Large-Scale Bundle-Size Pricing: A Theoretical Analysis

Tarek Abdallah
Kellogg School of Management, Northwestern University, tarek.abdallah@kellogg.northwestern.edu

Arash Asadpour
Lyft Inc., New York, aasadpour@lyft.com

Josh Reed
Leonard N. Stern School of Business, New York University, jreed@stern.nyu.edu

Bundle-size pricing (BSP) is a multidimensional selling mechanism where the firm prices the size of the bundle rather than the different possible combinations of bundles. In BSP, the firm offers the customer a menu of different sizes and prices. The customer then chooses the size that maximizes his surplus and customizes his bundle given his chosen size. While BSP is commonly used across several industries, little is known about the optimal BSP policy in terms of sizes and prices, along with the theoretical properties of its profit. In this paper, we provide a simple and tractable theoretical framework to analyze the large-scale BSP problem where a multiproduct firm is selling a large number of products. The BSP problem is in general hard as it involves optimizing over order statistics; however, we show that for large numbers of products, the BSP problem transforms from a hard multidimensional problem to a simple multiunit pricing problem with concave and increasing utilities. Our framework allows us to identify the main source of inefficiency of BSP: the heterogeneity of marginal costs across products. For this reason, we propose two new BSP policies, “clustered BSP” and “assorted BSP,” which significantly reduce the inefficiency of regular BSP. We then utilize our framework to study richer models of BSP, such as when customers have budgets.

1. Introduction

Many multiproduct firms engage in selling bundles with the intent of extracting a large consumer surplus. For example, the classical business model for cable TV involves selling large bundles of multiple TV channels. Airlines also usually bundle a seat reservation with multiple services, such as free bag check-in, access to the onboard entertainment system, and onboard meals, among others. However, in recent years, there has been a strong trend towards the unbundling of cable TV and airline tickets (Popper 2015, Owram 2014). This unbundling trend is usually justified by customers’ desire for the flexibility to pay for what they want rather than being tied to predesigned bundles (Flint 2015, Bhaskara 2015). Meanwhile, proponents of classical bundling argue that such practices simplify the selling process (Karp 2013, Vasel 2016).
The academic literature on consumer behavior emphasizes the importance of striking a balance between two key properties for a successful selling strategy: (i) simplicity (Freeman et al. 2012) and (ii) flexibility for customers to customize their purchases (Arora et al. 2008). In practice, several multiproduct firms may adopt different selling strategies even within the same industry. One common strategy is to price each product separately and let customers choose the bundle of products that they want. This form of selling is commonly referred to as component pricing. Component pricing is a relatively simple selling strategy that provides flexibility for customers to pay for what they want, but it suffers from inherent inefficiencies with regards to extracting a large consumer surplus (Adams and Yellen 1976). A somewhat opposite strategy is to restrict the sale of all the products to a single comprehensive bundle without allowing for individual product sales. This extreme form of bundling is commonly referred to as pure bundling. Pure bundling is a simple selling strategy, yet it can be inefficient when products have positive marginal costs (Abdallah 2019). In addition, pure bundling does not offer any flexibility for customers to customize their bundle.

However, bundling need not be synonymous with lack of customization. In fact, a multiproduct firm can offer full customization to its customers by selling every possible combination of its products as different bundles. This form of bundling is commonly referred to as mixed bundling. Mixed bundling is the most efficient bundle-selling strategy as it subsumes both pure bundling and component pricing as special cases. It also gives customers full flexibility to pay for what they want. However, mixed bundling is very complicated for both the firm and the customers as it involves selling an exponential number of bundles. More specifically, a firm with $N$ different products may potentially sell $2^N - 1$ different bundles. Therefore, the question remains as to whether there exists a simple selling strategy that mitigates the inefficiencies akin to pure bundling and component pricing yet provides customers with some flexibility to pay for what they want.

One potential candidate is the bundle-size pricing (BSP) strategy. In the BSP selling strategy, a multiproduct firm sets prices as a function of the quantity of products included in the bundle (i.e., the bundle size) regardless of which products are chosen. Hence, for a given bundle size, the customer is free to customize her bundle by choosing any number of products less than that size (e.g., buy any 2 shirts for $80). BSP is a simple selling strategy as it involves setting prices for a relatively small number of bundles compared to mixed bundling. More specifically, in the presence of $N$ different products, a firm adopting a BSP policy sets prices for $N$ different bundle sizes ranging from 1 through $N$. This is in contrast with the $2^N - 1$ prices required by mixed bundling. Furthermore, by its virtue, BSP allows a great deal of flexibility for customers to customize their bundle subject to the offered sizes.
BSP is used in several retailing industries and restaurants and is also extensively used for selling digital products such as customized cable TV, internet subscriptions, and data packages for phone plans. In fact, several companies have been able to successfully adopt a BSP strategy as a new form of subscription service where the service is priced based on the amount of access. Some examples of such companies are shown in Figure 1. A notable example is, Scribd, which amid criticism of the failing unlimited-subscription model for e-books, announced in 2015 a major change to its business model from an “all-you-can-read” model (pure bundling) to a BSP model that limits readers to only three books per month (Berkowitz 2016). Scribd justified this move by citing the increasing cost of offering e-books, stating on their website\(^1\) the following:

“As with any library (or business) there are certain costs involved in maintaining it. Books need to be purchased (or in our case - licensed).... For the first time ever, demand for ebooks is right on par with physical books, and to that end - ebooks are just as expensive (if not more so!).”

Despite its common use in practice, there is limited literature on the theoretical properties of BSP. In this paper, we present a theoretical framework to analyze and characterize the optimal BSP policy for a multiproduct firm that sells a large number of products. In this regard, our contribution is twofold. First, from a methodological standpoint, we provide a simple and tractable theoretical framework for analyzing the BSP problem as the number of products grows large. Our framework is based on showing that the BSP problem, while being a hard multidimensional problem, reduces in the limit to a simple multiunit monopolistic pricing problem with a nondecreasing concave utility that is easy to solve.

Our second contribution is utilizing our developed framework to shed some light on the theoretical properties of the optimal BSP policy under richer models and extensions. In the case of equal
marginal costs, and as the number of products grows large, we show that BSP can asymptotically achieve the profit under mixed bundling. In fact, both BSP and mixed bundling asymptotically achieve perfect price discrimination (i.e., first best). Meanwhile, we uncover the main source of BSP inefficiency: the heterogeneity of the marginal costs across products. More specifically, we show that the efficiency of BSP decreases as the heterogeneity in marginal costs increases, and hence, BSP can no longer achieve perfect price discrimination (which can be achieved by mixed bundling). However, using our framework, we propose two new BSP policies, “clustered BSP” and “assorted BSP,” which can reduce the inefficiency of the (regular) BSP policy when the products have heterogeneous marginal costs.

It is worth mentioning that regardless of whether marginal costs are homogeneous or heterogeneous, our framework reveals a nice property about the simplicity of the optimal BSP policy for large numbers of items, where it is sufficient to offer only a limited number of bundle sizes. Yet, this comes at the cost of offering customers less flexibility. More specifically, when customers have unlimited budgets and draw their valuations from a common underlying distribution, the asymptotically optimal BSP policy involves selling only one size for all the customers. Nonetheless when customers have heterogeneous budgets, one size is no longer asymptotically optimal, yet we can still characterize the asymptotically optimal BSP strategies with multiple sizes (even when the firm has no information on these budgets).

2. Related Literature
The bundling problem is related to the general multidimensional mechanism design problem (see Pekeč and Rothkopf (2003) or Cramton et al. (2006)). The literature on bundling dates back to a short note by Stigler (1963). The first serious attempt to analyze the mixed bundling problem was by Adams and Yellen (1976), who used a graphical approach with two items in order to illustrate how a firm can extract a large consumer surplus using bundles. Since then there has been a large literature in economics and marketing on bundling. However, due to the combinatorial complexity of mixed bundling, the classical literature has mainly focused on the analysis of two-item settings (see Venkatesh and Mahajan (2009) for a summary of the main insights in this literature). The bundling literature is also related to that on add-on pricing; see for example, Ellison (2005), Wang et al. (2015), and Cui et al. (2017).

For a long time, it was a challenge to extend the analysis beyond a two-item setting, until the seminal paper by Bakos and Brynjolfsson (1999). Motivated by the rise of digital markets and companies that sell a large number of digital products, Bakos and Brynjolfsson (1999) analyzed the bundling problem for a large number of items that have zero marginal costs, or “information goods.” They showed that, in the case of independent and additive item valuations, a seller of a large
number of information goods can come close to extracting all of the consumer surplus by simply selling all of the items as a pure bundle. Later, Geng et al. (2005) questioned the additive-valuation assumption and argued that bundles usually exhibit a subadditive valuation, where customers have a decreasing marginal valuation for each additional item. They further showed that under a particular discounted utility model, the results by Bakos and Brynjolfsson (1999) do not hold.

In the case of items with positive marginal costs, Bakos and Brynjolfsson (1999) argued using a counterexample that the efficiency of a pure bundle in extracting a large consumer surplus is highly compromised. Abdallah (2019) revisited the large-scale pure bundling problem but with nonnegative marginal costs. He characterized the asymptotically optimal pricing policy and provided a simple way to compute a tight lower bound on the potential loss from pure bundling. For the special case of information goods, Abdallah (2019) showed that the results by Bakos and Brynjolfsson (1999) hold for almost any correlation structure among the items’ valuations. He further extended his analysis to subadditive and superadditive valuations under a general utility model and characterized regimes under which the critique by Geng et al. (2005) holds. Chen et al. (2017) studied a more general distribution-free pricing problem, where a monopolist has limited distributional information on the customers’ valuation. They applied their robust pricing policy to the pure bundling problem and obtained tight bounds for a finite number of products.

In parallel to the bundling literature, there is an extensive literature on nonlinear pricing, which focuses on the role of multipart tariffs in extracting a large consumer surplus (see, for example, Wilson (1993)). The most widely studied tariff in this literature is the two-part tariff, which includes a lump-sum fee (subscription fee) and a per-product price (Oi 1971). For a large number of products, Armstrong (1999) showed that in the presence of nonnegative marginal costs, a cost-based two-part tariff where products are priced at their respective marginal costs comes close to extracting all of the consumer surplus as the number of items grows large. In the case of zero marginal costs, the two-part tariff involves only the tariff and hence is equivalent to pure bundling.

It is worth mentioning that any two-part tariff policy can be implemented by mixed bundling where each of the bundles is priced at the subscription fee plus the sum of the per-product prices. Therefore, in the presence of positive marginal costs, mixed bundling can also achieve the expected profit under perfect price discrimination as the number of items grows large. However, despite the appealing theoretical properties of the two-part tariff, it has been argued empirically and through experiments that customers tend to derive lower utility from the two-part tariff selling mechanism relative to simpler selling mechanisms such as pure bundling or flat-fee pricing mechanisms (Train et al. 1989, Iyengar et al. 2011, Lambrecht and Skiera 2006).

Several researchers have analyzed different forms of simple bundle selling mechanisms. For example, Ma and Simchi-Levi (2015) propose a new form of bundling called bundling with disposal,
where the customers purchase the full bundle but are allowed to return any item they want for a rebate that is equal to the marginal cost. Despite BSP being a simple bundling mechanism, the related literature is quite limited and dates back to Spence (1980), who demonstrated how a firm can extract a large consumer surplus by simply pricing based on the quantity purchased without discriminating among the items purchased. Hitt and Chen (2005) studied this type of quantity-dependent pricing in the context of bundling, which they call “customized bundling.” Their analysis is mainly based on deterministic valuations, where they use comparative statics to analyze “customized bundling” versus other selling mechanisms. Later, using extensive numerical simulations, Chu et al. (2011) demonstrated that a BSP mechanism, despite its simpler form, can very well approximate the optimal profit under mixed bundling. Xue et al. (2015) studied a related problem of pricing personalized bundles, where customers are quoted a price after they select their items.

Based on the above literature review, it can be concluded that very little is known about the theoretical properties of the optimal BSP policy. For this reason, we next present a simple and tractable theoretical framework to analyze the optimal BSP policy as the number of items grows large, as in Bakos and Brynjolfsson (1999), Armstrong (1999), and Abdallah (2019).

3. Model

Our setup for the bundling model is as follows. A firm is interested in selling \( N \geq 1 \) distinct products to \( M \geq 1 \) customers. Since we assume that there are no inventory constraints for the firm and that customers draw their valuations independently from the same multidimensional distribution, it suffices to set \( M = 1 \). The customer’s valuations for products 1 through \( N \) are given by the random vector \( X = (X_1, \ldots, X_N) \). For any given subset \( S \subseteq \{1, \ldots, N\} \) of the products, the customer’s valuation for that subset of products is additive in the sense that it is given by \( \sum_{n \in S} X_n \).

In the most general form of the bundling selling mechanism, the firm sets prices for all \( 2^N \) subsets of the products and the customer purchases the subset or bundle that maximizes her surplus (with the no-purchase also being an option). In the present paper we will, however, be interested in a simpler mechanism, referred to as bundle-size pricing (BSP). Under BSP, the firm sets prices that only depend on the number of items in a bundle and not on the products themselves. The customer is free to place any products she desires in a bundle of a particular size.

Let \( (X_{(1)}, \ldots, X_{(n)}) \) denote the order statistic corresponding to the random vector \( X \) of the customer’s product valuations, where \( X_{(1)} \leq \ldots \leq X_{(n)} \). Then, by the additive valuation assumption, the customer’s valuation for a bundle of size \( 1 \leq n \leq N \) is given by

\[
V_N(n) = \sum_{k=0}^{n-1} X_{(N-k)}.
\]
Given a menu $p = (p(1), \ldots, p(N))$ of bundle-size prices determined by the firm, the customer will purchase the bundle size that maximizes her surplus. This is given by

$$\zeta(X, p) \in \arg\max_{n=0, \ldots, N} (V_N(n) - p(n)),$$

where we set $V_N(0) = p_0 = 0$, so that $n = 0$ corresponds to the no-purchase option.

