Large-scale Bundle Size Pricing: A Theoretical Analysis

Tarek Abdallah, Arash Asadpour, Josh Reed

Last Update January 25, 2017

Department of Information, Operations, and Management Sciences
Leonard N. Stern School of Business
New York University
New York, NY 10012

Bundle size pricing (BSP) is a multi-dimensional 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 multi-product 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 multi-dimensional problem to a simple multi-unit pricing problem with concave and increasing utilities. Our framework allows us to identify the main source of inefficiency of BSP that is the heterogeneity of marginal costs across products. For this reason, we propose two new BSP policies called “clustered BSP” and “assorted BSP” that significantly reduce the inefficiency of regular BSP. We then utilize our framework to study richer models of BSP such as when customers have budgets and when there exists multiple customer types.

1. Introduction

Bundling is a pervasive selling practice across various markets. Many multi-product 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 check-in bags, access to the on-board entertainment system, and on-board meals among others. However, in recent years, there has been a strong trend towards the unbundling of the cable TV and airline tickets (Popper 2015, Owram 2014). This unbundling trend is usually justified by customers’ needs for the flexibility
to pay for what they want rather than being tied to pre-designed bundles (Flint 2015, Bhaskara 2015). In fact, several cable TV providers such as Verizon Fios and Dish Network have recently introduced a new service called custom TV that allows a certain degree of flexibility for customers to design their personalized bundle of TV channels (Welch 2015, Morran 2016).

Meanwhile, some industries which have traditionally sold products separately are moving towards a subscription model which bundles access to a large number of products or services. For example, in 2013, Adobe announced a major shift in its business model where it would no longer sell its products separately as standalone software. Instead, it now offers a subscription-only service where customers gain access to a large portfolio of its software in return for a monthly subscription fee (Shankland 2013). This move proved very successful for Adobe as it has enjoyed significant increases in its revenues over the past years. In fact, in the third quarter of 2016, Adobe reported a record revenue which was mainly boosted by its subscription-only model (Minayai 2016). Yet, the shift to this form of bundling has upset some of Adobe’s customers to an extent that some of them signed an online petition urging Adobe to abandon its subscription-only model (Schoffstall 2013).

Based on the above anecdotal evidence, it seems that there is a trade-off between the classical bundling practices and customers’ need for the flexibility to choose what they want to pay for. However, proponents of classical bundling practices argue that it simplifies the selling process. In fact, there are several news articles that discuss how customers are becoming increasingly agitated by the recent unbundling practices in the airlines industry (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 multi-product 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). Another strategy is to restrict selling all the products to a single comprehensive bundle without allowing 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 2015). In addition, pure bundling does not allow any flexibility for customers to customize their bundle.

1 Adobe now also offers subscription service for individual software.
However, bundling need not be synonymous with lack of customization. In fact, a multi-product 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 nests both pure bundling and component pricing as special cases. It also allows full flexibility for customers 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 decide what they want to pay for.

One potential candidate is the bundle size pricing (BSP) strategy. In the BSP problem, a multi-product 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 his 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.

Despite its common use in practice, there is limited literature on the theoretical properties of BSP. The BSP problem was first studied under the name “customized bundling” by Hitt and Chen (2005) who performed comparative statics analysis to understand some properties of the BSP problem. Later, Chu et al. (2011) ran extensive numerical simulations for a setting with up to 5 products to understand the performance of the optimal BSP policy. According to their numerical analysis, Chu et al. (2011) report that, on average, the optimal BSP policy achieved 98% of the profit of mixed bundling. But they notice that the efficiency of BSP decreases with an increase in marginal costs to an extent that BSP can be less profitable than component pricing for large marginal costs. However, Chu et al. (2011) note the limitations of making a general conclusion based solely on numerical simulations.

