined as

$$V_t(s^t, a^t) = \sum_{i|d_i \in (s_1^t, a_1^t]} \min\{s_1^t + a_2^t e_{i,t-1}, a_1^t\} \mathbb{1}_{\{W_i \ge \min\{s_1^t + a_2^t e_{i,t-1}, a_1^t\}\}} + V_{t+1}^*(a^t). \tag{15}$$

The revenue-to-go is written as follows:

$$V_t^*(s^t) = \max_{a^t \in \mathcal{A}(s^t)} V_t(s^t, a^t), \tag{16}$$

where the boundary condition is $V_{S+1}^* = 0$. Thus, the dynamic program is solved from stages $t = S$ to $t = 1$, obtaining the optimal value as $z^{DPPB} = V_1^*((0, 0, \overline{v}))$.

In addition to the price and price-bundle optimization, we are interested in analyzing the setting in which only bundles are optimized while keeping prices fixed to their current values. In other words, what is the impact in revenues due to a change in prices, bundles, and both? The formulation of this DP that considers only the plans' bundles as decision variables, can be simply obtained from the formulation of DPPB, and therefore it is omitted for the sake of brevity. The DP that optimizes just on bundles is denoted by DPB.

5. Numerical Results

We illustrate the performance of the solution methods presented in Section 4 by applying SAA over a set of $M = 1,000$ instances. Each instance is composed of a sample of $N = 1,000$ customers, where each of them has a download and willingness to pay. (We picked $N = 1,000$ since we found that it was a good balance between the number of customers and the solution time. Despite the fact that the solution time is not the main bottleneck for the type of industry our example applies to in this case, we still need to solve $1,000$ instances for the 15 combinations of parameters ν and κ, and for each solution method.) The described process is done for each tuple of parameters $(\nu, \kappa) \in \{1/3, 2/3, 1, 4/3, 5/3\} \times \{1/4, 1/3, 1/2\}$.

This section is divided into four parts: (i) we analyze the DPP optimality gap on the instances solved and on the expected revenue (Problem (10)); (ii) we analyze the obtained optimal prices from DPP for different parameters (κ, ν) of the problem; (iii) we study the performance of these prices with respect to the current prices used by the company, taking into consideration cases where the real parameters (as the number of unobserved customers and the coefficient of variation of customers' WTP) are wrongly assumed; and, finally, (iv) we compare the performance of optimizing on prices, on bundles, as well as jointly on both.

5.1. DPP Performance — Optimality Gap

In this subsection, we show the optimality gap of the DPP solution: (i) for each of the scenarios considered in the solved instances and (ii) with respect to the expected revenue of the optimization problem (10). These optimality gaps are denoted as **Opt-Gap-I** and **Opt-Gap-II**, respectively. **Opt-Gap-I** can be easily computed by using the UBDP and then computing the average gap among the solved instances. Namely, for each combination of (ν, κ), we average the optimality gap over the (M) instances run. For each instance, the upper bound used is computed with the UBDP method described in Section 4.1. In order to compute **Opt-Gap-II**, we state the following lemma:

LEMMA 1. *An upper bound of Problem* (10) *with* $(1-\alpha) \in (0,1)$ *confidence is given by*

$$z_{N,M}^{UB} := \overline{z}_{N,M}^{UBDP} + t_{\alpha,M-1}\hat{\sigma}_{N,M}^2, \tag{17}$$

where $\overline{z}_{N,M}^{UBDP} := \frac{1}{M}\sum_{m=1}^{M} z_N^{UBDP,m}$, $\hat{\sigma}_{N,M} := \frac{1}{M}\left(\frac{1}{M-1}\sum_{m=1}^{M}\left(z_N^{UBDP,m} - \overline{z}_{N,M}^{UBDP}\right)^2\right)$, $z_N^{UBDP,m}$ *is the upper bound of the* m^{th} *Sample Optimization Problem* (11) *when using the upper bound dynamic program,* N *denotes the number of customers in the SAA problem (i.e., Problem* (11)*),* M *represents the number of times this problem is solved for different samples, and* $t_{\alpha,M-1}$ *is the CDF at* $1-\alpha$ *of a t-distribution with* $M-1$ *degrees of freedom.*

Proof. See Appendix J. □

Let z^{DPP} be the objective function of Optimization Problem (10) evaluated at the average prices obtained from the SAA. More precisely, for a given tuple of (κ, ν), we average the M fixed and variable prices obtained from the SAA problems (i.e., instances of Optimization Problem (11)) and evaluate these average prices in the objective of Problem (10). From Lemma 1, we can state with a confidence level of $1-\alpha$ that an upper bound of the optimality gap is given by $\left(z_{N,M}^{UB} - z^{DPP}\right)/z^{DPP}$. Table 3 shows **Opt-Gap-I** and **Opt-Gap-II** for each considered scenario for (κ, ν). We can observe from the results of **Opt-Gap-I** that, on average, the DPP attains a solution within 3.8% from the optimal solution of the instances considered. It is worth noting that optimality gaps are to some degree inversely related to the spread of customers' willingness to pay (κ). The reason for this is that, when customer valuations are closer together (i.e., lower values of κ), it is more likely that more customer valuations will become cluttered in between the grid of prices used in the UBDP method, leading to a higher upper bound.