The firm incurs a nonnegative marginal cost $c_n \geq 0$ for each item $n = 1, \ldots, N$ sold. The cost for the firm to offer a bundle of a particular size depends on the items the customer chooses to place in it. Let $\tau : (1, 2, \ldots, N) \mapsto (1, 2, \ldots, N)$ be a permutation function such that $X_{\tau(n)} = X(n)$ for $n = 1, \ldots, N$. In other words, $\tau(n)$ is the index of the item corresponding to the $n$th order statistic of the valuations of the customer. In order to break ties in the case in which a customer values multiple items identically, we assume that the customer prefers the item with the lowest index amongst all such items. Thus, $\tau$ is uniquely defined, and the cost for the firm to sell a bundle of size $n$ is a random variable given by

$$c_N(n) = \sum_{k=0}^{n-1} c_{\tau(N-k)}.$$  \hfill (1)

Putting the above together, we now have that given a menu $p = (p(1), \ldots, p(N))$ of bundle-size prices, the firm’s realized profit is given by

$$\pi(p) = \sum_{n=1}^{N} (p(n) - c_N(n))1\{\zeta(X, p) = n\}.$$  \hfill (2)

Assuming that the firm is risk-neutral and interested in maximizing its expected profit, its problem is to find the optimal price menu $p \in (\mathbb{R}_+ \cup \infty)^{N+1}$ solving the optimization problem

$$[\text{BSP}] \quad \sup_{p \in (\mathbb{R}_+ \cup \infty)^{N+1}} \mathbb{E}[\pi(p)] = \sup_{p \in (\mathbb{R}_+ \cup \infty)^{N+1}} \sum_{n=1}^{N} \mathbb{E}[(p(n) - c_N(n))1\{\zeta(X, p) = n\}].$$ \hfill (3)

In general, there may be multiple optimal solutions to the above optimization problem, and we denote by $\mathcal{P}^*$, its set of optimal solutions. Hence, letting $p^* \in \mathcal{P}^*$ it follows that $\mathbb{E}[\pi(p^*)]$ is the maximum expected profit that the firm may obtain under a BSP policy.

Solving for $\mathbb{E}[\pi(p^*)]$ is difficult since the optimization problem (3) is not necessarily concave and may not admit closed forms. To the best of our knowledge, Chu et al. (2011) are the only ones who solve this problem using numerical simulations for up to 5 items. Our approach is different in the sense that we develop a theoretical framework to study the BSP problem as the number of items grows large.
3.1. The Asymptotic Regime

The asymptotic regime, we consider in this paper is similar to the one developed in Bakos and Brynjolfsson (1999). In particular, we begin as in Bakos and Brynjolfsson (1999) by assuming that $N$, the number of products that the firm offers, approaches infinity. As an example, one can take Netflix, which has over 5,000 titles in its online streaming catalog, or a cable TV provider such as Verizon, which offers hundreds of channels for its subscribers to watch.

Furthermore, we assume that the fraction of items whose valuations are below any value $x$ is relatively stable and its limit exists (in the weak sense). That is, the normalized sum of valuations exhibits statistical concentration. In particular, denote by

$$F_{X,N}(x) = \frac{1}{N} \sum_{n=1}^{N} 1\{X_n \leq x\}, \quad x \geq 0 \quad (4)$$

the empirical distribution function of the components of $X$, and let $F_X$ denote the distribution function of a nonnegative random variable with finite mean $\mu$.

**Assumption 1.** The sequence of empirical distribution functions $\{F_{X,N}, N \geq 1\}$ satisfies

$$F_{X,N}(x) \Rightarrow F_X(x) \quad \text{for} \quad x \geq 0 \quad \text{as} \quad N \to \infty. \quad (5)$$

Notice that since $F_X$ is nondecreasing, we also have uniform convergence. This assumption is a weaker version of the well-known Glivenko-Cantelli Lemma and does not require that the components of $X$ be i.i.d. (see Billingsley (1999) for further discussion). In order to simplify our proofs, we also make the technical assumption on $F_X$ that

$$\int_{0}^{\infty} (F_X(x)(1-F_X(x)))^{1/2}dx < \infty. \quad (6)$$

This condition (6) is satisfied by most known distributions and is only slightly more restrictive than requiring a finite second moment on $F_X$, but in certain cases the two conditions are equivalent (see Feller (1971) for further discussion).

**Assumption 2.** The collection of random variables $\{X_n, n \geq 1\}$ is uniformly integrable, and the weak law of large number (WLLN) holds where

$$\sum_{n=1}^{N} X_n/N \Rightarrow \mu \quad \text{as} \quad N \to \infty. \quad (7)$$

We also assume that fraction of marginal costs that fall below any value $c$ is relatively stable and its limit exists (this level of generality will be used in Sections 4.3 and 4.4). In order to technically state our assumption, denote

$$F_{C,N}(c) = \frac{1}{N} \sum_{n=1}^{N} 1\{c_n \leq c\}, \quad x \geq 0, \quad (8)$$
denote the empirical function of the marginal costs normalized, and let $F_C$ denote the distribution function of a nonnegative random variable with finite mean $\mu_C$. We then have the following assumption.

**Assumption 3.** The marginal costs are bounded and the sequence of empirical distribution functions $\{F_{C,N}, N \geq 1\}$ satisfies

$$F_{C,N}(c) \Rightarrow F_C(c) \text{ for } c \geq 0 \text{ as } N \to \infty. \quad (9)$$

Taken together, we refer to the above three assumptions as the large-offering regime.

We use the following asymptotic notation to describe the behavior of functions relative to $N$.

**Definition 1 (Asymptotic Notation).** Let $f(N), h(N) : N \to \mathbb{R}$. We say

$$f(N) \in \omega_+(h(N)) \text{ if } \liminf_{N \to \infty} f(N)/h(N) \to +\infty, \text{ and}$$

$$f(N) \in o(h(N)) \text{ if } \limsup_{N \to \infty} f(N)/h(N) \to 0.$$

Throughout the paper, our focus is on studying the asymptotic properties of a sequence of BSP menus in the large-offering regime. Hence, let $\{p_N, N \geq 1\}$ be a sequence of pricing menus where $p_N \in \mathbb{R}_+^N$ for each $N \geq 1$. We then have the following.

**Definition 2 (Asymptotic Optimality).** A sequence $\{p_N, N \geq 1\}$ of BSP menus is asymptotically optimal if

$$E[\pi(p_N)]/E[\pi(p^*]) \to 1 \text{ as } N \to \infty. \quad (10)$$

We are also interested in comparing the expected profit of an asymptotically optimal BSP policy with the expected profit under perfect price distribution, given by

$$\text{PPD} = \sum_{n=1}^{N} E[(X_n - c_n)^+],$$

which is the expected profit that the firm would obtain under perfect discrimination in which it knew in advance each of the customers’ item valuations. In fact, it is the maximum possible profit that the firm may achieve under any selling mechanism, not just BSP.

Armstrong (1999) has shown that in the current model setup and under certain regularity conditions, if the firm considers the cost-based two-part tariff policy where the firm prices the tariff at slightly below $\sum_{n=1}^{N} E[(X_n - c_n)^+]$ and allows the customers to buy any item at its respective marginal cost $c_n$, then the ratio of the optimal expected profit divided by $\sum_{n=1}^{N} E[(X_n - c)^+]$ goes to 1 as the number of items $N$ goes to infinity. However, as mentioned in the literature review, several behavioral-research studies suggest that customers in practice exhibit lower utilities compared to simpler bundling strategies due to the “pain of paying.” In this paper, we are interested in characterizing the asymptotic gap between the expected profit of the relatively simple BSP selling strategy relative to perfect price discrimination (or equivalently, the two-part tariff by Armstrong (1999)).
4. Large-Scale Bundle-Size Pricing

4.1. General Approach

In general, our framework is based on showing that as $N \to \infty$, the uncertainty in the normalized BSP valuations and costs is resolved and the limiting problem becomes a simple deterministic pricing problem. Hence, under the base model where customers have unlimited valuations and draw valuations from the same distribution, there exists a simple asymptotically optimal BSP policy where only a single size is offered rather than $N$ sizes (in Section 5.3, we show that when customers have random budgets, more than a single size is needed).

More specifically, letting

$$\bar{V}_N(t) = \frac{1}{N} V_N([Nt])$$

$$\bar{c}_N(t) = \frac{1}{N} c_N([Nt]),$$

we show that there exist nondecreasing and continuous (deterministic) functions $\bar{V}(t)$ and $\bar{c}(t)$ such that $\bar{V}_N(t) \Rightarrow \bar{V}(t)$ and $\bar{c}_N(t) \Rightarrow \bar{c}(t)$ as $N \to \infty$. Hence, the asymptotically optimal BSP size is given by $t^* \in \arg\max \bar{V}(t) - \bar{c}(t)$. We formalize this in the following theorem.

**Theorem 1.** Assume there exist nondecreasing functions $\bar{V}(t)$ and $\bar{c}(t)$ such that $\bar{V}_N(t) \Rightarrow \bar{V}(t)$ and $\bar{c}_N(t) \Rightarrow \bar{c}(t)$ for any $t \in [0,1]$ as $N \to \infty$. An asymptotically optimal BSP policy is to offer a **single size** $[Nt^*]$, where $t^* \in \arg\max \bar{V}(t) - \bar{c}(t)$. Moreover, let $p_N \in \mathbb{R}^N$ be the pricing menu defined by

$$p_N(n) = \begin{cases} N\bar{V}(t^*) - g(N) & \text{if } n = [Nt^*], \\ +\infty & \text{if } n \neq [Nt^*], \end{cases}$$

where $g(N) \in \omega^+ \left( \sqrt{\mathbb{E} \left[ (V_N([Nt^*]) - N\bar{V}(t^*))^2 \right]} \right) \cap o(N)$, the sequence $\{p_N, N \geq 1\}$ is asymptotically optimal, and we have

$$\lim_{N \to \infty} \frac{\mathbb{E}[\pi(p_N)]}{\mathbb{E}[\pi(p^*)]} = \lim_{N \to \infty} \frac{\mathbb{E}[\pi(p_N)]}{N(\bar{V}(t^*) - \bar{c}(t^*))} = 1.$$

A few comments are in order regarding the choice of the function $g(N)$. First of all, there exists a continuum of asymptotically optimal pricing policies; however, $g(N)$ needs to grow faster than $\sqrt{\mathbb{E} \left[ (V_N([Nt^*]) - N\bar{V}(t^*))^2 \right]}$ but slower than $N$. Notice that by Assumption 2, we have $\bar{V}(1) = \mu$, and roughly speaking, $\mathbb{E} \left[ (V_N([Nt^*]) - N\bar{V}(t^*))^2 \right]$ represents the variance of the sum of the top $[Nt^*]$ order statistics of $X$, which in general does not admit a closed form. For such a function $g(N)$ to exist, the variance of the sum of top $[Nt^*]$ order statistics needs to grow slower than $N^2$, or equivalently, the valuations should not be perfectly correlated. That is, there exists sufficient mixing such that statistical concentration in the average valuations occurs in the
large-offering regime. We will show in Section 4.2 that when the valuations are i.i.d., then \( g_N \in \omega_+ \left( \sqrt{N} \right) \cap o(N) \) and in Section 5.4 that the function that achieves the fastest convergence rate when \( t^* = 1 \) is \( g(N) = \sigma \sqrt{N \log N} = \sqrt{\text{Var}(V_N(N) \log N)} \). This suggests that the best choice of \( g(N) \) is \( \sqrt{\text{Var}(V_N([N t^*]))} \log N \), where \( \text{Var}(V_N([N t^*])) \) can be simulated numerically. This is consistent with the extensive numerical simulations done by Abdallah (2019) regarding \( g(N) \) in the case of the pure bundle, i.e., \( t^* = 1 \).

Theorem 1 provides a general overview of our framework in the large offering regime, where in the limit the BSP problem becomes a simple deterministic problem. However, the main technical challenge remains to establish the convergence and characterize \( \bar{V}(t) \) and \( \bar{c}(t) \) as functions of the model primitives. To do this, we will consider two cases: (1) homogeneous and (2) heterogeneous costs. The case of homogeneous costs is relatively easier since the cost for any bundle of size \( n \) is deterministic and is given by \( n \cdot c \), and hence \( \bar{c}_N(t) = \bar{c}(t) = t \cdot c \). In this case, roughly speaking, as long as the valuations are not perfectly correlated, we can prove our results. The case of heterogeneous costs is more challenging since the cost curve is a random variable and is not necessarily independent of the customer’s valuation; for instance, items with higher marginal costs can have higher valuations and are more likely to be included in a bundle. Therefore, we need more structure on the relationship between the costs and the valuations in order to characterize the limiting cost curve of the bundle. For this reason, in Section 4.3 we study the case of item-dependent marginal costs with i.i.d. valuation, and finally in Section 4.4 we discuss a more general case of item-dependent marginal costs with correlated valuations.

4.2. Homogeneous Marginal Costs

In this basic model, we assume that \( c_n = c \geq 0 \) for each \( n = 1, \ldots, N \). Let \( F_X^{-1}(s) = \inf\{x \geq 0 : F_X(x) \geq s\} \) for \( s \in (0, 1) \) be the quantile function of \( F_X \) (see Chapter 21 of Van der Vaart (2000)). We then have the following result.

**Theorem 2.** *In the presence of identical marginal costs \( c \geq 0 \), for \( 0 \leq t \leq 1 \) and as \( N \to \infty \), we have*

\[
\bar{V}_N(t) \Rightarrow \bar{V}(t) = \int_0^t F_X^{-1}(1 - s)ds, \\
\bar{c}_N(t) \Rightarrow \bar{c}(t) = t \cdot c.
\]

*Moreover, \( t^* = \sup\{s \in [0, 1] : F_X^{-1}(1 - s) \geq c\} \), and if \( c \) is a continuity point of \( F_X \), \( t^* = 1 - F_X(c) \). Also,*

\[
\lim_{N \to \infty} \frac{\mathbb{E} [\pi(p_N^*)]}{\text{PPD}} = 1.
\]
Theorem 2 provides a closed-form characterization of $\bar{V}(t)$ and $c(t)$ as well as the optimal size $t^*$. It also implies that BSP asymptotically achieves perfect price discrimination under homogeneous marginal costs. We note that this generalizes the result of Bakos and Brynjolfsson (1999) from asymptotic perfect price discrimination of pure bundling when $c = 0$ to asymptotic perfect price discrimination of BSP when $c \geq 0$, where BSP reduces to a pure bundling policy when $c = 0$.