In this paper, we present a theoretical framework to analyze and characterize the optimal BSP policy for a multi-product firm that sells a large number of products. In this regard, our contribution is two-fold. 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 the showing that the BSP problem, while being a hard multi-dimensional problem, reduces in the limit to a simple multi-unit monopolistic pricing problem with a non-decreasing concave utility à la Mussa and Rosen (1978) that is easy to solve. We further provide a complete characterization of the limiting valuation curve based on the model primitives along with characterizing the convergence rate of the profits. This simple framework allows us to study richer models and extensions of the BSP problem and provides a deeper understanding of the performance of BSP relative to mixed bundling.

Our second contribution is utilizing our developed framework to understand the theoretical properties of the optimal BSP policy under various settings including ones that have not yet been explored in the bundling literature. We first study the BSP problem under non-negative marginal costs. In the case of equal marginal costs, we confirm the main insight of Chu et al. (2011) regarding the ability of BSP to closely approximate the profits under mixed bundling. In particular, for equal marginal costs and as the number of products grows large, the optimal BSP policy can asymptotically achieve the profit under mixed bundling. In fact, both BSP and mixed bundling asymptotically achieve perfect price discrimination (i.e. first best). We point out here that the above property holds for any level of equal marginal costs. That is, unlike the insights from Chu et al. (2011), in the case of equal marginal costs, an increase or decrease in the marginal cost level has no effect on the asymptotic performance of BSP relative to mixed bundling.

Meanwhile, we uncover the main source of inefficiency for BSP which is 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). In fact, the asymptotic ratio of the BSP profits relative to mixed bundling can be much lower than 98% as reported by Chu et al. (2011) depending on the level of heterogeneity in the marginal costs. The intuition behind this source of inefficiency is very simple and relates to the underlying structure of BSP where prices are set for the sizes of the bundle regardless of which items are purchased. Hence, due to this limitation, a firm that adopts a BSP policy cannot properly discriminate between customers who choose items with high marginal costs from those who choose items with low marginal costs. This creates major inefficiencies for BSP relative to mixed bundling. However, using our framework, we propose two new BSP policies, “clustered BSP” and “assorted BSP”, that can significantly reduce the inefficiency of the (regular) BSP policy when the products have heterogeneous marginal costs. We further extend our analysis to settings where customers have heterogeneous budgets that are unknown to the firm and where customers draw their valuations from different distributions.
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 less flexibility to the customers. 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. However, when customers have heterogeneous budgets that are unknown to the firm, we show that the asymptotically optimal BSP policy is no longer to offer a single size. Instead, the firm should offer a pricing curve for a range of different bundle sizes. In the presence of budgets but items with equal marginal costs, the firm can still asymptotically achieve perfect price discrimination. Meanwhile, when there are two customer types that draw their valuations from different underlying distributions, we show that as long as it is profitable to discriminate between the two types, the asymptotically optimal BSP policy is to offer the low valuation customers with a low bundle size for which it extracts all their surplus. At the same time, the firm also offers the high valuation customers an efficient bundle size for which they retain some positive surplus due to the “information rent”. In this case, the optimal BSP policy cannot asymptotically achieve perfect price discrimination, yet we show that it can asymptotically achieve second-degree price discrimination (i.e. second best) as long as the items have equal marginal costs.

2. Related Literature

The literature on bundling dates back to a short note by Stigler (1963). The first serious attempt to analyze the mixed bundling problem is due to 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).

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, dubbed as “information goods”. They showed that, in the case of independent and additive item valuations, a seller of a

---

2 In the case when it is not profitable to discriminate between the two types, the optimal BSP policy is either to price out the low valuation customers and sell an efficient size for the high valuation customers whereby extracting all their surplus or the optimal BSP policy is to bunch them as a single type and sell the efficient size of the low valuation customers that extracts all the low type’s surplus but retains positive surplus for the high types.
large number of information goods items 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 sub-additive 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 counter-example that the efficiency of a pure bundle in extracting a large consumer surplus is highly compromised. Abdallah (2015) revisited the large-scale pure bundling problem but with non-negative marginal costs. He characterized the asymptotically optimal pricing policy and provided a simple and easy to compute tight lower bound on the potential loss from pure bundling. For the special case of information goods, Abdallah (2015) 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 sub-additive and super-additive valuations under a general utility model and characterized regimes under which the critique by Geng et al. (2005) holds.