Opt-Gap-II from Table 3 is computed using a confidence level $(1-\alpha)$ of 95%. It can be seen that the DPP solution lies within 17.93% of the optimal solution. In general, the optimality gap increases when assuming higher values of unobserved customers and variability of customers' WTP. The latter is because scenarios with higher variability have more dispersion in their upper bound attained among the SAA instances, which increases the last term of Equation (17), thus leading

Scenario		Gap [%]		Time [s]	
κ	ν	**Opt-Gap-I**	**Opt-Gap-II**	DPP	UBDP
1/4	1/3	3.13	9.63	75	100
	2/3	3.39	5.77	60	79
	1	3.47	8.71	68	87
	4/3	3.64	8.64	62	79
	5/3	3.75	17.93	55	71
1/3	1/3	2.67	9.67	128	174
	2/3	3.16	10.35	98	130
	1	3.39	11.48	83	111
	4/3	3.52	11.65	79	105
	5/3	3.55	9.26	76	100
1/2	1/3	2.06	11.28	265	359
	2/3	2.77	11.09	144	196
	1	3.08	11.43	130	174
	4/3	3.31	13.26	120	162
	5/3	3.41	11.54	105	140

Table 3 Opt-Gap-I and Opt-Gap-II for each pair (κ,ν). Opt-Gap-I is computed as $(100/M)\sum_{h=1}^{M}\left(z_h^{UBDP}-z_h^{DPP}\right)/z_h^{UBDP}$ where z_h^{DPP} is the objective value attained under instance h, and z_h^{UBDP} is the respective upper bound attained with $UBDP$. Opt-Gap-II is computed as $100\left(z_{N,M}^{UB}-z^{DPP}\right)/z^{DPP}$.

to a higher optimality gap. In addition, we can observe from Table 3 that SAA prices obtained from instances with more uncertainty are likely to be more suboptimal than their counterpart in instances with less variability. It is worth to remark that the values shown for **Opt-Gap-II** in Table 3 are an upper bound of the optimality gap. In addition, even if we solve the instances with an exact method, there would still be an optimality bound of the obtained average solution with respect to the expected revenue of Optimization Problem (10).

5.2. Optimal Prices

We analyze the optimal prices obtained with the DPP for the different scenarios considered. It is important to analyze the obtained prices for each scenario in order to understand the impact of the different assumptions on the suggested prices, and hence on the obtained revenues. This analysis is key when looking at cases in which reality (in this case, customers' downloads and WTP) does not match the precise values of the assumed parameters. More on this is shown in the next section. Figure 4 shows the average fixed and variable prices obtained with DPP as a fraction of the current prices charged by the company. We observe the effect on the optimal prices of the assumed (i) number of unobserved customers and (ii) dispersion of customers' WTP. With respect to (i), as expected, prices tend to decrease when assuming more unobserved customers. The intuition is that more unobserved customers will result, on average, in a lower WTP. Consequently, assuming more of these customers will result in a decrease in prices in order to capture more revenue from