We elaborate on the impact of cost on the optimal BSP in Figure 2. We provide 4 illustrations of the limiting valuation curve $\bar{V}(t)$ relative to the marginal cost curve: the case in which pure bundling is optimal, two cases in which a bundle of proportion $0 < t^* < 1$ is optimal, and finally the case in which it is optimal for the firm to offer no bundles at all.

We close this section by providing a concrete example to illustrate the applicability of our results. Consider an online content provider offering a large number of books, songs or movies for sale. The provider has $N = 1,000$ titles in its library and each item has a marginal cost of $c = 80 \text{¢}$. Suppose that a customer valuation vector for the 1,000 items has an empirical distribution that is approximated on $0 \leq x \leq $1 by $F_X = \sqrt{x}$ for $0 \leq x \leq $1. In this case, we have from Theorem 2 that the proportion of items that the firm should offer in a bundle is given by $t^* = 1 - F_X(c) = 10.56\%$, which corresponds to a bundle of about 105 items. The annual price that the firm should set is approximately $1,000 \bar{V}(t^*) = $94.82, at a cost to the firm of $105 \times 80 \text{¢} = $84, thereby yielding an annual profit per customer of around $10.36. Equivalently, the content provider can offer a monthly subscription service of up to $105/12 \approx 9$ items per month for a monthly price of $7.9.

4.3. Heterogeneous Marginal Costs with IID Valuations

In many cases the firm may offer products with substantially varying marginal costs. We therefore now consider the general marginal cost setup. First, we make the simplifying assumption that
the components of the customer valuation vector are i.i.d. with common distribution $F_X$. In the
subsection that follows, we remove this assumption.

**Theorem 3.** In the presence of heterogeneous marginal costs and i.i.d. valuations, for $0 \leq t \leq 1$
and as $N \to \infty$, we have

$$
\bar{V}_N(t) \Rightarrow \bar{V}(t) = \int_0^t F_X^{-1}(1-s)ds,
$$
$$
\bar{c}_N(t) \Rightarrow \bar{c}(t) = t \cdot \bar{c},
$$

where $\bar{c} = \int_0^\infty cdF_C(c)$. Moreover, $t^* = \sup\{s \in [0,1] : F_X^{-1}(1-s) \geq \bar{c}\}$ and if $c$ is a continuity point
of $F_X$, $t^* = 1 - F_X(\bar{c})$. Also,

$$
\lim_{N \to \infty} \frac{\mathbb{E}[\pi(p_N^*)]}{N\mathbb{E}[(X_1 - \bar{c})^+]} = 1. \quad (13)
$$

Theorem 3 implies that in the case of heterogeneous marginal costs the asymptotically optimal
BSP policy is the same as that with homogeneous marginal costs equal to $\bar{c}$. However in this case
the asymptotically optimal policy does not achieve perfect price discrimination, PP. This can be
seen by noting that by Jensen’s inequality we have that

$$
\mathbb{E}[(X_1 - \bar{c})^+] \leq \int_{\mathbb{R}^+} \mathbb{E}[(X_1 - c)^+]dF_C(c), \quad (14)
$$

where the right-hand side of the inequality is the limiting normalized expected profit under perfect
price discrimination.

Now consider the previous example of an online content provider with a catalog of 1,000 items.
As before, customers have (in this case i.i.d.) a limiting valuation distribution for each item given
by $F_X(x) = \sqrt{x}$ for $0 \leq x \leq 1$. Suppose now that instead of the marginal cost of each item being
80¢, it is now either 80¢ + δ or 80¢ − δ, where $0 \leq \delta \leq 20$¢ is equally divided among the 1,000 items.
In this case, if one uses Theorem 3 to subtract the asymptotically optimal expected profit under
BSP from the profit expected under perfect price discrimination, one obtains

$$
1,000 \times \frac{1}{3} \left[((0.80 + \delta)^{3/2} - (0.80)^{3/2}) - (0.80)^{3/2} - (0.80 - \delta)^{3/2}\right] \geq 0, \quad (15)
$$
a quantity that is increasing in $\delta$.

We now generalize the above formula. Let

$$
\bar{\Delta}_N = \frac{1}{N} \left(\sum_{n=1}^N \mathbb{E}[(X_n - c_n)^+] - \mathbb{E}[\pi(p_N^*)]\right)
$$

be the gap between the expected profit under perfect price discrimination and the optimal BSP
policy. We then have the following.
Corollary 1. In the present setup, \( \bar{\Delta}_N \to \bar{\Delta} \) as \( N \to \infty \), where

\[
\bar{\Delta} = \int_0^\infty \int_0^\infty (\text{sgn}(x - \bar{c}) \cdot (c - x))^+ dF_C(c) dF_X(x).
\] (16)

We next provide an upper and lower bound on \( \bar{\Delta} \). Let \( S_C \) denote the support of \( F_C \), and set \( c_u = \sup S_C \) and \( c_l = \inf S_C \) to be the upper and lower boundaries of \( S_C \). Also, let \( \sigma^2_C \) denote the variance of \( F_C \). We then have the following result.

Corollary 2. Let \( \delta_1 = \inf \{ f_X(x) : x \in S_C \} \) and \( \delta_2 = \sup \{ f_X(x) : x \in S_C \} \), and assume that \( F_X \) is continuous with density \( f_X \). We then have the bounds

\[
\delta_1 \sigma^2_C / 2 \leq \bar{\Delta} \leq \min\{\delta_2 \sigma^2_C / 2, c_u - c_l\}.
\] (17)

The above bound is tight, as can be shown in the following example. Suppose that \( F_X \) is a uniform distribution on \([a, b]\), where \( c_l \leq a \leq b \leq c_u \). Then, \( \delta_1 = \delta_2 = 1/(b - a) \) and \( \bar{\Delta} = \sigma^2_C / (2(b - a)) \). Note that the value of \( \sigma^2_C \) provides a measure of heterogeneity in the marginal costs, and the length of \( S_C \), which is given by \( c_u - c_l \), represents another measure of heterogeneity in the marginal costs.

4.4. Cost-Dependent Valuations

In most cases, the valuations and the costs are correlated in the sense that items with higher costs tend to have higher valuations. For this reason, we now also assume that each item’s valuation distribution depends on its marginal cost and, in particular, is proportional to it. Technically speaking, this assumption is stated as follows.

Assumption 4. For each \( n = 1, \ldots, N \), we have that \( X_n = c_n Z_n \), where \( \{Z_n ; n = 1, \ldots, N\} \) is an i.i.d. sequence of random variables with common continuous distribution \( F_Z \) that has a finite mean.

We continue to consider the case where items have heterogeneous marginal costs satisfying Assumption 3. The following is our result.

Theorem 4. In the presence of cost-dependent valuations, for \( 0 \leq t \leq 1 \) and as \( N \to \infty \), we have

\[
\bar{V}_N(t) \Rightarrow \bar{V}(t) = F^{-1}_X(1 - s) ds,
\]

\[
\bar{c}_N(t) \Rightarrow \bar{c}(t) = \int_0^\infty c (1 - F_Z(F_X^{-1}(1 - t)/c)) dF_C(c),
\]

where \( F_X(x) = \int_0^\infty F_Z(x/c) dF_C(c) \). Moreover, \( t^* = \arg \max \{ t \in [0,1] : \bar{V}(t) - c(t) \} \). Also,

\[
\lim_{N \to \infty} \frac{\mathbb{E}[\pi(p^*_N) \mathbb{I}[t^* = t]]}{N \left( \int_0^\infty \int_{F^{-1}_X(1 - s)/c}^\infty c(z - 1) dF_Z(z) dF_C(c) \right)} = 1.
\] (18)
In the present setting, the limiting normalized profit under perfect price discrimination is given by

$$\frac{1}{N} \text{PPD} = \frac{1}{N} \sum_{n=1}^{N} (X_n - c_n)^+ = \frac{1}{N} \sum_{n=1}^{N} c_n (Z_n - 1)^+ \Rightarrow \bar{c} \mathbb{E}[(Z_1 - 1)^+] \text{ as } N \to \infty.$$ 

As is shown below, in general, asymptotic perfect price discrimination cannot be achieved in the present setup; however, in certain special cases, it can (for example, if \( F_Z(1) = 0 \)). More generally, it turns out that a necessary and sufficient condition can be provided for when asymptotic perfect price discrimination is achieved. Let \( S_C, S_Z \subset \mathbb{R}_+ \) denote the supports of \( F_C \) and \( F_Z \), respectively. Define \( z_u = \inf(S_Z \cap [1, \infty)) \) and \( z_l = \sup(S_Z \cap [0, 1]) \). Loosely speaking, \( z_u \) represents the smallest value greater than 1 that \( Z_n \) may achieve, and \( z_l \) represents the largest value less than 1 that \( Z_n \) may achieve. Also denote by \( c_l = \inf S_C \) and \( c_u = \sup S_C \) the lower and upper limits of the support of \( F_C \), respectively. We further assume that \( c_l > 0 \) and \( c_u < \infty \). We then have the following result.

**Proposition 1.** In the present setup, the following statements are equivalent:

(i) Asymptotic perfect price discrimination may be achieved.

(ii) The inequality \( z_l c_u \leq z_u c_l \) holds.

Specifically, if the inequality in (ii) holds, then asymptotic perfect price discrimination is achieved by setting \( t^* = 1 - F_Z(1) \).

Proposition 1 implies that perfect price discrimination may only be achieved if \( F_Z \) places zero mass in some neighborhood of 1 or if \( F_C \) is degenerate.

5. **Extensions**

We now discuss several extensions to the baseline model to show the robustness of our framework and discuss some interesting implications of such variants. To simplify the analysis, we will assume that valuations are i.i.d.

5.1. **Improving BSP for Heterogeneous Costs**

As discussed in Section 4.3, BSP is inefficient in extracting the consumer surplus relative to PPD. We next analyze two new variants of BSP that can help improve its efficiency.

**Clustered Bundle-Size Pricing.** In clustered BSP the firm clusters items with equal marginal costs and does BSP on each cluster separately. In particular, suppose that products can have \( K > 1 \) distinct marginal costs denoted by \( c_1, \ldots, c_K \), where we assume that \( c_1 < \ldots < c_K \). We say that an item is of type \( k \) if its marginal cost is equal to \( c_k \). Now let \( N_k \) denote the number of items of type \( k \), where \( N_1 + \ldots + N_K = N \), the total number of items. In clustered BSP, the firm allows customers to select a bundle of size \( 0 \leq n_k \leq N_k \) from the type \( k \) items for each \( k = 1, \ldots, K \).
Because customers may not create bundles across different item types and all items of the same type have identical marginal costs, it follows that the optimal clustered BSP menu decomposes into \( k \) separate BSP problems with constant marginal costs. In particular, suppose that \( N_k/N \to \alpha_k \) as \( N \to \infty \). Hence, applying Theorem 2 we have the following result.

**Proposition 2.** If \( N_k/N \to \alpha_k \) for \( k = 1, \ldots, K \), then \( K \)-BSP achieves asymptotic perfect price discrimination.

Scribd’s new business model closely resembles clustered BSP with 4 clusters: news & magazines, ebooks, audiobooks, and documents. Scribd could have potentially created a larger number of clusters in order to improve the efficiency of its selling strategy; however, this may have had an adverse effect by increasing its customers’ cognitive burden thus decreasing their utility. For this reason, we next introduce a selling strategy that avoids any clustering.

**Assorted Bundle-Size Pricing.** In assorted BSP, the firm only allows customers to create a bundle out of a preselected subset (referred to as an assortment) of its items. Bundles of the same size have the same price as in BSP. The goal of this approach is for the firm to avoid selling high-cost items at a relatively low average price. A solution to the assorted BSP problem is characterized by the subset of items offered, the offered bundle sizes, and the corresponding prices for each bundle. This approach can be particularly appealing when the firm wants to avoid offering a complicated menu of prices as in clustered BSP. We note that even though as mentioned above Scribd primarily offers a clustered BSP selling strategy, it recently reduced the size of its catalog of romance books due to the large costs involved in offering these titles. Hence, Scribd has adopted an assorted BSP model within the ebooks’ cluster.

We now discuss the limiting version of the optimization problem that the firm must solve in order to find the optimal assorted BSP menu. Suppose as in clustered BSP that products have only \( K > 1 \) distinct marginal costs denoted by \( c_1, \ldots, c_K \), where \( c_1 < \ldots < c_K \), and that \( N_k/N \to \alpha_k \) as \( N \to \infty \). The firm must then decide on an assortment vector \( y = (y_1, y_2, \ldots, y_K) \), where \( y_k \in [0, \alpha_k] \) for each \( k = 1, \ldots, K \). Note that \( y_k \) represents the fraction of type \( k \) items that will be offered to the customer. Assuming that an assortment \( y \) is selected, the limiting weighted average cost for selling an item is then given by

\[
\bar{c}(y) = \frac{\sum_{k=1}^{K} y_k c_k}{\sum_{k=1}^{K} y_k}.
\]

(19)

Then, a similar analysis to the above shows that, asymptotically, the expected profit of the optimal assorted BSP policy can be written as

\[
\max \left( \sum_{k=1}^{K} y_k \right) \left( \int_0^t F_X^{-1}(1 - s)ds - \bar{c}(y) \cdot t \right)
\]

where \( F_X \) is the cumulative distribution function of the marginal costs.
The objective function of this mathematical program is continuous and the feasible set is compact. Moreover, since $F_X$ is continuous, the asymptotically optimal bundle size is given by $t^* = 1 - F_X(\bar{c}(y_1^*, y_2^*, \ldots, y_K^*))$. Thus, the optimization problem can be simplified to

$$
\max \left( \sum_{k=1}^{K} y_k \right) \left( \int_{0}^{1-F_X(\bar{c}(y))} F_X^{-1}(1-s) ds - \bar{c}(y) \cdot (1 - F_X(\bar{c}(y))) \right)
$$

(20)

The following proposition characterizes the structure of the optimal solution to the limiting assorted BSP problem.