In parallel to the bundling literature, there is an extensive literature on non-linear pricing which focuses on the role of multi-part 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 non-negative marginal costs, a cost-based two-part tariff where products are priced at their respective marginal cost 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. However, despite being a simple bundling mechanism, the
literature on BSP 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 do 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. Recently, Mirrokni and Nazerzadeh (2015) analyzed a new form of selling contracts in the ads-exchange industry which they call “preferred deals”. Essentially, the preferred deals contract is a BSP mechanism in which advertisers are offered deals to purchase a fraction of the ad impressions sent by the firm for a fixed price.

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 (2015).

3. Bundle Size Pricing with Zero Marginal Costs

In order to present the notation and layout some general concepts, we start with describing the most basic bundle size pricing (BSP) problem where the items have zero marginal costs and the customers have i.i.d. valuations. We consider a setting in which a monopolistic risk-neutral firm is selling \( N \geq 1 \) differentiated indivisible items to a market of \( M \geq 1 \) customers. In this basic model, we assume that each of the \( n = 1, \ldots, N \), items has a zero marginal cost for the firm to obtain or produce. Customers have heterogeneous non-negative valuations for the \( N \) different items. We denote by \( X_{m,n} \) the value that customer \( m = 1, \ldots, M \), places on item \( n = 1, \ldots, N \). We assume that the collection of the non-negative random variables \( \{X_{m,n}; m = 1, \ldots, M; n = 1, \ldots, N\} \) is i.i.d. with common distribution \( F_X \) that has a finite mean \( \mu \) and variance \( \sigma^2 \). Consistent with the bundling literature and throughout this paper, we assume that the customers have unit demand for each item and that their valuations for any bundle is additive, i.e. a customer’s valuation for any subset of items is simply the sum of his valuations for the items.

In the BSP problem, the firm sets the prices based on the size of the bundle regardless of which items are chosen by the customers. In other words, the firm offers a menu of \( N \) different bundle sizes that range from \( n = 1 \) through \( N \), along with their corresponding price vector \( (p(1), \ldots, p(N)) \in (\mathbb{R}_+ \cup \infty)^N \). We allow the possibility that \( p(n) = \infty \) for some \( n \in \{1, \ldots, N\} \). Since the customers’ valuations are always finite, this corresponds to the case in which the firm decides not to offer a bundle of size \( n \).
For each bundle size \( n = 1, \ldots, N \), the customers are allowed to customize their bundle by choosing any \( n \) out of the \( N \) offered items. The value which customer \( m = 1, \ldots, M \), places on a bundle of size \( n \) is given by the optimal value of the integer program

\[
\max \sum_{r=1}^{N} I_{m,r} X_{m,r}
\]

subject to \( \sum_{r=1}^{N} I_{m,r} = n \)

\( I_{m,r} \in \{0,1\} \) for \( r = 1, \ldots, N \).

This is a 0-1 knapsack problem and its solution can be described in terms of the order statistics of the vector of valuations \( X_m = (X_{m,1}, \ldots, X_{m,N}) \). Specifically, for each \( n = 1, \ldots, N \), we denote by \( X_{m,(n)} \) the \( n \)th order statistic of \( (X_{m,1}, \ldots, X_{m,N}) \). That is, \( X_{m,(n)} \) is the \( n \)th smallest value in the vector \( (X_{m,1}, \ldots, X_{m,N}) \). Then we have that \( X_{m,(1)} \leq X_{m,(2)} \leq \ldots \leq X_{m,(N)} \) and the optimal value of the previous integer program is given by

\[
V_{m,N}(n) = \sum_{k=0}^{n-1} X_{m,(N-k)}.
\] (1)