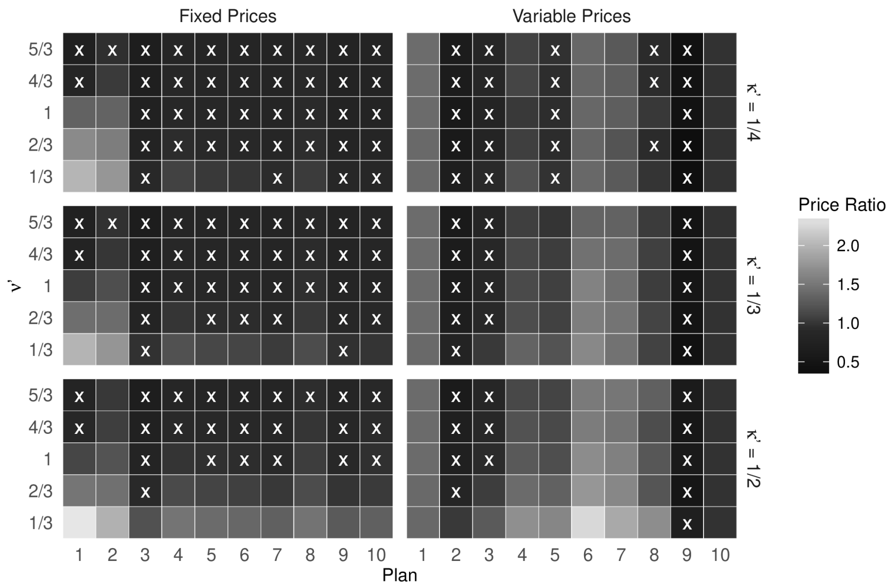

Figure 4 **Fixed (left panel) and Variable (right panel) Price Ratios between DPP and current prices computed as $\frac{f^{DPP}_{s,(\kappa',\nu')}}{f^0_s}$ and $\frac{v^{DPP}_{s,(\kappa',\nu')}}{v^0_s}$ respectively, where $f^{DPP}_{s,(\kappa',\nu')}$ and $v^{DPP}_{s,(\kappa',\nu')}$ are the average fixed and variable price for plan s among the M instances when assuming $(\kappa,\nu) = (\kappa',\nu')$. Values below 1 are marked with "$\times$". The last column is fixed to 1 since this variable price is null.**

this segment. In the case of (ii), assuming higher variations of customers' WTP will lead to a higher dispersion of prices when considering different levels of unobserved customers. To observe this, note that for every column (i.e., price), the price dispersion is much wider among the five rows of the bottom panels of Figure 4 compared to the dispersion among the five rows of the top panels. As a result, we can conjecture that revenues are likely to be more negatively impacted when misspecifying the number of unobserved customers when customers' WTP variation is high. Finally, prices are not that afar (i.e., most ratios are close to one) when assuming a high number of unobserved customers for different variations of customers' WTP; however, the contrary holds when the number of unobserved customers is low. We observe from the prices obtained that some of the current prices should be decreased. An example of the latter are the variable prices of plans 2 and 9. Figure 4 shows that their current value is higher than it should be in almost all scenarios. For plan 2, the decrease in the variable price is compensated by an increase in its fixed price (in almost all the scenarios). The current variable price of plan 9 is the same as that of plan 8, and due to the non-increasing rule of variable prices, there is room for decreasing the former. In general, the direction of price changes of the fixed prices with respect to the actual prices is more sensitive to the number of unobserved customers, as in the case of variable prices. The intuition behind this is that changes in the fixed prices have a greater effect on consumers than variable prices, since, as an example, an increase in a plan's fixed price will raise the tariff for all customers purchasing it. This is in contrast to an increase in the variable price, which would only impact some of the consumers. Also, we can see in Figure 4 that the price changes are more pronounced in general for the variable prices than for the fixed prices. One explanation is that variable prices have a much smaller magnitude than fixed prices; hence, the former price ratios are more pronounced than the latter ones.

5.3. DPP vs. Actual Prices

It is also of interest to analyze the performance of the obtained SAA prices with respect to the current prices the company charges. We proceed with the comparison by considering the cases in which some of the problem parameters have been incorrectly assumed. More specifically, let us rename the assumed model parameters as (κ', ν') (instead of (κ, ν)) to denote the parameters that are used to estimate customers' valuations and download distributions (by using Algorithm 3.1). Recall that the latter distributions are used for sampling instances of the problem to be solved with SAA with DPP. For each tuple (κ', ν'), we compute the average optimal prices among the M sampled instances. The real values of the model parameters will be denoted by (κ, ν). These last, (κ, ν), are used for evaluating the performance (by evaluating the objective function of Problem (11) ten million times in order to resemble the objective function of Problem (10)) of the average optimal prices obtained from the SAA. Note that the assumed parameters (κ', ν') might not necessarily match the real values of these (κ, ν).

Let $z^{DPP}_{(\kappa,\nu),(\kappa',\nu')}$ be the objective function of Optimization Problem (10) under the real parameters (κ, ν) evaluated at the average DPP prices assuming (a possibly different tuple of) parameters (κ', ν'). Also, denote $z^0_{(\kappa,\nu)}$ as the same objective function evaluated at the current prices $(\mathbf{f^0}, \mathbf{v^0})$ for the real parameters (κ, ν). The heatmap in Figure 5 shows the performance in terms of revenue of the DPP with respect to the current prices for different combinations of assumed and real parameters. In Figure 5, each row (column) represents an assumed scenario (the real scenario). We observe that DPP prices outperform current prices in most of the cases; however, there are some cases where misspecifying some of the problem parameters can decrease revenues. We separate the analysis of the cost of under- and overestimating (i) the number of unobserved customers and (ii) customers' WTP, while still observing the relationship between both.