**Proposition 3.** There exists an optimal solution $(y_1^*, y_2^*, \ldots, y_K^*)$ to the optimization problem (20) for the assorted BSP problem that is a corner point of the feasible polytope and has a nested-in-cost structure. That is, an index $0 \leq K^* \leq K$ can be found such that $y_k^* = \alpha_k$ for $1 \leq k \leq K^*$ and $y_k^* = 0$ for $K^* < k \leq K$.

Proposition 3 implies that under assorted BSP, the asymptotically optimal assortment $y^*$ consists of all items whose marginal costs are less than or equal to $c_{K^*}$. Then, the optimal bundle size is found by assuming that all items in the assortment have identical marginal costs of $\bar{c}(y^*)$.

We now briefly compare the limiting expected profits of (regular) BSP, clustered BSP, and assorted BSP using our running example. Recall that we assume that a firm is offering $N = 1,000$ products with i.i.d. valuations with distribution function $F(x) = \sqrt{x}$ for $0 \leq x \leq 1$. The marginal costs for these items are either $80\text{¢} - \delta$ or $80\text{¢} + \delta$, each with probability $1/2$, where $0 \leq \delta \leq 20\text{¢}$. From Section 4.2 we have that if the firm uses the optimal (regular) BSP strategy assuming a marginal cost equal to the average of $80\text{¢}$, then a single bundle of about 105.57 items is offered at a price of about $94.819 and the expected profit is $10.36. On the other hand, suppose for concreteness that $\delta = 15\text{¢}$, so that the marginal cost of an item is either $65\text{¢}$ or $95\text{¢}$ each with probability $1/2$. Then, using (14) we have that the expected profit under perfect price discrimination is given by $16.66$. This may be achieved by using clustered BSP. On the other hand, solving the optimization problem (20), one finds that the optimal policy under assorted BSP is for the firm to restrict its offerings to only the 500 products with a lower marginal cost of $65\text{¢}$ and to allow customers to select a bundle of approximately 96.85 items. The optimal price of the bundle is about $79.3255$, and the expected profit to the firm will be around $16.3489$. Hence, the firm is able to achieve 98.11% of the profit under perfect price discrimination by using assorted BSP.
5.2. Nonadditive Valuations

In the baseline, we have assumed that the customer’s valuation for any bundle is additive. We now discuss how our analysis can be generalized to include subadditive or superadditive valuations so as to capture substitutability or complementarity effect.

When the valuation of the bundle is subadditive, then for any subset of items $S \subseteq \{1, \ldots, N\}$, we can define a concave function $G : \mathbb{R}_+ \mapsto \mathbb{R}_+$ that maps the additive valuation of the bundle to a subadditive valuation. That is, the customer’s valuation for a bundle of items $S$ is now given by $G(\sum_{n \in S} X_n)$. We also require that $G$ be continuous and increasing. For example, $G(\sum_{n \in S} X_n) = \sqrt{\sum_{n \in S} X_n}$. Similarly, in the case of superadditive valuations, $G$ can be chosen to be an increasing convex function. For example, $G(\sum_{n \in S} X_n) = (\sum_{n \in S} X_n)^2$.

Since $G$ is increasing, the customer’s valuations for a bundle of size $n$ is given by

$$G(V_N(n)) = G\left(\sum_{k=0}^{n-1} X_{(N-k)}\right).$$

(21)

Since $G$ is continuous, the rest of the asymptotic results hold by the continuous-mapping theorem. However, we note that the normalization should be modified accordingly.

5.3. Bundle-Size Pricing with Budgets

In this section, we consider the case where the customer has a budget constraint $b_N > 0$ that limits her ability to pay for a bundle. Our setting is general in the sense that we assume that the customer’s budget $b_N$ is private information and that the firm has no information on its value. For ease of exposition, we assume homogeneous marginal costs and that $\bar{b}_N = b_N/N \to \bar{b} \geq 0$ as $N \to \infty$.

For a given BSP menu $p = (p(0) = 0, p(1), \ldots, p(N))$ set by the firm, the customer will purchase the bundle of an appropriate size in order to maximize her surplus subject to her budget constraint. That is, she will purchase a bundle of size

$$\zeta(X, p, b) \in \arg \max_{\{n=0, \ldots, N: p(n) \leq b\}} (V_N(n) - p(n)).$$

(22)

In return, the firm’s realized profit is given by the random variable

$$\pi(p, b) = \sum_{n=1}^{N} (p(n) - nc) 1\{\zeta(X, p, b) = n\};$$

(23)

and the firm’s risk-neutral pricing problem is given by the optimization problem

$$\max_{p \in (\mathbb{R}_+ \cup \infty)^{N+1}} \sum_{n=1}^{N} (p(n) - nc) \mathbb{P}({\zeta(X, p, b) = n}).$$

Now denote by $V_N(n, b) = V_N(n) \wedge b_N$ the restricted utility of the customer with a budget of $b_N$ for a bundle of size $n$. The restricted utility for a bundle of size $n = 1, \ldots, N$ is intuitively
the customer’s maximum willingness to pay for the bundle given a budget of $b$. Moreover, it is straightforward to establish the following upper bound on the BSP profit:

$$\frac{\mathbb{E}[\pi(p, b)]}{\mathbb{E}[\sup_{n=0,1,\ldots,N} (V_N(n, b) - nc)]} \leq 1 \quad \text{for } p \in (\mathbb{R}_+ \cup \infty)^N,$$

(24)

where the denominator is an upper bound on the expected profit even if the firm knows the exact value of the budget $b_N$.

Since a customer’s willingness to pay differs depending upon her budget $b_N$, it is natural to expect that the firm will offer a variety of different bundle sizes (and prices), each for a customer of a specific budget. This is confirmed by the following theorem.

**Theorem 5.** In the presence of identical marginal costs $c \geq 0$ with an unknown customer budget $b_N \geq 0$, let $p_N \in \mathbb{R}^N$ be the pricing menu defined by

$$p_N([Nt]) = \begin{cases} NV(t) - h(t)g(N) & \text{if } t \leq t^*, \\ +\infty & \text{if } t > t^*, \end{cases}$$

where $g \in \omega^+(\sqrt{N}) \cap o(N)$, $t^* = \sup\{s \in [0,1]: F_X^{-1}(1-s) \geq c\}$, and $h(\cdot)$ is a strictly increasing continuous function. Then, the sequence \(\{p_N, N \geq 1\}\) is asymptotically optimal, and we have that

$$\lim_{N \to \infty} \frac{\mathbb{E}[\pi(p_N, b_N)]}{\mathbb{E}[\sup_{n=0,1,\ldots,N} (V_N(n, b_N) - nc)]} = 1.$$  

(25)

![Figure 3](image.png)

**Figure 3** BSP in the presence of presence of high and low budget customers.

We note that the policy does not depend on the budget level since the firm has no information about it. Also the assumption of a single customer is without loss of generality, as our result still holds if different customers have different budgets.
The intuition behind Theorem 5 is that since the function $h$ is increasing, the pricing curve is constructed such that the surplus of a customer with a budget $b$ is increasing for all of the bundles that the customer can afford. Hence, customers will self select and purchase the minimum of either the maximum bundle that their budget will allow or $\lfloor Nt^* \rfloor$. This is illustrated in Figure 3 for a population consisting of two customer types with either high or low budgets.

5.4. Convergence Rate
Recall from Theorem 2 that in the case of zero marginal costs $c = 0$, the asymptotically optimal BSP policy is to offer a full bundle of size $N$. Moreover, any sequence of pricing policies given by $p(N) = N\mu - g(N)$, where $g(N) \in \omega_+(\sqrt{N}) \cap o(N)$ is asymptotically optimal. We now provide a tighter characterization of the optimal function $g$ and also specify the converge rate of the limit in (13). To keep the analysis simple, we only consider the case of zero marginal costs and i.i.d. valuations with a common distribution $F_X$ with finite mean $\mu$ and variance $\sigma^2$. The analysis can be extended to nonzero identical marginal cost setting. In order to state our result, let $p^*(N)$ be an optimal solution to the pure bundling problem, and define $g^*(N) = N\mu - p^*(N)$. Also, assume that Cramer’s condition holds.

**Assumption 5 (Cramer’s Condition).** There exists a constant $K > 0$ such that $\mathbb{E}[\exp(tX_1)] < \infty$ for $|t| < K$.

We then have the following.

**Theorem 6.** Given Assumption 5 and $c = 0$, we have that

$$\frac{g^*(N)}{\sigma\sqrt{N\log N}} \to 1 \quad \text{as} \quad N \to \infty,$$

and

$$\frac{PPD - \mathbb{E}[\pi(p^*)]}{\sigma\sqrt{N\log N}} \to 1 \quad \text{as} \quad N \to \infty.$$

Note that Theorem 6 implies that the optimal pricing policy for the pure bundle is $p^*(N) = N\mu - \sigma\sqrt{N\log N} + o(\sqrt{N\log N})$, and that the ratio of the profits relative to perfect price discrimination is $\mathbb{E}[\pi(p^*)]/N\mu = 1 - \sigma\sqrt{\log N/N} + o(\sqrt{\log N/N})$.

6. Computational Results
In this section we compare the performance of our single-size BSP policy against the simulation-optimization methodology used in Chu et al. (2011), which solves a nonlinear optimization problem to find the “optimal” prices for all bundle sizes. We follow the simulation setup of Chu et al. (2011) with minor modifications to make sure the that problem can scale in $N$. We note that Chu et al.
Abdallah et al. (2011) have done extensive numerical simulations to compare BSP to other pricing policies such as component pricing, pure bundling, and mixed bundling. For this reason, we limit our scope to comparing our single-size BSP policy to the simulation-optimization approach by Chu et al. (2011).

We consider six different cases for the number of items \( N \in \{3, 4, 5, 10, 51000\} \). We simulate a heterogeneous population of size \( M = 10,000 \), where the customers’ valuations are drawn from the same multivariate distribution. We assume that all customers have additive valuations. We use five different parametric distributions to define the marginal distribution of each item’s valuation, as summarized in Table 1. Each simulation run uses the same distribution for all the items. Each marginal distribution is parametrized by \( \varphi \) where \( \varphi \) is a random number between 1 and 5.

<table>
<thead>
<tr>
<th>Marginal Distribution</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform(0, ( \varphi ))</td>
<td>( X_n ) follows a uniform distribution on the support ([0, \varphi]).</td>
</tr>
<tr>
<td>Lognormal(( \varphi ), 1)</td>
<td>( X_n ) follows a lognormal distribution with log mean ( \varphi ) and log variance 1.</td>
</tr>
<tr>
<td>Exponential(1/( \varphi ))</td>
<td>( X_n ) follows an exponential distribution with mean ( \varphi ).</td>
</tr>
<tr>
<td>Truncated Normal(( \varphi ), 1)</td>
<td>( X_n ) follows a truncated normal distribution at zero with mean ( \varphi ) and variance 1.</td>
</tr>
<tr>
<td>Truncated Logit (( \varphi ), 1)</td>
<td>( X_n ) follows a truncated extreme value distribution with location ( \varphi ) and scale 1.</td>
</tr>
</tbody>
</table>

In order to simulate correlated random variables, we use the Gaussian Copula, similarly to Chu et al. (2011) and Abdallah (2019). We use a correlation proxy \( \rho \in \{-1, -0.5, 0, 0.5, 1\} \). Finally, we consider four different cases for marginal costs, where the marginal costs are given by \( c_n = \gamma \ast \mu \), for all \( n = 1, \ldots, N \), with \( \gamma \in \{0, 0.3, 0.5, 0.8\} \). The summary of all parameters is given in Table 2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Different Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market size ( M )</td>
<td>10,000</td>
</tr>
<tr>
<td>Number of Items ( N )</td>
<td>3, 4, 5, 10, 50, 100, and 1000</td>
</tr>
<tr>
<td>Distribution Type</td>
<td>uniform, lognormal, exponential, and truncated normal</td>
</tr>
<tr>
<td>Distribution Parameter ( \varphi )</td>
<td>~([1, 5])</td>
</tr>
<tr>
<td>Correlation Proxy ( \rho )</td>
<td>(-1, -0.5, 0, 0.5, 1)</td>
</tr>
<tr>
<td>Marginal Cost Coefficient ( \gamma )</td>
<td>0, 0.3, 0.5 and 0.8</td>
</tr>
</tbody>
</table>

We consider two main BSP pricing policies based on our asymptotic results, which offer a single size \([N t^*]\) at

\[
\begin{align*}
(1) \quad (\text{log-pricing}) \quad & p_N([N t^*]) = N \bar{V}(t^*) - \sqrt{\text{VAR}(V_n([N t^*]))} \log(N) \\
(2) \quad (\text{loglog-pricing}) \quad & p_N([N t^*]) = N \bar{V}(t^*) - \sqrt{\text{VAR}(V_N([N t^*]))} \log \log(N),
\end{align*}
\]
The ratio of profits of a BSP with one size relative to N sizes, as in Chu et al. (2011): (left) log log-pricing with simulated $t^*$ and (right) log-pricing with closed forms.

where $\text{VAR}(V_N([Nt^*]))$ is calculated numerically. The log-pricing strategy is consistent with our analysis of convergence rate in Section 5.4. We chose the loglog-pricing strategy to improve the performance on small sizes, since the log-pricing strategy prices suggests low prices for small sizes.

For both pricing strategies we consider 4 different options to determine $t^*$ and $\bar{V}(t^*)$:

1. Set $t^* = 1 - F_X(c)$ and $\bar{V}(t^*) = \int_{t^*}^1 F_X^{-1}(1 - s)ds$ as in Section 4.
2. Set $t^* = \frac{1}{N} \arg\max_n N\bar{V}(n/N) - n \cdot c$ and set $\bar{V}(t^*) = \int_{t^*}^1 F_X^{-1}(1 - s)ds$.
3. Set $t^* = 1 - F_X(c)$, and numerically calculate $\bar{V}(t^*)$ as the normalized average value of the sum of top $\lceil Nt^* \rceil$.
4. Numerically calculate $\bar{V}(t)$ as the normalized average value of the sum of top $t$, and set $t^* = \frac{1}{N} \arg\max_n \bar{V}(n) - n \cdot c$.