The vector \( (V_{m,N}(1), \ldots, V_{m,N}(N)) \) now represents the values which customer \( m \) places on bundles of size 1 through \( N \). We denote the no-purchase option as bundle size 0 for which we set \( V_m(0) = p(0) = 0 \) for \( m = 1, \ldots, M \). Hence, the customers always retain a non-negative surplus. Given a vector of bundle size prices \( p = (p(0), p(1), \ldots, p(N)) \), each customer will purchase the bundle of an appropriate size so as to maximize his particular surplus. That is, customer \( m \) purchases a bundle of size

\[
\zeta(X_m, p) \in \arg\max_{n=0,\ldots,N} (V_{m,N}(n) - p(n)),
\]

and the firm’s realized profit (assuming zero marginal costs) in the BSP problem is given by the random variable

\[
\pi(p) = \sum_{m=1}^{M} \sum_{n=1}^{N} p(n)1\{\zeta(X_m, p) = n\}.
\] (2)

Recall next that the firm is risk-neutral and hence is interested in maximizing its expected profit. Due to the i.i.d. assumptions on the customer valuations, the firm’s expected profit can be written as

\[
\mathbb{E}[\pi(p)] = M \cdot \sum_{n=1}^{N} p(n) \mathbb{P}(\zeta(X_1, p) = n),
\]
and hence the firm’s profit maximization problem reduces to solving the optimization problem

$$
\max_{p \in (\mathbb{R}_+ \cup \infty)^{N+1}} \sum_{n=1}^{N} p(n) P(\zeta(X_1, p) = n).
$$

In general, there may be multiple optimal solutions to the above problem and so we denote by $P^*$ its set of optimal solutions. Moreover, letting $p^* \in P^*$ we have that $\mathbb{E}[\pi(p^*)]$ is the maximum expected profit which the firm may obtain under a BSP policy with zero marginal costs. It is worth mentioning that by setting $p(n) = +\infty$ for $n = 1, ..., N - 1$ and allowing $p(N)$ to be a free variable, the BSP problem reduces to the pure bundling problem where the firm sells all the items as a single comprehensive bundle only. Hence, pure bundling is is a special form of BSP. However, in the pure bundling problem, customers do not have a choice over which items to place in their bundles.

Solving for $\mathbb{E}[\pi(p^*)]$ is in general hard since the optimization problem (3) is not necessarily concave and does not have 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.

We first provide an upper bound for $\pi(p)$. Notice that we have $\{\zeta(X_m, p) = n\} \subseteq \{V_m,N(n) > p(n)\}$ for $n = 1, ..., N$. Now it follows from (2) together with the fact that $V_m(1) \leq V_m(2) \leq ... \leq V_m(N)$, that we obtain the perfect price discrimination bound

$$
\pi(p) \leq \sum_{m=1}^{M} V_{m,N}(N) = \sum_{m=1}^{M} \sum_{n=1}^{N} X_{m,n}.
$$

Now taking expectations on both sides of the above, we obtain the simple bound

$$
\mathbb{E}[\pi(p^*)] \leq MN\mu,
$$

where $MN\mu$ is the expected profit under perfect price discrimination. In fact, the expected profit under perfect price discrimination is an upper bound on any selling strategy not only BSP. Nevertheless, in the present setting, Bakos and Brynjolfsson (1999) show that the ratio of the expected optimal pure bundling profit to the expected profit under perfect price discrimination goes to one when the number of items $N$ becomes large. Still they left open the question of how to either compute or approximate the optimal price $p^*$. This question was recently answered by Abdallah (2015) who, letting $\omega_+(N^{1/2}) = \{h(N) : \liminf_{N \to \infty} h(N)/N^{1/2} = +\infty\}$, showed that the asymptotically optimal pure bundle price is $N\mu - g(N)$ for some $g(N) \in \omega_+(N^{1/2}) \cap o(N)$.