(i) Misspecifying unobserved customers The three subplots of Figure 5 in the diagonal (from top-left to bottom-right) illustrate the instances where customers' WTP distribution is correctly assumed, but not necessarily the number of unobserved customers. We can see that overestimating and underestimating the number of unobserved customers could lead to a loss in revenue in some of the cases. As mentioned in Section 5.2, in the former case the firm charges lower prices than it should, whereas in the latter case the opposite happens. In addition, this revenue loss

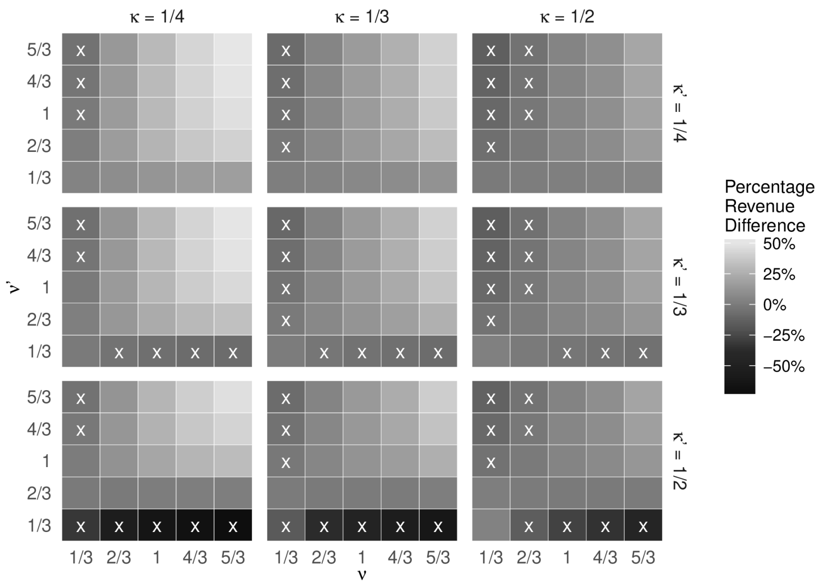

Figure 5 **Revenue difference is computed as** $\frac{z^{DPP}_{(\kappa,\nu),(\kappa',\nu')} - z^{0}_{(\kappa,\nu)}}{z^{0}_{(\kappa,\nu)}}$**. Assumed parameters** (κ',ν') **are in rows, while real parameters** (κ,ν) **are in columns. Cases where the revenue difference is negative are marked with "×".**

effect is more likely to happen in cases where customers' WTP coefficient of variation is high (see bottom-right panel of Figure 5). Another interesting observation that can be made from Figure 5 is how, for a given state of reality (i.e., a column of Figure 5), the revenue outcome (which can be either a gain or a loss) remains almost unchanged for different assumed values of customers' WTP variation when considering a high number of unobserved customers. On the contrary, if we assume a few number of unobserved customers, the revenue outcome will be highly dependent on the customers' WTP variation assumed. To visualize the latter, the top three rows for each κ' from Figure 5 illustrate how revenues are barely affected when assuming a high number of unobserved customers ($\nu' = 5/3$) for different WTP coefficients of variation assumed. However, assuming a low number of unobserved customers ($\nu' = 1/3$) leads to very different revenues depending on the WTP coefficient of variation assumed. An explanation of this can be obtained when looking at the prices in Figure 4, where there is little price difference between the top rows of the top and bottom panels, whereas the opposite holds for the bottom rows of the same two panels. Then, assuming a few number of unobserved customers leads to prices that are highly dependent on the level of variation of customers' WTP.

(ii) Misspecifying customers' WTP We observe from Figure 5 that assuming a higher variability for customers' WTP might induce higher risks in terms of revenues, which in most of the cases does not pay off. Indeed, when comparing each correlative *cell* (i.e., case) of the top three panels of Figure 5 with respect to their corresponding *cell* in the lower three panels, revenues can be severely reduced in the latter for some of the cases, while they are barely augmented for some other cases. For instance, if reality is such that $(\kappa,\nu)=(1/2,1)$, assuming $(\kappa',\nu')=(1/2,1)$ results in a slightly higher revenue than assuming $(\kappa',\nu')=(1/4,1)$; however, if reality is such that $(\kappa,\nu)=(1/2,1/3)$, revenues will be drastically decreased in the latter case (when assuming $(\kappa',\nu')=(1/2,1)$ instead of $(\kappa',\nu')=(1/4,1)$). The intuition behind this is that, as observed in Section 5.2, the magnitude of optimal prices has a negative relation with the number of unobserved customers. Moreover, price differences are more pronounced when assuming a higher coefficient of variation of customers' WTP. Then, pricing wrongly is more pronounced in the latter cases and therefore induces more revenue losses. As a result, assuming a lower WTP variation results in a more conservative and therefore robust strategy.