The motivation for exploring option 2 is that for small N, a low $t^*$ might lead to offering no bundle, or equivalently, a size 0. For example, for $N = 3$, if $t^* = 1 - F_X(c) < 1/3$ then $\lceil Nt^* \rceil = 0$. Regarding option 3, we wanted to to explore the effect of not having a well defined $F_X$. Finally, option 4 captures both effects. In total, we explored eight different pricing policies.

We compare our pricing policies in terms of the ratio of profit of each our eight BSP policies relative to the profit using the nonlinear optimization approach by Chu et al. (2011). The performance of the eight different policies is available in Figure A1 in the appendix. In Figure 4, we highlight the performance of two particular policies: log log-pricing with simulated $t$ (option 2) for a low number of items ($N \leq 50$), and the log-pricing with $t^* = 1 - F_X(c)$ (option 1) for a medium to high number of items ($N > 50$).
Figure 5  Computational time of our eight pricing policies combined (left) and nonlinear optimization as in Chu et al. (2011) (right).

We observe that the log log-pricing policy with a single size is very competitive with the $N$ sizes and prices obtained by Chu et al. (2011) even for small $N \in \{3, 4, 5, 10, 50\}$. Meanwhile, the log pricing policy outperforms all pricing policies for medium to high number of items $N \geq 50$.

Finally, we compare the computational time of our asymptotic pricing policies with the nonlinear optimization approach by Chu et al. (2011) in Figure 5. We observe that our approach is more efficient even when the firm experiments with all eight pricing policies combined relative to using a nonlinear optimization approach.

7. Conclusion

Although bundle-size pricing (BSP) is a simple form of bundling, it is a hard multidimensional problem that involves optimizing over order statistics. In this paper, we have presented a simple and tractable theoretical framework to study the BSP problem. Our framework is based on studying the large-scale BSP problem for a multiproduct firm that sells a large number of products. We show that in the limit, the BSP problem transforms from a hard multidimensional problem to a simple multi-unit monopolistic pricing problem with concave nondecreasing utility. This allows us to provide closed-form solutions for the asymptotically optimal sizes and prices along with characterizing the asymptotically optimal BSP profit relative to more complicated bundling policies such as mixed bundling.

More specifically, we show that when customers draw their valuations from the same distribution, then, as the number of items grows large, the asymptotically optimal BSP involves selling only one size. However, regarding the performance of BSP to relative mixed bundling, we show that it highly depends on the properties of the products' marginal costs. In particular, when the marginal
costs are equal across products, both BSP and mixed bundling asymptotically achieve the expected profit under perfect price discrimination, regardless of the level of marginal costs. However, BSP suffers from a serious limitation when the marginal costs are heterogeneous across products. In particular, in the presence of heterogeneous marginal costs, BSP can no longer asymptotically achieve perfect price discrimination (which can be achieved by mixed bundling).

In order to overcome the limitations of BSP in the presence of heterogeneous marginal costs, we propose two new BSP policies, which we call “clustered BSP” and “assorted BSP.” In clustered BSP, the items are clustered into groups of products with homogeneous marginal costs, and BSP is then applied for each cluster separately. Meanwhile, in assorted BSP, the firm first decides on an optimal assortment of products to offer to the customers and then uses BSP on this offered assortment. Both policies can significantly improve the performance of BSP relative to mixed bundling. In fact, we show that clustered BSP can asymptotically achieve perfect price discrimination.

Our framework also allows us to study richer models of BSP. We extend our analysis to a setting where customers have heterogeneous budgets that are unknown to the firm and customers draw their valuations from different distributions. In this case, one size is no longer optimal, but we can still characterize the asymptotically optimal BSP with multiple sizes.

Finally, we highlight some limitations of our framework, which is based on an asymptotic analysis where the number of items grows large. We note here that we provide an exact characterization of the convergence rate, which shows that the convergence is very fast in the case of i.i.d. valuations. In this regard, our framework, while not being exact, can be viewed as a reasonable approximation to a setting where a multiproduct firm sells medium to large number of items, as is the case with digital products such as movies, music, online articles, etc. or business-to-business settings such as ads-exchange and large-scale contracts. In addition, this framework allows us the uncover the “first-order” effects that govern the performance of BSP. Of course, characterizing the optimal BSP policy and understanding its properties for a finite number of item remains an open yet challenging problem.

Endnotes

References


Large-Scale Bundle-Size Pricing:
A Theoretical Analysis

APPENDIX

Tarek Abdallah
Kellogg School of Management, Northwestern University,
tarek.abdallah@kellogg.northwestern.edu

Arash Asadpour
Lyft Inc., New York, a.asadpour@gmail.com

Josh Reed
Leonard N. Stern School of Business, New York University, jreed@stern.nyu.edu

In this Appendix, we present the proofs of the statements in the main body of the paper.

Proofs of Section 4.

Proof of Theorem 1. Letting $\bar{\pi}(p) = \frac{1}{N} \pi(p)$, then for any $p \in \mathbb{R}_+^N$, we have $\bar{\pi}(p) \leq \sup_{t \in [0,1]} \bar{V}_N(t) - \bar{c}_N(t)$. Moreover, since $\bar{V}(t)$ and $\bar{c}(t)$ are monotone continuous functions then the convergence holds in $D([0,1], \mathbb{R})$ and hence $\bar{V}_N(\cdot) - \bar{c}_N(\cdot) \Rightarrow \bar{V}(\cdot) - \bar{c}(\cdot)$ in $D([0,1], \mathbb{R})$ as $N \to \infty$. Therefore, by interchanging the limit and the sup, we get $\sup_{t \in [0,1]} \bar{V}_N(t) - \bar{c}_N(t) \Rightarrow \sup_{t \in [0,1]} \bar{V}(t) - \bar{c}(t)$ as $N \to \infty$. Also by the dominated convergence theorem, the limit holds in expectation as well. Hence, we obtain

$$\limsup_{N \to \infty} \mathbb{E}[\bar{\pi}(p^*)] \leq \bar{V}(t^*) - \bar{c}(t^*). \tag{A1}$$

Next, noting that in general $c_N(n)$ and $1\{\zeta(X, p) = n\}$ in (3) are not necessarily independent, then under the pricing policy described in the theorem, the firm’s expected profit is bounded below by

$$\mathbb{E}[\bar{\pi}(p_N)] \geq \left(\bar{V}(t^*) - g(N)/N - \mathbb{E}[\bar{c}_N(t^*)]\right) \mathbb{P}(V_N(\lfloor N t^* \rfloor) > N\bar{V}(t^*) - g(N)). \tag{A2}$$

Since $\bar{c}_N(t)$ is bounded, then by the bounded convergence theorem we have that $\mathbb{E}[\bar{c}_N(t^*)] \to \bar{c}(t)$ as $N \to \infty$. Now noting that $g(N) \in o(N)$, we obtain

$$\bar{V}(t^*) - g(N)/N - \mathbb{E}[\bar{c}_N(t^*)] \Rightarrow \bar{V}(t^*) - \bar{c}(t^*) \text{ as } N \to \infty. \tag{A3}$$

Next, note that

$$\mathbb{P}(V_N(\lfloor N t^* \rfloor) > N\bar{V}(t^*) - g(N)) \geq 1 - \mathbb{P}\left(V_N(\lfloor N t^* \rfloor) - N\bar{V}(t^*)^2 \geq g^2(N)\right)$$

$$\geq 1 - \frac{\mathbb{E}\left(V_N(\lfloor N t^* \rfloor) - N\bar{V}(t^*)^2\right)}{g^2(N)} \tag{Markov’s Ineq.}$$

$$\to 1 \text{ as } N \to \infty \tag{A4}$$

where the convergence holds since $g(N) \in \omega_+$. 

$\sqrt{\mathbb{E}\left(V_N(\lfloor N t^* \rfloor) - N\bar{V}(t^*)^2\right)}$. 
Combining (A2), (A3) and (A4), we obtain that

$$\lim_{N \to \infty} \inf E[\bar{\pi}(p_N)] \geq \bar{V}(t^*) - \bar{\epsilon}(t^*). \quad (A5)$$

Finally, combining (A1) and (A5), we obtain

$$\lim_{N \to \infty} \mathbb{E}[\bar{\pi}(p_N)] = \lim_{N \to \infty} \mathbb{E}[\bar{\pi}(p^*)] = \bar{V}(t^*) - \bar{\epsilon}(t^*). \quad \Box$$

**Proof of Theorem 2**  The proof that $\bar{c}_N(t) \Rightarrow t \cdot c$ as $N \to \infty$ is straightforward and hence is skipped. Next, Recall from Assumption 2 that

$$\bar{V}_N(1) \Rightarrow \mu = \int_0^1 F_{X_N}^{-1}(1-s)ds \quad \text{as} \quad N \to \infty,$$

where the equality follows by an appropriate change-of-variables. Now notice that for each $n = 1, ..., N$, we have $X_{(n)} = F_{X_N}^{-1}(t)$ for $t \in ((n-1)/N, n/N]$ (see, for instance, Section 21 of Van der Vaart (2000)). Hence, using the definition of $\bar{V}_N$ it follows after some algebra that

$$\bar{V}_N(t) = \int_0^t F_{X_N}^{-1}(1-s)ds, \quad 0 \leq t \leq 1.$$  

However, by Assumption 1 and similar arguments to Lemma 21.2 of Van der Vaart (2000), we have that $F_{X_N}(t) = F_{X_N}(t)$ for every $t \in (0, 1)$ where $F_{X_N}(t)$ is continuous. Moreover, since $F_{X_N}^{-1}$ is monotonic it has at most a countable number of discontinuities, and hence by the bounded convergence theorem (see Lieb and Loss (2001)), and for each $t > 0$, we have that,

$$\bar{V}_N(1) - \bar{V}_N(t) = \int_t^1 F_{X_N}^{-1}(1-s)ds \Rightarrow \int_t^1 F_{X_N}^{-1}(1-s)ds \quad \text{as} \quad N \to \infty. \quad (A7)$$

Placing (A6) and (A7) together, we now obtain that for any $t \in [0, 1], \bar{V}_N(t) \Rightarrow \bar{V}(t)$, as $N \to \infty$. Moreover, it is straightforward to establish that $\bar{V}(\cdot)$ and $\bar{\epsilon}(\cdot)$ are continuous non-decreasing functions. Moreover,, the expression of $t^*$ follows from the FOC for the maximzer of $\bar{V}(t) - \bar{\epsilon}(t)$.

We are now left to show that $\frac{1}{N}\sum_{n=1}^{N} (X_n - c) = \frac{1}{N} \sum_{n=1}^{N} (X_{(n)} - c)$.

Recalling that $X_{(n)} = F_{X_N}^{-1}(t)$ for $t \in ((n-1)/N, n/N]$, we get

$$\frac{1}{N} \sum_{n=1}^{N} (X_n - c) = \int_0^1 (F_{X_N}^{-1}(1-s) - c)ds$$

$$\Rightarrow \int_0^1 (F_{X_N}^{-1}(1-s) - c)ds \quad \text{as} \quad N \to \infty,$$

where the convergence follows from the dominated convergence theorem. Furthermore, by the uniform integrability assumption, the convergence holds in expectation. Now letting $t^* = \sup\{s \in [0, 1] : F_{X_N}^{-1}(1-s) \geq c\}$ and since $F_{X_N}^{-1}(1-s)$ is non-increasing, we obtain

$$\int_0^1 (F_{X_N}^{-1}(1-s) - c)ds = \int_0^{t^*} (F_{X_N}^{-1}(1-s) - c)ds$$

$$= \bar{V}(t^*) - t^*c. \quad \Box$$
Proof of Theorem 3. The proof that $\bar{V}_N(t) \Rightarrow \bar{V}(t) = \int_0^t F_X^{-1}(1-s)ds$ as $N \to \infty$ follows from Theorem 2. Regarding the cost curve, note that for each $n = 1, \ldots, N$, we have that $\mathbb{E}[\bar{c}_N(t)] = t \cdot \bar{c}_N \to t \cdot \bar{c}$ as $N \to \infty$, where the convergence follows from Assumption 3. Hence to show that $\bar{c}_N(t) \Rightarrow t \cdot \bar{c}$, it suffices to show that $\text{Var}(\bar{c}_N(t)) \to 0$. Note that after some algebra we may write

$$\text{Var}(\bar{c}_N(t)) = \frac{1}{N^2} \sum_{k=0}^{[Nt]-1} \mathbb{E}[(c_{r(N-k)} - \bar{c}_N)^2] + \frac{2}{N^2} \sum_{k=0}^{[Nt]-1} \sum_{\ell=0}^{k-1} \mathbb{E}[(c_{r(N-k)} - \bar{c}_N)(c_{r(N-\ell)} - \bar{c}_N)],$$

where $\tau$ is the permutation function that maps the order statistics to their corresponding item indices.

Regarding the first term, recalling that $\mathbb{P}(\tau(n) = k) = 1/N$ for $k = 1, \ldots, N$, we have that

$$\frac{1}{N^2} \sum_{k=0}^{[Nt]-1} \mathbb{E}[(c_{r(N-k)} - \bar{c}_N)^2] = \frac{[Nt]}{N^2} \mathbb{E}[(c_{r(1)} - \bar{c}_N)^2] \to 0$$
as $N \to \infty$.