Since pure bundling is a special case of the BSP policy, then setting the bundle size prices to

$$
\mathcal{P}(n) = \begin{cases} 
\infty & \text{if } n = 1, ..., N - 1, \\
N\mu - g(N) & \text{if } n = N,
\end{cases}
$$

will achieve the asymptotically optimal price. This approach is numerically very fast even for large number of items. Moreover, this will be our approach in the numerical simulations section.
where \( g(N) \in \omega_+(N^{1/2}) \cap o(N) \), we obtain from the bound in (4) above and the results by Abdallah (2015) that

\[
\lim_{N \to \infty} \frac{\mathbb{E}[\pi(P)]}{MN\mu} = \lim_{N \to \infty} \frac{\mathbb{E}[\pi(p^*)]}{MN\mu} = 1.
\]

Therefore, in the BSP problem with zero marginal costs, the firm asymptotically achieves the expected profit under perfect price discrimination by only selling a pure bundle. However, in more complex settings such as those considered next where there are positive marginal costs, budget constraints, or multiple customer types, pure bundling is not asymptotically optimal and perfect price discrimination cannot always be achieved. We therefore focus on providing asymptotically optimal BSP policies in such situations while at the same time providing a general theoretical framework which may be applied to even more complex settings.

4. Bundle Size Pricing with Positive Marginal Costs

We now study a more general bundle size pricing problem in which the firm incurs a positive marginal cost for each item selected by a customer in his purchased bundle. We study three different cases where in each case we add an extra layer of complexity. We start by considering the case of i.i.d. valuations where items have identical positive marginal costs. We then extend the analysis to the case of i.i.d valuations but heterogeneous marginal costs, and we end the section by analyzing the case of cost-dependent valuations along with heterogeneous marginal costs.

4.1. Identical Positive Marginal Costs

We first consider the setting in which each item has an identical positive marginal cost \( c > 0 \) associated with it. We assume that the marginal cost for an item is incurred by the firm only if the item is purchased by a customer. Hence, a purchased bundle of size \( n = 1, \ldots, N \), costs \( nc \) for the firm to sell. Using the same setup and notation as in Section 3, if the firm selects a price vector \( p = (p(0), \ldots, p(N)) \in (\mathbb{R}_+ \cup \infty)^{N+1} \), then its realized profit is given by the random variable

\[
\pi(p) = \sum_{m=1}^{M} \sum_{n=1}^{N} (p(n) - nc)\mathbb{1}\{\zeta(X_m, p) = n\}.
\]

(6)

The firm’s objective is to choose a price vector \( p \) in order to maximize its expected profit. Hence, taking expectations on both sides (6), we obtain that in order to maximize its expected profit, the firm must choose a price vector which solves the optimization problem

\[
\max_{p \in (\mathbb{R}_+ \cup \infty)^{N+1}} \mathbb{E}[\pi(p)] = M \cdot \max_{p \in (\mathbb{R}_+ \cup \infty)^N} \sum_{n=1}^{N} (p(n) - nc)\mathbb{P}(\zeta(X_1, p) = n),
\]

and we denote by \( P^* \) the set of optimal solutions to the above optimization problem.
We now proceed to obtain an upper bound on the firm’s expected profit. First, note that as in Section 3, we have that \( \{\zeta(X_m, p) = n\} \subseteq \{p(n) < V_{m,N}(n)\} \) for each \( m = 1, ..., M, \) and \( n = 1, ..., N. \) Hence, we immediately obtain from (6) that

\[
\pi(p) \leq \sum_{m=1}^{M} \sum_{n=1}^{N} (V_{m,N}(n) - nc) 1\{\zeta(X_m, p) = n\},
\]

from which it is straightforward to deduce that

\[
\pi(p) \leq \sum_{m=1}^{M} \sup_{n=0,1,...,N} (V_{m,N}(n) - nc). \tag{7}
\]