5.4. Price and Bundle Optimization

Similar to the price optimization, we run $M=1,000$ instances for scenarios with $\nu\in\{1/3,2/3,1,4/3,5/3\}$ and $\kappa\in\{1/4,1/3,1/2\}$ with DPB and DPPB. Let z_h^X denote the revenue obtained from solving instance $h\in\{1,\dots,M\}$ when using solving method $X\in\{DPP,DPB,DPPB\}$; or $X=0$ if this revenue is computed with actual prices. Table 4 shows, for

		Avg. Instance Rev. Improvement [%]			Expected Rev. Improvement [%]		
κ	ν	DPB	DPP	DPPB	DPB	DPP	DPPB
1/4	1/3	4.1	11.2	13.8	1.7	4.5	5.6
	2/3	10.3	24.8	27.7	7.6	17.0	18.3
	1	18.2	39.3	42.6	17.1	31.1	32.7
	4/3	25.1	49.7	53.7	25.9	42.9	44.8
	5/3	34.8	63.3	67.7	30.2	50.8	52.8
1/3	1/3	2.9	8.1	10.7	1.1	1.2	2.2
	2/3	4.8	14.9	17.8	2.5	7.3	8.6
	1	10.6	25.6	28.8	7.4	16.2	17.5
	4/3	16.3	36.0	39.5	13.7	25.9	27.1
	5/3	25.4	49.8	53.6	24.5	41.7	42.8
1/2	1/3	3.3	13.5	16.2	1.9	4.2	5.0
	2/3	2.9	8.9	11.7	0.4	0.6	1.4
	1	4.4	14.0	16.9	1.8	5.4	6.1
	4/3	8.0	20.8	23.9	3.9	10.0	10.6
	5/3	13.2	30.9	34.4	11.2	21.2	21.6

Table 4 Average Instance Revenue Improvement is computed as the average among all M instances of the percentage gap between the revenues attained by DPB, DPP, and DPPB versus using current prices. Namely, $(100/M)\sum_{h=1}^{M}\left(z_h^X-z_h^0\right)/z_h^0$ for $X\in\{DPP,DPB,DPBP\}$. Expected Revenue Improvement is computed as the expected revenue percentage difference using optimal prices and/or bundles of DPP, DPB, and DPPB versus the expected revenue with current prices.

each pair (κ, ν), the percentage revenue difference between the solving methods versus using current prices: (I) averaged over the M instances and (II) applying the SAA prices/bundles solution in the expectation of the optimization problem. The three following observations can be made. (i) Optimizing on prices is clearly more beneficial than doing so on bundles. Intuitively, prices have more flexibility than bundles to capture more revenue from customers, due to the fixed and variable tariffs of each plan. For example, when having customers with lower WTPs, optimizing on prices allows for decreasing the prices of the plans making them affordable to new customers, and thereby obtaining the corresponding additional revenue. In contrast, optimizing on bundles has more difficulties in terms of capturing these customers. To achieve the latter, one option could be to increase the plans' bundle values so that cheaper plans (i.e., those with low fixed prices) become attractive to these customers with low valuations. Nevertheless, since prices are fixed to the current values, there is still revenue left on the table, as is shown in Table 4. Also, note that the price optimization problem has twice the decision variables compared to those in the bundle optimization problem. (ii) In terms of the revenue attained on the solved instances, we observe that jointly optimizing over prices and bundles outperforms methods that optimize on prices or bundles alone. However, in terms of expected value, the objective function DPPB is just as good, although just slightly better, than DPP. Then, optimizing just on prices results in a significant increase in revenues, leaving little margin of improvement when bundles are also considered (along with prices) as decision variables. (iii) Scenarios that assumed more unobserved customers have a wider revenue improvement, as is shown in Table 4. This is because in these cases there are several additional potential customer purchases at the optimized prices/bundles, which translate into more revenues in the end. It is worth mentioning that the results obtained, especially (i), should be read with some caveat. Although we found in this application that optimizing only on prices outperforms optimizing on bundles, this might not be the case in other applications. Indeed, if the bundles of current plans are close to the optimal ones (when optimizing on prices and bundles), while prices of the current plans are far away from the optimal ones, optimizing on prices might be more promising than optimizing on bundles. The opposite is true if current prices are close to the optimal ones, and bundles are far away from the optimal ones.