Regarding the second term, note that for $1 \leq i, j \leq N$ with $i \neq j$, we have that

$$\mathbb{P}(\tau(i) = j) = 1/(N-1) \text{ for } \ell = 1, \ldots, N \text{ with } \ell \neq k,$n

and so

$$\frac{2}{N^2} \sum_{k=0}^{[Nt]-1} \sum_{\ell=0}^{k-1} \mathbb{E}[(c_{r(N-k)} - \bar{c}_N)(c_{r(N-\ell)} - \bar{c}_N)] = \frac{N-1}{N} \mathbb{E}[(c_{r(1)} - \bar{c}_N)(c_{r(2)} - \bar{c}_N)].$$

Moreover, conditioning on $\tau(1)$ we obtain that

$$\mathbb{E}[c_{r(2)}|c_{r(1)}] = \frac{N}{N-1} \bar{c}_N - \frac{1}{N-1} c_{r(1)}.$$n

Therefore,

$$\frac{N-1}{N} \mathbb{E}[(c_{r(1)} - \bar{c}_N)(c_{r(2)} - \bar{c}_N)] = \frac{1}{N} \mathbb{E}[(c_{r(1)} - \bar{c}_N)^2] \to 0 \text{ as } N \to \infty.$$n

Hence, $\text{Var}(\bar{c}_N(t)) \to 0$ as desired. Finally, the choice of $t^*$ and the asymptotic profit follows similarly to Theorem 2. □

Proof of Corollary 1. From Theorem 3, it is straightforward to establish that

$$\bar{\Delta}_N \to \int_0^\infty \int_0^\infty (x-c)^+ dF_C(c)dF_X(x) - \int_0^\infty (x-\bar{c})^+ dF_X(x) \text{ as } N \to \infty,$$

where the interchange of the integration holds by Tonelli’s theorem. Now, for $x \geq \bar{c}$ it follows by some algebra that

$$\int_0^\infty \int_0^\infty (x-c)^+ dF_C(c)dF_X(x) - \int_0^\infty (x-c)^+ dF_X(x) = \int_0^\infty \int_0^\infty (c-x)^+ dF_C(c)dF_X(x).$$

On the other hand, for $x < \bar{c}$ we have

$$\int_0^\infty \int_0^\infty (x-c)^+ dF_C(c)dF_X(x) - \int_0^\infty (x-c)^+ dF_X(x) = \int_0^\infty \int_0^\infty (c-x)^+ dF_C(c)dF_X(x).$$

Putting the above together now yields the result. □
**Proof of Corollary 2.** Regarding the lower bound, by Tonelli’s theorem and some algebra,

\[ \Delta = \int_{\xi}^{\infty} \int_{\xi}^{c} (c-x) dF_X(x) dF_C(c) + \int_{c}^{\infty} \int_{\xi}^{c} (x-c) dF_X(x) dF_C(c) \]

\[ \geq \int_{\xi}^{\infty} \int_{\xi}^{c} (c-x) \delta_1 dx dF_C(c) + \int_{c}^{\infty} \int_{\xi}^{c} (x-c) \delta_1 dx dF_C(c) \]

\[ = \delta_1 / 2 \int_{0}^{\infty} (c-c)^2 dF_C(c) = \delta_1 \sigma_C^2 / 2. \]

The proof of the upper bound \( \delta_2 \sigma_C^2 / 2 \) follows analogously, replacing \( \delta_1 \) by \( \delta_2 \) and reversing the inequality.

As for the upper bound \( c_u - c_l \), we have

\[ \Delta = \int_{\xi}^{\infty} \int_{\xi}^{c} (c-x) dF_X(x) dF_C(c) + \int_{c}^{\infty} \int_{\xi}^{c} (x-c) dF_X(x) dF_C(c) \]

\[ \leq c_u - c + c - c_l = c_u - c_l. \quad \square \]

**Proof of Theorem 4** In order to show that \( \tilde{V}_N(t) \to \tilde{V}(t) \) as \( N \to \infty \) it suffices to show that for the empirical distribution of \( X \), \( \mathbb{P}_{X,N} \), we have \( \mathbb{P}_{X,N}(x) \to F_X(x) = \int_{\mathbb{R}} F_Z(x/c) dF_C(c) dc \) for any \( x \geq 0 \) (the rest follows similarly to Theorem 2).

Note that by the same arguments as used in the proof of the Glivenko-Cantelli theorem (see Theorem 19.1 of Van der Vaart (2000)), it suffices to show that for each \( x \geq 0 \) we have \( \mathbb{P}_{X,N}(x) \to F_X(x) \) as \( N \to \infty \). Next, since by Assumption 4, \( X_n = c_n Z_n \) for each \( n = 1, \ldots, N \), it follows after some algebra that for each \( x \geq 0 \) we may write

\[ \mathbb{P}_{X,N}(x) = \frac{1}{N} \sum_{n=1}^{N} 1\{Z_n \leq x/c(n)\} = \frac{1}{N} \Phi(N,x) + \int_{0}^{\infty} F_Z(x/c) d\mathbb{P}_{C,N}(c), \quad (A9) \]

where

\[ \Phi(N,x) = \sum_{n=1}^{N} (1\{Z_n \leq x/c(n)\} - F_Z(x/c(n))). \]

However, by the i.i.d. assumption on the sequence \( \{Z_n, n = 1, \ldots, N\} \), it is clear that the process \( \Phi(x) = \{\Phi(x), N \geq 1\} \) is a square-integrable martingale with predictable quadratic variation

\[ \langle \Phi(x) \rangle_N = \sum_{n=1}^{N} F_Z(x/c(n))(1 - F_Z(x/c(n))) = N \int_{0}^{\infty} F_Z(x/c)(1 - F_Z(x/c)) d\mathbb{P}_{C,N}(c), \]

for \( N \geq 1 \). On the other hand, by Assumption 3 and since \( F_Z \) is bounded and continuous, it follows that

\[ \int_{0}^{\infty} F_Z(x/c)(1 - F_Z(x/c)) d\mathbb{P}_{C,N}(c) \to \int_{0}^{\infty} F_Z(x/c)(1 - F_Z(x/c)) dF_C(c) \quad \text{as} \quad N \to \infty, \]

and so by the strong law of large numbers for martingales (see Section 2.6 of Liptser and Shiryaev (2012)), it follows that \( \mathbb{P} \)-a.s., \( (1/N)\Phi(N,x) \to 0 \) as \( N \to \infty \).

In addition, it follows from Assumption 3 and the assumption that \( F_Z \) is continuous, that

\[ \int_{0}^{\infty} F_Z(x/c) d\mathbb{P}_{C,N}(c) \to \int_{0}^{\infty} F_Z(x/c) dF_C(c) \quad \text{as} \quad N \to \infty. \]

Now the result immediately follows from the decomposition \( (A9) \).

Next regarding \( \tilde{e}_N(t) \Rightarrow \bar{e}(t) \), recall that as in (1) we may write

\[ \bar{e}_N(t) = \frac{1}{N} \sum_{n=0}^{\lfloor Nt \rfloor - 1} c_{e(t-n)}, \quad \text{for} \quad t \geq 0. \quad (A10) \]
However, since $F_Z$ is continuous, then the above is equivalent to summing over the marginal costs of all the items for which their valuation is larger than the valuation of the $\lfloor Nt \rfloor$th order statistic. In particular, we have that $\mathbb{P}$-a.s.,

$$
\sum_{n=0}^{\lfloor Nt \rfloor - 1} c_{r(n-n)} = \sum_{n=1}^{N} c_{n} 1\{X_n > X_{\lfloor Nt \rfloor + 1} \},
$$

where we set $X_{(N+1)} = \infty$. Now let

$$
\bar{\varepsilon}(t) = \frac{1}{N} \sum_{n=1}^{N} c_{n} (1\{X_n > X_{\lfloor Nt \rfloor + 1} \} - 1\{X_n > F_X^{-1}(1-t) \}), \quad t \in [0,1],
$$

and

$$
\bar{\delta}(t) = \frac{1}{N} \sum_{n=1}^{N} c_{n} (1\{X_n > F_X^{-1}(1-t) \} - (1 - F_Z(F_X^{-1}(1-t)/c_n))), \quad t \in [0,1].
$$

It then follows that we may write

$$
\bar{e}_n(t) = \frac{1}{N} \sum_{n=1}^{N} c_{n} (1 - F_Z(F_X^{-1}(1-t)/c_n)) + \bar{\varepsilon}(t) + \bar{\delta}(t), \quad t \in [0,1].
$$

Moreover, by Assumption 3 and using similar techniques used in proving $F_{X,N}(x) \Rightarrow F_X(x)$, $x \geq 0$, one may show that

$$
\frac{1}{N} \sum_{n=1}^{\lfloor Nt \rfloor} c_{n} (1 - F_Z(F_X^{-1}(1-t)/c_n)) \Rightarrow \int_{0}^{\infty} c(1 - F_Z(F_X^{-1}(1-t)/c)) dF_C(c) \quad \text{as} \quad N \to \infty.
$$

Hence, in order to complete the proof it suffices now to show that $\bar{\varepsilon}(t) \to 0$ and $\bar{\delta}(t) \to 0$ as $N \to \infty$.

Regarding $\bar{\varepsilon}(t)$, we show $\bar{\varepsilon}(t) \overset{L_1}{\to} 0$ as $N \to \infty$, which in return implies convergence in probability. First, notice that we have the following bound

$$
\mathbb{E}[\|\bar{\varepsilon}(t)\|] \leq \frac{1}{N} \sum_{n=1}^{N} c_{n} E[1\{X_n > X_{\lfloor Nt \rfloor + 1} \} - 1\{X_n > F_X^{-1}(1-t) \}].
$$

(A11)

Meanwhile, we have that $X_{\lfloor Nt \rfloor + 1} = F_X^{-1}(1-t)$, where $F_X^{-1}$ is the quantile function of the empirical distribution of the realized valuations $\{X_n; n=1, \ldots, N\}$. In addition, we have that

$$
1\{X_n > X_{\lfloor Nt \rfloor + 1} \} - 1\{X_n > F_X^{-1}(1-t) \} = 1\{\min(X_{\lfloor Nt \rfloor + 1}, F_X^{-1}(1-t)) < X_n < \max(X_{\lfloor Nt \rfloor + 1}, F_X^{-1}(1-t))\}
$$

Therefore, noting that $X_n = c_n Z_n$ and substituting into (A11), we obtain

$$
\mathbb{E}[\|\bar{\varepsilon}(t)\|]
\leq \frac{1}{N} \sum_{n=1}^{N} c_{n} E\left[1\{\min(F_X^{-1}(1-t), F_X^{-1}(1-t)) < c_n Z_n < \max(F_X^{-1}(1-t), F_X^{-1}(1-t))\}\right]
= \int_{0}^{\infty} c \left[ F_Z(\max(F_X^{-1}(1-t), F_X^{-1}(1-t))/c) - F_Z(\min(F_X^{-1}(1-t), F_X^{-1}(1-t))/c) \right] dF_C(c).
$$

However, since $F_{X,N}(x) \Rightarrow F_X(x)$, $x \geq 0$, and $F_X$ is continuous, we get that $F_X^{-1}(1-t) \Rightarrow F_X^{-1}(1-t)$ for $t \in (0,1)$ as $N \to \infty$. Therefore, by the continuous mapping theorem and for every $c > 0$, we get

$$
F_Z(\max(F_X^{-1}(1-t), F_X^{-1}(1-t))/c) - F_Z(\min(F_X^{-1}(1-t), F_X^{-1}(1-t))/c) \to 0 \quad \text{as} \quad N \to \infty.
$$
Consequently, it follows from Assumption 3 that the right hand side in the above also converges to 0 as $N \to \infty$. Hence, we get that $E[|\bar{\epsilon}(t)|] \to 0$ for $t \in (0,1)$ as $N \to \infty$, which in turn implies that $\bar{\epsilon}(t) \Rightarrow 0$ for $t \in (0,1)$ as $N \to \infty$. For $t = 0$ and $t = 1$, it is straightforward to establish the convergence result.

Regarding $\tilde{\delta}(t)$, note that since $\{X_n, 1 \leq n \leq N\}$ is i.i.d. it is straightforward to verify that $E[\tilde{\delta}(t)] = 0$

Moreover, using Assumption 3 we have that

Hence, it follows that $\tilde{\delta}(t) \Rightarrow 0$ as $N \to \infty$, and consequently $\tilde{\epsilon}_N(t) \Rightarrow \bar{\epsilon}(t)$ as $N \to \infty$.

Finally, using an appropriate change-of-variables and letting $u = F_X^{-1}(1-t)$, the limiting normalized BSP profit can be written as

$$V(t) - \bar{\epsilon}(t) = \int_{0}^{\infty} \int_{l}^{\infty} c(z-1) d\bar{F}_Z(z) d\bar{F}_C(c),$$ 

(A12)

as desired. □

**Proof of Proposition 1.** We first prove the if part. Consider the following bundle size $t = 1 - F_X(z_u c_l)$. Since $z_l c_u \leq z_u c_l$, then we have that $z_l \leq z_l c_u / c \leq z_u c_l / c \leq z_u$ for every $c \in [c_l, c_u]$ where $F_Z(z_l) = F_Z(z_u) = F_Z(1)$. Hence, it can be shown that $u = F_X^{-1}(F_X(z_u c_l)) = z_u c_u$. It now follows from (A12) that the limiting normalized profit under the proposed pricing policy is given by

$$V(t) - \bar{\epsilon}(t) = \int_{c_l}^{c_u} \int_{z_l c_u}^{\infty} c F_Z(z-1) d\bar{F}_Z(c) = E[(Z_1 - 1)^+] \bar{\epsilon}(1),$$

where the second equality is due to the fact that for $c \in [c_l, c_u]$ we have $z_l \leq z_l c_u / c \leq z_u$. Finally, since $z_l \leq z_u c_l / c \leq z_u$ for every $c \in [c_l, c_u]$, then it can be shown that $F_X(z_u c_l) = F_X(1)$ and hence $t = 1 - F_X(1)$ which concludes the proof of the if part.

We now prove the only if part. Assume that perfect price discrimination is asymptotically achieved for a given size $t^\star$ but $z_l c_u > z_u c_l$. Hence, we get that $z_l < 1 < z_u$ where $F_X(z_l) = F_X(z_u) = F_X(1)$. Let $u = F_X^{-1}(1-t^\star)$. Next, we consider three different cases for the values of $u$.

**Case I** ($u \leq z_u c_l$) We note that since $z_l c_u > z_u c_l$, then there exists $\hat{c} \in (c_l, c_u)$ such that $z_u c_l \geq z_l c$ for every $c \in [c_l, \hat{c}]$ and $z_l c \geq z_u c$ for every $c \in [\hat{c}, c_u]$. Hence, we have that $z_l \leq z_l c_u / c \leq z_u$ for $c \in [c_l, \hat{c}]$ and $z_l > z_u c_l / c$ for $c \in [\hat{c}, c_u]$. After some algebra, we get that

$$V(t) - \bar{\epsilon}(t) \leq E[(Z_1 - 1)^+] \bar{\epsilon}(1) + \int_{c_l}^{z_u} \int_{z_l c_u}^{\infty} c f_Z(z-1) d\bar{F}_Z(c).$$

However, since $u/c < z_l$ for $c \in [\hat{c}, c_u]$, we get that $\int_{c_l}^{z_u} \int_{z_l c_u}^{\infty} c f_Z(z-1) d\bar{F}_Z(c) < 0$ which is a contradiction.