6. Price Dependent Consumption

We extend the described model so that the consumers' download amounts depend on the price. Each consumer i has a reference download value, $r_i > 0$, which is not observed. This represents their ideal consumption level at null price. Then, given a vector of fixed and variable prices, consumers choose their downloads by solving a utility maximization problem. Since customers might not foresee the precise download amount, for each customer i we include a multiplicative (non-negative) noise, denoted by Y_i, which follows a log-Normal$(-\frac{a^2}{2}, a)$ so that its mean is equal to 1, where a is a positive parameter to estimate. Putting things together, the utility maximization problem that consumers solve is

$$\max_{d>0, s\in\mathcal{S}} w_i - \mathbb{E}\left[\frac{1}{2c}\left(\ln(dY_i) - \ln(r_i)\right)^2\right] - \frac{a^2}{2c} - f_s - v_s\mathbb{E}\left[\max\{dY_i - b_s, 0\}\right], \tag{18}$$

where the parameter $c > 0$ is inversely related with customer sensitivity to deviations from the reference download consumption. The first, second, third, and fourth terms in Equation (18) represent the consumer willingness to pay, the expected disutility from deviations from the reference download, a constant term for ease of manipulating the expression, and the expected cost of the chosen plan, respectively. Note that, if $(c, a) = (0, 0)$ and $r_i = d_i$ (i.e., the reference downloads of observed customers are equal to their observed download amounts) we obtain the same setting as before, in which consumer consumption is fixed. The decision variables are the download amount d, and the plan. The reference downloads, similarly to before, follow a mixture of log-Normal distributions with parameters $\tau = \{(\mu_j, \sigma_j, \eta_j)\}_{j=1}^J$, and the willingness to pay follows a distribution that depends on the reference download and the parameters θ. We use the same function form $\gamma(\theta, d)$ that achieved the best performance in the price-independent consumption setting. Then, the parameters to estimate are: τ, θ, a, and c. The utility maximization problem is described in more detail in Appendix L, as is the log-likelihood of the problem, as well as the estimation of the problem parameters by maximizing separately on τ by using the EM algorithm, and on θ, a, and c by using a gradient ascent method. Appendix M shows step by step the sampling of data to create the instances solved using SAA. We use a Nearest-Neighbor (NN) heuristic to solve the pricing problem. Unlike the previous case (price-independent consumption), the optimization problem is more difficult and cannot be addressed with the dynamic programming formulation used before. In NN, we use a grid of fixed and variable prices. For each current price solution, we evaluate the expected revenue in the *neighboring prices* moving to the best adjacent solution if there is an improvement, and stopping if there is not. A neighboring price is such that it is equal in all fixed and variable prices to the current price, except on one fixed or variable price whose value is adjacent in the price grid to the corresponding fixed or variable price of the current price we are at. More details of the NN heuristic used are provided in Appendix O.

Figure 6 shows the revenue differences when wrongly assuming the number of unobserved customers and/or the coefficient of variation of customers' WTP. Unlike the setting with price insensitive consumption, we observe now that the revenues obtained with the NN prices outperform the ones obtained with current prices in most of the cases, even when the assumed parameters are not the real ones. However, we note that this increase in revenue is of less magnitude than the

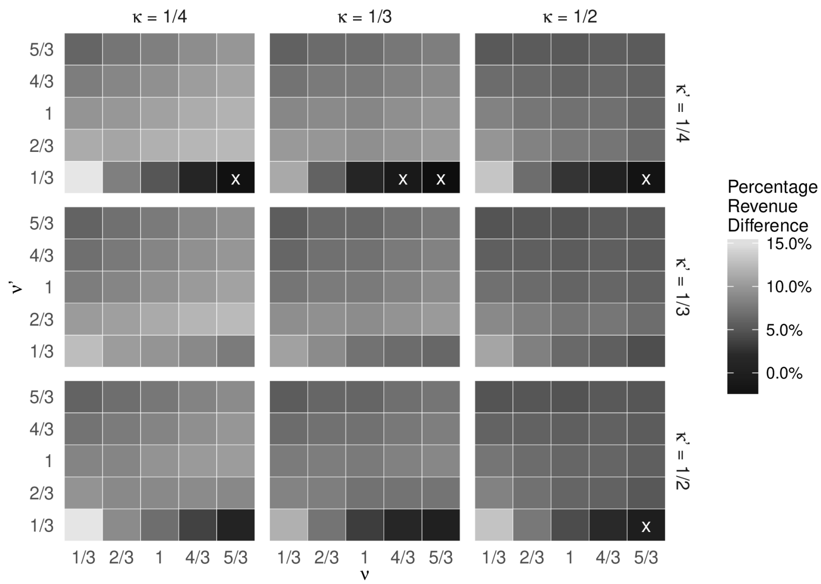

Figure 6 **Revenue difference is computed as $\frac{z^{NN}_{(\kappa,\nu),(\kappa',\nu')} - z^{0}_{(\kappa,\nu)}}{z^{0}_{(\kappa,\nu)}}$. Assumed parameters (κ', ν') are in rows, while real parameters (κ, ν) are in columns. Cases where the revenue difference is negative are marked with "$\times$".**

one shown in Figure 5. The reason for this is that in the price-dependent model, consumers might purchase other plans, contrary to the price insensitive situation where consumers have a fixed consumption amount and are, therefore, less likely to shift their consumption to other plans. As a result, the price-dependent consumption model leads to more cases of revenue increase compared to the insensitive consumption case.