**Case II** ($u \geq z_u c_l$) Since $z_l c_u > z_u c_l$, then there exists $\tilde{k} \in (c_l, c_u)$ such that $z_l c_u > z_l c$ for every $c \in [c_l, \tilde{k}]$ and $z_l c_l \leq z_l c$ for every $c \in [\tilde{k}, c_u]$. Hence, we have that $z_l c_u / c > z_u$ for any $c \in [c_l, \tilde{k}]$ and $z_l \leq z_l c_u / c \leq z_u$ for $c \in [\tilde{k}, c_u]$. Again after some algebra, we get that

$$V(t) - \bar{\epsilon}(t) \leq E[(Z_1 - 1)^+] \bar{\epsilon}(1) - \int_{\tilde{k}}^{c_u} \int_{z_l c_u}^{\infty} c f_Z(z-1) d\bar{F}_Z(c).$$

However, since $u/c > z_u$ for $c \in [\tilde{k}, c_u]$, we get that $\int_{\tilde{k}}^{c_u} \int_{z_l c_u}^{\infty} c f_Z(z-1) d\bar{F}_Z(c) > 0$ which is a contradiction.

**Case III** ($z_l c_u > u > z_u c_l$) We have that $u/c_u > z_l$ and $u/c_l > z_u$. Consequently there exists $c_1, c_2 \in (c_l, c_u)$ where $c_1 < c_2$ such that
As a result, we obtain
\[
F_k = \frac{\partial y_k}{\partial y_k} \mathcal{L}^{[0,\alpha_k)}
\]
After some algebra, the limiting BSP profit can be written as
\[
\tilde{V}(t) - \bar{c}(t) = \mathbb{E}[\mathcal{Z}(1) - \int_{c_1}^{u/c} \mathcal{F}(z) dz - \int_{u/c}^{z_1} \mathcal{F}(z) dz] + \int_{z_1}^{u/c} \mathcal{F}(z) dz + \int_{u/c}^{z_1} \mathcal{F}(z) dz.
\]
However, since \(u/c > z_1\) for any \(c \in [c_1, c_1]\) and \(u/c < z_1\) for any \(c \in (c_2, c_3]\), then
\[
\int_{c_1}^{u/c} \mathcal{F}(z) dz - \int_{u/c}^{z_1} \mathcal{F}(z) dz + \int_{u/c}^{z_1} \mathcal{F}(z) dz < 0,
\]
which is a contradiction. \(\square\)

Proofs of Section 5

**Proof of Proposition 3.** Before providing the proof of Proposition 3, we first state a lemma. Let \(H : \prod_{k=1}^{K} [0, \alpha_k] \rightarrow \mathbb{R}\) be the function defined by
\[
H(y_1, y_2, \ldots, y_K) = \left( \sum_{k=1}^{K} y_k \right) \left( \int_{0}^{1-P(\bar{c}(y))} F_X^{-1}(s) ds - \bar{c}(y) \cdot \left( 1 - F_X(\bar{c}(y)) \right) \right).
\]

**Lemma A1.** For each \((y_1, y_2, \ldots, y_K) \in \prod_{k=1}^{K} [0, \alpha_k]\), we have
\[
\frac{\partial^2}{\partial y_k^2} H(y_1, y_2, \ldots, y_K) > 0 \text{ for } 1 \leq k \leq K.
\]

**Proof of Lemma A1.** By using the change-of-variables xxx, we may write
\[
H(y_1, y_2, \ldots, y_K) = \left( \sum_{k=1}^{K} y_k \right) \left( \int_{\bar{c}(y)}^{1} (s - \bar{c}(y)) dF_X(s) \right).
\]
Now, since we have assumed that \(F_X\) is uniformly continuous with density \(f_X\), taking the second partial derivative of \(H\) with respect to \(y_k\), we obtain
\[
\frac{\partial^2}{\partial y_k^2} H(y_1, y_2, \ldots, y_K) = - \int_{\bar{c}(y)}^{1} \left( \frac{\partial}{\partial y_k} \bar{c}(y) + \sum_{k=1}^{K} \frac{\partial^2}{\partial y_k^2} \bar{c}(y) \right) dF_X(s)
\]
\[
+ \left( \sum_{k=1}^{K} y_k \right) \left( \frac{\partial}{\partial y_k} \bar{c}(y) \right)^2 f_X(\bar{c}(y)).
\]
However, using the definition of \(\bar{c}(y)\) from (19), it is straightforward to verify that
\[
2 \frac{\partial}{\partial y_k} \bar{c}(y) + \sum_{k=1}^{K} \frac{\partial^2}{\partial y_k^2} \bar{c}(y) = 0.
\]
As a result, we obtain
\[
\frac{\partial^2}{\partial y_k^2} H(y_1, y_2, \ldots, y_K) = \left( \sum_{k=1}^{K} y_k \right) \left( \frac{\partial}{\partial y_k} \bar{c}(y) \right)^2 f_X(\bar{c}(y)) \geq 0.
\]
\(\square\)
We now provide the proof for Proposition 3. We first prove that there exists an optimal solution that is a corner point in the set of feasible solutions of (20). The proof is by contradiction. Suppose that no such optimal solution exists. Let \( \beta : \prod_{k=1}^{K} [0, \alpha_k] \to \mathbb{N} \) be defined by \( \beta(y) = |\{k : y_k \in \{0, \alpha_k\}\}| \). In words, \( \beta(y) \) represents the number of \( y_k \)'s that are equal to either 0 or \( \alpha_k \). We refer to \( \beta(y) \) as the *binding number* of \( y \). Since the objective function of (20) is bounded and continuous, and the set of feasible solutions is compact, there exists at least one global optimal solution. Let \( y^* \) be the global optimal solution with the highest binding number. Since by assumption \( y^* \) is not a corner point, we have \( \beta(y^*) < K \). Let \( k \) be such that \( 0 < y_k^* < \alpha_k \) and define two solutions \( y^* \) and \( \bar{y}^* \) by

\[
\begin{align*}
  y^* &= (y_1^*, y_2^*, \ldots, y_{k-1}^*, 0, y_{k+1}^*, \ldots, y_K^*), \\
  \bar{y}^* &= (y_1^*, y_2^*, \ldots, y_{k-1}^*, \alpha_k, y_{k+1}^*, \ldots, y_K^*).
\end{align*}
\]

Note that \( y^* = (1 - y_k^*/\alpha_k)y_k^* + (y_k^*/\alpha_k)\bar{y}^* \). Due to the result of Lemma A1, we have

\[
\frac{\partial^2}{\partial y_k^2} H(y_1^*, y_2^*, \ldots, y_{k-1}^*, y_k, y_{k+1}^*, \ldots, y_K^*) \geq 0.
\]

Hence, \( H(y^*) \leq (1 - y_k^*/\alpha_k)H(y^*) + (y_k^*/\alpha_k)H(\bar{y}^*) \), which implies that at least one of \( y^* \) and \( \bar{y}^* \) is also a global optimum solution of (20). However, the binding number of both \( y^* \) and \( \bar{y}^* \) is equal to \( \beta(y^*) + 1 \). This contradicts with the choice of \( y^* \) as a global optimum solution with the highest binding number. Therefore, \( y^* \) is a corner point.

Now we prove that there exists a global optimum solution with the nested-in-cost structure. The proof is again by contradiction. Suppose that no such solution exists. For every corner point solution \( y = (y_1, y_2, \ldots, y_K) \in \prod_{k=1}^{K} \{0, \alpha_k\} \) that does not satisfy the nested property, we define the *irregularity index* of \( y \) by \( \gamma(y) = \min\{k : y_k = 0\} \). Since \( y \) does not satisfy the nested property, then there exists \( l > \gamma(y) \) such that \( y_l = \alpha_l \). Consider a corner point global optimum solution \( y^* \) with the highest irregularity index and let \( k = \gamma(y^*) \). We then distinguish between two different cases and in each case prove a contradiction.

**Case I** \( (\alpha_k < \sum_{k=k+1}^{K} y_k^*) \) Construct a new solution \( \tilde{y} \) such that

i. \( \tilde{y}_k = y_k^* = \alpha_k \), for \( 1 \leq k < \hat{k} \),

ii. \( \tilde{y}_k = \alpha_k \),

iii. \( 0 \leq \tilde{y}_k \leq y_k^* \), for \( \hat{k} < k \leq K \), and

iv. \( \sum_{k=\hat{k}+1}^{K} \tilde{y}_k = (\sum_{k=\hat{k}+1}^{K} y_k^*) - \alpha_k \).

Due to the construction of \( \tilde{y} \), we have \( \sum_{k=1}^{K} \tilde{y}_k = \sum_{k=1}^{K} y_k^* \). Moreover, since the \( \alpha_k \)'s are increasing in \( k \), we have that \( c(\tilde{y}) \leq c(y^*) \). Hence, by using (A14) we obtain that \( H(\tilde{y}_k) \geq H(y_k^*) \). Therefore, \( \tilde{y} \) is also a global optimal solution of (20).

We use the convexity result of Lemma A1 similar to the above to round all non-binding \( y_k \)'s for \( k > \hat{k} \) to either 0 or \( \alpha_k \), and obtain another global optimum solution \( \hat{y}^* \) that is a corner point. However, the first \( \hat{k} \) entries of \( \hat{y}^* \) are non-zero, and so \( \gamma(\hat{y}^*) \geq \hat{k} + 1 \). This contradicts the assumption that \( y^* \) is a global optimal solution with the highest irregularity index.

**Case II** \( (\alpha_k \geq \sum_{k=k+1}^{K} y_k^*) \) Consider a solution \( \check{y} \) such that
i. \( \hat{y}_k = y_\alpha^* = \alpha_k \), for \( 1 \leq k < \hat{k} \),

ii. \( \hat{y}_k = \sum_{k=1}^{K} y'_k \), and

iii. \( \hat{y}_k = 0 \), for \( \hat{k} < k \leq K \).

Similar to the first case, we have \( \sum_{k=1}^{K} \hat{y}_k = \sum_{k=1}^{K} y'_k \) and \( \bar{c}(\hat{y}) \leq \bar{c}(y^*) \). We use the convexity result of Lemma A1 to round \( \hat{y}_k \) to either 0 or \( \alpha_k \) and obtain a new global optimum solution \( \hat{y}^* \). However, \( \hat{y}^* \) is nested by its construction, which is a contradiction.

Therefore, both cases lead to a contradiction and hence there exists a global optimal solution to (20) that is a corner point and has a nested structure. \( \square \)

Proof of Theorem 5. It is sufficient by (24) to prove the limit (25). In order to do, we prove that the numerator and denominator on the lefthand side of (25) converge to the same limiting value. We begin with the denominator. Denote by \( \bar{V}_N(t, b_N) = \bar{V}_N(t) \wedge b_N \) the normalized restricted utility. By Theorem 2, \( \bar{V}_N(t) \Rightarrow \bar{V}(t) \) as \( N \to \infty \) and since \( \bar{V}(t) \) is a continuous non-decreasing function the convergence holds in \( D([0,1], \mathbb{R}) \). Hence, by the continuous mapping theorem, \( \bar{V}_N(\cdot, b_N) \Rightarrow \bar{V} \wedge b \) as \( N \to \infty \) and so, by a second application of the the continuous mapping theorem and uniform integrability,

\[
E \left[ \sup_{0 \leq t \leq 1} (\bar{V}_N(t, b_N) - ct) \right] \Rightarrow \sup_{0 \leq t \leq 1} (\bar{V}(t) \wedge b - ct) \quad \text{as} \quad N \to \infty.
\]

Regarding the term on the right-hand side above, using the concavity of \( \bar{V} \) it may be shown by solving for the first order condition that \( \bar{V}(t) \wedge b - ct \) is maximized at \( t^*_g = \min(t_b, t^*) \) where \( t^* \) is as given in the theorem and \( t_b = \inf \{ t \geq 0 : \bar{V}(t) > b \} \).

We next show that the numerator on the left-hand side of (25) converges to the same value. Denote by \( \bar{\zeta}(X, p_N, b_N) = \frac{1}{N} \zeta(X, p_N, b_N) \) the normalized bundle size choice of the customer that maximizes her surplus under the pricing policy \( p_N \) given in the Theorem. To prove the result, it suffices to show that \( \bar{\zeta}(X, p_N, b_N) \Rightarrow t^* \wedge t_b \) as \( N \to \infty \). Suppose not. Then, since the sequence \( \{ \bar{\zeta}(X, p_N, b_N), N \geq 1 \} \) is bounded and hence relatively compact, there must exist a subsequence \( \{ N_k, k \geq 1 \} \) along which \( \bar{\zeta}(X, p_{N_k}, b_{N_k}) \Rightarrow \hat{t} \) as \( k \to \infty \), where \( \hat{t} \neq t^* \wedge t_b \). Note that it has to be the case that \( \hat{t} < t^* \wedge t_b \). To complete the proof, we will show that the consumer’s surplus is lower if \( \bar{\zeta}(X, p_{N_{\hat{t}}} b_{N_{\hat{t}}}) \Rightarrow \hat{t} \) as \( k \to \infty \) instead of \( \bar{\zeta}(X, p_{N_k}, b_{N_k}) = t^* \wedge t_b \) for \( k \geq 1 \) which is a contradiction to the assumption that consumers are surplus maximizers.