7. Conclusions

We have presented a framework that allows us to study a variation of the two-price tariff problem from a data-driven perspective. We focused on two main aspects: (i) the estimation of unobserved data from parametric probability distributions, and (ii) the formulation and solution of the price, bundle, and price-bundle optimization problems. In addition, we show how the model described can also be extended to a price-dependent consumption setting. We applied our approach to a satellite service provider's problem who uses this pricing structure. Furthermore, we compute the potential revenues by using our approach compared to the current prices used by the company. Throughout the paper, the problem is described for the particular case of this satellite service provider; however, the methodology can be directly applied to settings that share a similar pricing structure.

There are several lines of research that could be followed up in future related work. For example, the estimation methodology could be extended to other settings with multivariate variables either for consumption (equivalent to download in our case) and/or WTP, or other customer' covariates of the particular context. Similarly, future works could incorporate information, such as capturing proxy information of unobserved customers in, for example, an online context where no purchases can be tracked.

Acknowledgments

We would like to thank the satellite service provider for their generous support that made this work possible. This work was also partly supported by NSF grants CMMI-1162034 and CMMI-1563343. This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02). The authors gratefully acknowledge financial support from CONICYT PIA/BASAL AFB180003.

References

Afeche, Philipp, Opher Baron, Joseph Milner, Ricky Roet-Green. 2019. Pricing and prioritizing time-sensitive customers with heterogeneous demand rates. *Operations Research* .

Alberini, Anna. 1995. Optimal designs for discrete choice contingent valuation surveys: Single-bound, double-bound, and bivariate models. *Journal of Environmental Economics and management* **28**(3) 287–306.

Bernstein, Fernando, Leslie M Marx. 2005. Reservation profit levels and the division of supply chain profit.

Cachon, Gérard P, Martin A Lariviere. 2005. Supply chain coordination with revenue-sharing contracts: strengths and limitations. *Management Science* **51**(1) 30–44.

Cameron, Trudy A, Michelle D James. 1987. Estimating willingness to pay from survey data: an alternative pre-test-market evaluation procedure. *Journal of Marketing Research* **24**(4) 389–395.

Chen, Boxiao, Xiuli Chao, Cong Shi. 2015. Nonparametric algorithms for joint pricing and inventory control with lost-sales and censored demand. *Mathematics of Operations Research, Major Revision* .

Dag, Ali, Kazim Topuz, Asil Oztekin, Serkan Bulur, Fadel M Megahed. 2016. A probabilistic data-driven framework for scoring the preoperative recipient-donor heart transplant survival. *Decision Support Systems* **86** 1–12.

Dempster, Arthur P, Nan M Laird, Donald B Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)* 1–38.

Dobson, Gregory, Shlomo Kalish. 1988. Positioning and pricing a product line. *Marketing Science* **7**(2) 107–125.

Essegaier, Skander, Sunil Gupta, Z John Zhang. 2002. Pricing access services. *Marketing Science* **21**(2) 139–159.

Feldstein, Martin S. 1972. Equity and efficiency in public sector pricing: the optimal two-part tariff. *The Quarterly Journal of Economics* 176–187.

Fibich, Gadi, Roy Klein, Oded Koenigsberg, Eitan Muller. 2017. Optimal three-part tariff plans. *Operations Research* **65**(5) 1177–1189.

Gallego, Guillermo, Özalp Özer. 2002. Optimal use of demand information in supply chain management. *Supply Chain Structures*. Springer, 119–160.

Hu, Kejia, Jason Acimovic, Francisco Erize, Douglas J Thomas, Jan A Van Mieghem. 2018. Forecasting new product life cycle curves: Practical approach and empirical analysis: Finalist–2017 m&som practice-based research competition. *Manufacturing & Service Operations Management* **21**(1) 66–85.

Huh, Woonghee Tim, Retsef Levi, Paat Rusmevichientong, James B Orlin. 2011. Adaptive data-driven inventory control with censored demand based on kaplan-meier estimator. *Operations Research* **59**(4) 929–941.

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

Iyengar, Raghuram, Sunil Gupta. 2009. 16 nonlinear pricing. *Handbook of Pricing Research in Marketing* **355**.

Jedidi, Kamel, Z John Zhang. 2002. Augmenting conjoint analysis to estimate consumer reservation price. *Management Science* **48**(10) 1350–1368.