We proceed as follows. By comparing the consumer surplus under both choices we have,

\[
N_k (\bar{V}_{N_k}(\bar{\zeta}(X, p_{N_k}, b_{N_k}))) - \bar{V}(\bar{\zeta}(X, p_{N_k}, b_{N_k}))) + h(\bar{\zeta}(X, p_{N_k}, b_{N_k}))g(N_k) - N_k (\bar{V}_{N_k}(t^* \wedge t_b) - \bar{V}(t^* \wedge t_b)) - h(t^* \wedge t_b)g(N_k).\]

Dividing by \( g(N) \) and rearranging, we get

\[
\frac{\sqrt{N_k} (\bar{V}_{N_k}(\bar{\zeta}(X, p_{N_k}, b_{N_k}))) - \bar{V}(\bar{\zeta}(X, p_{N_k}, b_{N_k})))}{g(N_k) / \sqrt{N_k}} - \frac{\sqrt{N_k} (\bar{V}_{N_k}(t^* \wedge t_b) - \bar{V}(t^* \wedge t_b))}{g(N_k) / \sqrt{N_k}} + h(\bar{\zeta}(X, p_{N_k}, b_{N_k})) - h(t^* \wedge t_b).
\]

(A15)

However, it follows from Stigler (1974) that \( \sqrt{N_k} (\bar{V}_{N_k} - \bar{V}) \) converges in distribution to a standard normal distribution as \( k \to \infty \). But since \( g(N_k) / \sqrt{N_k} \to \infty \) as \( k \to \infty \) it follows that the first two terms in (A15) \( \Rightarrow 0 \). Therefore, the difference in surplus \( \Rightarrow h(\hat{t}) - h(t^* \wedge t_b) < 0 \) as \( k \to \infty \) where the inequality holds since \( h(\cdot) \) is a strictly increasing function which completes the proof. \( \square \)
Proof of Theorem 6. Before presenting the proofs of the theorem, we introduce some additional convergence rate definitions. Let \( f(N) : N \to \mathbb{R} \). Then, we define \( O(h(N)) = \{ f(N) : \limsup_{N \to \infty} \frac{|f(N)|}{\Omega(N)} < +\infty \} \), \( \Omega_+(h(N)) = \{ f(N) : \liminf_{N \to \infty} \frac{|f(N)|}{\Omega(N)} > 0 \} \). Then, (h(N)) = O(h(N)) \cap \Omega_+(h(N)). We now state the following important result.

**Lemma A2.** For \( c = 0 \), we have that \( g^*(N) \in \omega_+(N^{1/2}) \cap o(N^\beta) \) for any \( \beta > 1/2 \).

**Proof:** Suppose first that \( g^*(N) \notin \omega_+(\sqrt{N}) \). Then, by the definition of \( \omega_+(\sqrt{N}) \), there must exist a subsequence \( \{N_k\} \) such that \( g^*(N_k)/\sqrt{N_k} \to c < \infty \). However, in this case

\[ \lim_{N_k \to \infty} \frac{\mathbb{E}[\pi(p)]}{N_k} = \lim_{n \to \infty} \left( 1 - \frac{g^*(N_k)}{N_k} \right) \mathbb{P}(V_{N_k}(N_k) - N_k \mu < -g^*(N_k)) \]

\[ = \lim_{N_k \to \infty} \mathbb{P} \left( \frac{V_{N_k}(N_k) - N_k \mu}{\sigma \sqrt{N_k}} > -\frac{g^*(N_k)}{\sigma \sqrt{N_k}} \right), \]

where here we have used the fact that since \( g^*(N) \notin \omega_+(\sqrt{N}) \),

\[ \lim_{N \to \infty} \left( 1 - \frac{g^*(N)}{N} \right) = 1. \]

On the other hand, by the central limit theorem we have that \( (V_{N_k}(N_k) - N_k \mu)/(\sigma \sqrt{N_k}) \to \mathcal{N}(0,1) \), as \( N \to \infty \). Hence, since \( g^*(N_k)/\sqrt{N_k} \to c < \infty \) it follows that

\[ \lim_{N_k \to \infty} \mathbb{P} \left( \frac{V_{N_k}(N_k) - N_k \mu}{\sigma \sqrt{N_k}} > -\frac{g^*(N_k)}{\sigma \sqrt{N_k}} \right) = \mathbb{P}(N(0,1) > -c) < 1, \]

from which we obtain that

\[ \limsup_{N_k \to \infty} \frac{\mathbb{E}[\pi(p)]}{N_k} < 1, \]

which is a contradiction. Thus, \( g^*(N) \in \omega_+(\sqrt{N}) \).

Next, suppose that \( g^* \notin o(N^\beta) \) for some \( \beta > 1/2 \). If \( g^* \notin o(N) \), then \( g^* \) cannot be optimal since

\[ \lim_{N_k \to \infty} \left( 1 - \frac{g^*(N_k)}{N_k} \right) \neq 1. \]

On the other hand, suppose that \( g^* \notin o(N^\beta) \) for some \( \beta > 1/2 \) but that \( g^* \in o(N) \). Then, consider \( g^\beta_\sigma(N) = g^*(N)/N^\varepsilon \) for some \( 0 < \varepsilon < \beta - 1/2 \) and let \( p^\beta_\sigma \) be its associate price. Then, since \( g^\beta_\sigma < g^* \), it follows after some algebra that

\[ \mathbb{E}[\pi(p^\beta_\sigma)] - \mathbb{E}[\pi(p^*)] = (g^*(N) - g^\beta_\sigma(N)) \mathbb{P} \left( \frac{V_{N_k}(N_k) - N_k \mu}{\sigma \sqrt{N_k}} > -\frac{g^*(N_k)}{\sigma \sqrt{N_k}} \right) \]

\[ + g^\beta_\sigma(N) \mathbb{P} \left( \frac{g^*(N_k)}{\sigma \sqrt{N_k}} < \frac{V_{N_k}(N_k) - N_k \mu}{\sigma \sqrt{N_k}} < -\frac{g^*(N_k)}{\sigma \sqrt{N_k}} \right), \]

Now, regarding the second term, note that

\[ \mathbb{P} \left( \frac{g^*(N_k)}{\sigma \sqrt{N_k}} < \frac{V_{N_k}(N_k) - N_k \mu}{\sigma \sqrt{N_k}} < -\frac{g^*(N_k)}{\sigma \sqrt{N_k}} \right) < \mathbb{P}(V_{N_k}(N_k) - N_k \mu < -g^*(N_k)) \]

\[ = \exp \left( -\frac{(g^\beta_\sigma(N_k))^2}{2k \sigma^2} + o \left( \frac{(g^\beta_\sigma(N_k))^2}{k} \right) \right), \]

where the final equality follows by the moderate deviations principle since \( g^\beta_\sigma \in \omega_+(N_k^{1/2}) \cap o(N_k) \). Hence,

\[ g^\beta_\sigma(N) \mathbb{P} \left( \frac{g^*(N_k)}{\sigma \sqrt{N_k}} < \frac{V_{N_k}(N_k) - N_k \mu}{\sigma \sqrt{N_k}} < -\frac{g^*(N_k)}{\sigma \sqrt{N_k}} \right) \to 0 \quad \text{as} \quad N_k \to \infty. \]
Next, regarding the first time we have that since $g^*_k \in \omega_+ (N_k^{1/2})$ by the central limit theorem
\[
P \left( \frac{V_{N_k} (N_k) - N_k \mu}{\sigma \sqrt{N_k}} > - \frac{g^*(N_k)}{\sigma \sqrt{N_k}} \right) \to 1 \text{ as } N_k \to \infty.
\]
Moreover, $(g^*(N) - g^*_k (N)) = 0$ 

Putting the above together it follows that $\mathbb{E}[\pi(p^*_k)] - \mathbb{E}[\pi(p^*)] \to \infty$ as $N_k \to \infty$, which is a contradiction. Thus, $g^*(N) \in o(N^\beta)$ for every $\beta > 1/2$.

We are now ready to present the proof of Theorem 6.

In order to establish (27), we show that pricing at $p(N) = N\mu - \sigma \sqrt{N} \log N$ leads to the following limiting ratio
\[
\lim_{N \to \infty} \frac{N\mu - \mathbb{E}[\pi(p)]}{\sigma \sqrt{N} \log N} = 1.
\]
This implies that the ratio under price $p^*$ is upper bounded by 1. However, we show there is no price that can achieve a lower ratio than 1.

Notice that any $g(N) \in \omega_+(\sqrt{N}) \cap o(N^\beta)$ for all $\beta > \frac{1}{2}$ can be written as $g(N) = h(N) \sigma \sqrt{N} \log N$ where $h(N) \in \omega_+ (\frac{1}{\log N}) \cap o(\frac{N^\epsilon}{\log N})$ for any $\epsilon > 0$. Therefore, for any $p(N) = N\mu - h(N) \sigma \sqrt{N} \log N$, we have after some algebra that
\[
\mathbb{E}[\pi(p)] = N\mu - h(N) \sigma \sqrt{N} \log N
- R(g(N)) \left[ N\mu - h(N) \sigma \sqrt{N} \log N \right] \Phi(-h(N) \sqrt{\log N}),
\]
where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution and
\[
R(g(N)) = \frac{P \left( \frac{V_N (N) - N\mu}{\sigma \sqrt{N}} \leq - \frac{g(N)}{\sigma \sqrt{N}} \right)}{\Phi \left( - \frac{g(N)}{\sigma \sqrt{N}} \right)}.
\]

We now state a known result on the asymptotic expansion of the tail of a standard normal distribution.

**Lemma A3 (Asymptotic Expansion of the Tail of a Normal Distribution).**

\[
P \left( N(0, 1) \geq x \right) = \frac{1}{\sqrt{2\pi}} \frac{\exp(-\frac{x^2}{2})}{x} \sum_{l=0}^{k-1} (-1)^l (2l)! \frac{1}{x^{2l+1}} + (-1)^k D_k,
\]
where $D_k \in \Theta_+ (\frac{\exp(-\frac{x^2}{2})}{x})$ as $x \to \infty$.

Using the asymptotic expansion in Lemma A3 for $k = 1$, we get
\[
\Phi(-h(N) \sqrt{\log N}) = \frac{1}{\sqrt{2\pi h(N) \log N}} \exp \left( - \frac{(h(N))^2 \log N}{2} \right) - \Theta_+ \left( \frac{\exp(-h(N) \sqrt{\log N})}{(h(N) \sqrt{\log N})^2} \right)
= \frac{N^{-\frac{(h(N))^2}{2}}}{\sqrt{2\pi h(N) \log N}} (1 + o(1)).
\]

Substituting (A22) in (A21) and after some algebra, we get that
\[
\frac{N\mu - \mathbb{E}[\pi(p)]}{\sigma \sqrt{N} \log N} = h(N) + \frac{R(g(N)) \mu}{\sqrt{2\pi} \sigma} \sqrt{\frac{N}{\log N}} \frac{N^{-\frac{(h(N))^2}{2}}}{h(N) \sqrt{\log N}} (1 + o(1))
= h(N) + Q(g(N)) (1 + o(1)),
\]
where we define $Q(g(N)) = \frac{R(g(N))\mu}{\sqrt{2\pi}\sigma} \sqrt{\frac{N}{\log N} \frac{(h(N))^2}{h(N)\sqrt{\log N}}}$. Notice that we have

$$\log(Q(g(N))) = \log \left( \frac{R(g(N))\mu}{\sqrt{2\pi}\sigma} \right) + \frac{1}{2} \log N - \frac{1}{2} \log \log N - \frac{(h(N))^2}{2} \log N - \log h(N) - \frac{1}{2} \log \log N + o(1).$$  \hspace{1cm} (A23)

We now state Cramer’s moderate deviation principle.

**Lemma A4.** Given Assumption 5, then for any $g(N) \in \omega_+(\sqrt{N}) \cap o(N^{2/3})$, we have $R(g(N)) \rightarrow 1$ as $N \rightarrow \infty$.

This result dates back to Cramér (1938) (see Petrov (1975), Chapter VIII, Theorem 1). From Lemma A2, we have that $g^*(N)$ belongs to Cramer’s moderate deviation regime and clearly so does $g(N) = \sigma \sqrt{N \log N}$.

It now follows that for $p(N) = N \mu - \sigma \sqrt{N \log N}$ we get that $\lim_{N \rightarrow \infty} \log(Q(\sigma \sqrt{N \log N})) = -\infty$. This in return implies that $\lim_{n \rightarrow \infty} Q(\sigma \sqrt{N \log N}) = 0$ and

$$\lim_{N \rightarrow \infty} \frac{N \mu - E[\pi(p^*)]}{\sigma \sqrt{N \log N}} = 1.$$

Let $h^*(N) = g^*(N)/(\sigma \sqrt{N \log N})$. In order to establish (27), we show by contradiction that if $\lim_{n \rightarrow \infty} h^*(N) \neq 1$ then the ratio

$$\lim_{n \rightarrow \infty} \frac{N \mu - E[\pi(p^*)]}{\sigma \sqrt{N \log N}} > 1.$$

First, assume that $\tilde{h} = \lim \inf_{N \rightarrow \infty} h^*(N) > 1$. It follows from (A23) and Lemma A4, that $\limsup_{N \rightarrow \infty} \log Q(g^*(N)) = -\infty$ which implies that $\limsup_{N \rightarrow \infty} Q(g^*(N)) = 0$. Therefore, we get that

$$\limsup_{N \rightarrow \infty} \frac{N \mu - E[\pi(p^*)]}{\sigma \sqrt{N \log N}} = \tilde{h} > 1.$$

However, this contradicts the optimality of $p^*$ since pricing at $p(N) = N \mu - \sigma \sqrt{N \log N}$ leads to higher profits.

On the other hand, if $\limsup_{N \rightarrow \infty} h^*(N) < 1$ then from (A23) and Lemma A4, we get that $\liminf_{N \rightarrow \infty} \log(Q(g^*(N))) = +\infty$ which implies that $\liminf_{N \rightarrow \infty} Q(g^*(N)) = +\infty$. Therefore, we get that

$$\liminf_{N \rightarrow \infty} \frac{N \mu - E[\pi(p^*)]}{\sigma \sqrt{N \log N}} = +\infty.$$

Again, this contradicts the optimality of $p^*$ since setting $p(N) = N \mu - \sigma \sqrt{N \log N}$ leads to higher profits which establishes both (26) and (27). \hspace{1cm} \Box

**Numerical Simulations of All Eight Pricing Policies**

In Figure A1, we plot the ratio of profits of the 8 different pricing policies relative the profit obtained from non-linear optimization as in Chu et al. (2011).
Figure A1 The ratio of profits of a BSP with one size relative to N sizes as in Chu et al. (2011) for all 8 pricing policies.