Kaplan, Edward L, Paul Meier. 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**(282) 457–481.

Kohli, Rajeev, Vijay Mahajan. 1991. A reservation-price model for optimal pricing of multiattribute products in conjoint analysis. *Journal of Marketing Research* 347–354.

Kök, A Gürhan, Marshall L Fisher. 2007. Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research* **55**(6) 1001–1021.

Laird, Nan. 1978. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* **73**(364) 805–811.

Lambrecht, Anja, Katja Seim, Bernd Skiera. 2007. Does uncertainty matter? consumer behavior under three-part tariffs. *Marketing Science* **26**(5) 698–710.

Lewis, W Arthur. 1941. The two-part tariff. *Economica* **8**(31) 249–270.

Littlechild, Stephen C. 1975. Common costs, fixed charges, clubs and games. *The Review of Economic Studies* **42**(1) 117–124.

Liu, Jin-Tan, James K Hammitt, Jung-Der Wang, Jin-Long Liu. 2000. Mother's willingness to pay for her own and her child's health: a contingent valuation study in taiwan. *Health Economics* **9**(4) 319–326.

Miller, Klaus M, Reto Hofstetter, Harley Krohmer, Z John Zhang. 2011. How should consumers willingness to pay be measured? an empirical comparison of state-of-the-art approaches. *Journal of Marketing Research* **48**(1) 172–184.

Mitchell, Bridger M. 1978. Optimal pricing of local telephone service. *The American Economic Review* **68**(4) 517–537.

Musalem, Andrés, Marcelo Olivares, Eric T Bradlow, Christian Terwiesch, Daniel Corsten. 2010. Structural estimation of the effect of out-of-stocks. *Management Science* **56**(7) 1180–1197.

Ng, Yew-Kwang, Mendel Weisser. 1974. Optimal pricing with a budget constraint–the case of the two-part tariff. *The Review of Economic Studies* **41**(3) 337–345.

Oi, Walter Y. 1971. A disneyland dilemma: Two-part tariffs for a mickey mouse monopoly. *The Quarterly Journal of Economics* 77–96.

Oliveira, Fernando S, Carlos Ruiz, Antonio J Conejo. 2013. Contract design and supply chain coordination in the electricity industry. *European Journal of Operational Research* **227**(3) 527–537.

Rahmani, Morvarid, Guillaume Roels, Uday S Karmarkar. 2017. Collaborative work dynamics in projects with co-production. *Production and Operations Management* **26**(4) 686–703.

Rusmevichientong, Paat, Benjamin Van Roy, Peter W Glynn. 2006. A nonparametric approach to multi-product pricing. *Operations Research* **54**(1) 82–98.

Savaskan, R Canan, Shantanu Bhattacharya, Luk N Van Wassenhove. 2004. Closed-loop supply chain models with product remanufacturing. *Management Science* **50**(2) 239–252.

Schlereth, Christian, Tanja Stepanchuk, Bernd Skiera. 2010. Optimization and analysis of the profitability of tariff structures with two-part tariffs. *European Journal of Operational Research* **206**(3) 691–701.

Schmalensee, Richard. 1981. Monopolistic two-part pricing arrangements. *The Bell Journal of Economics* 445–466.

Shioda, Romy, Levent Tunçel, TGJ Myklebust. 2011. Maximum utility product pricing models and algorithms based on reservation price. *Computational Optimization and Applications* **48**(2) 157–198.

Tsay, Andy A, Steven Nahmias, Narendra Agrawal. 1999. Modeling supply chain contracts: A review. *Quantitative models for supply chain management*. Springer, 299–336.

Turnbull, Bruce W. 1974. Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association* **69**(345) 169–173.

Van Ryzin, Garrett, Gustavo Vulcano. 2015. A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Science* **61**(2) 281–300.

Varian, Hal R. 1985. Price discrimination and social welfare. *The American Economic Review* **75**(4) 870–875.

Vulcano, Gustavo, Garrett Van Ryzin, Richard Ratliff. 2012. Estimating primary demand for substitutable products from sales transaction data. *Operations Research* **60**(2) 313–334.

Wu, CF Jeff. 1983. On the convergence properties of the EM algorithm. *The Annals of Statistics* 95–103.

Yılmaz, Övünç, Pelin Pekgün, Mark Ferguson. 2017. Would you like to upgrade to a premium room? evaluating the benefit of offering standby upgrades. *Manufacturing & Service Operations Management* **19**(1) 1–18.

Zhu, Zhisu, Jiawei Zhang, Yinyu Ye. 2013. Newsvendor optimization with limited distribution information. *Optimization Methods and Software* **28**(3) 640–667.