Moreover, it follows from the definition of \( V_{m,N}(n) \) in (1) and some algebra that for each \( m = 1, ..., M, \) we have the identity

\[
\sup_{n=0,1,...,N} (V_{m,N}(n) - nc) = \sum_{n=1}^{N} (X_{m,n} - c)^+ , \tag{8}
\]

where \((a)^+ = \max(0, a)\) for \( a \in \mathbb{R}. \) Now combining (7) and (8) we obtain the bound

\[
\pi(p) \leq \sum_{m=1}^{M} \sum_{n=1}^{N} (X_{m,n} - c)^+.
\]

The quantity on the right hand side above is the 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 bundle size pricing.

Now note that the right hand side of the bound above is independent of \( p \) and that taking expectations on both sides of it we obtain that

\[
\frac{\mathbb{E}[\pi(p) ]}{MN \mathbb{E}[(X_{1,1} - c)^+]} \leq 1 \quad \text{for} \quad p \in (\mathbb{R}_+ \cup \infty)^N. \tag{9}
\]

In fact, Armstrong (1999) has shown that in the current model setup if a firm considers the optimal two-part tariff policy, then the ratio of the optimal expected profit divided by \( MN \mathbb{E}[(X_{1,1} - c)^+] \) goes to 1 as the number of items \( N \) goes to infinity. In our main result of this section, we show similarly that if a firm considers the optimal bundle sizing pricing policy, then the same ratio converges to 1 as well and the firm only needs to sell one bundle size. We proceed as follows.

Let \( F_X^{-1} : (0,1) \mapsto \mathbb{R} \) denote the quantile function of \( F_X \) defined by \( F_X^{-1}(p) = \inf\{x \geq 0 : F_X(x) \geq p\} \) for \( p \in (0,1) \) (see Chapter 21 of Van der Vaart (2000)). Next, let \( \bar{V} = (\bar{V}(t), 0 \leq t \leq 1) \) be defined by

\[
\bar{V}(t) = \int_{0}^{t} F_X^{-1}(1-s)ds \quad \text{for} \quad 0 \leq t \leq 1, \tag{10}
\]
where \( \bar{V}(0) = 0 \). We note that \( \bar{V}(t) \) is well defined as long as \( \mu < \infty \). Also, it is straightforward to show that \( \bar{V}(t) \) is a continuous, non-decreasing, and concave function.

For each \( m = 1, \ldots, M \), denote by

\[
\bar{V}_{m,N}(t) = \frac{1}{N} \cdot V_{m,N}([Nt]) \quad \text{for} \quad 0 \leq t \leq 1,
\]

the normalized valuation of a customer \( m \) for a bundle of size \( [Nt] \). We then have the following result (all proofs are relegated to the Appendix).

**Theorem 1.** In the presence of identical marginal costs \( c > 0 \), assume that \( 1 - t^* = F^{-1}_X(c) \) is a continuity point of \( F^{-1}_X \). Then, setting

\[
\mathcal{P}(n) = \begin{cases} 
N \mathbb{E}[\bar{V}_{1,N}(t^*)] - g(N) & \text{if } n = [Nt^*], \\
+\infty & \text{if } n \neq [Nt^*],
\end{cases}
\]

where \( g(N) \in \omega_+((\sqrt{N}) \cap o(N) \), we have that

\[
\lim_{N \to \infty} \frac{\mathbb{E}[\pi(\mathcal{P})]}{M \mathbb{E}[(X_{1,1} - c)^+] = \lim_{N \to \infty} \frac{\mathbb{E}[\pi(p^*)]}{M \mathbb{E}[(X_{1,1} - c)^+]} = 1. \tag{11}
\]

We note that the expectation term \( \mathbb{E}[\bar{V}_{1,N}(t^*)] \) appearing in the asymptotically optimal pricing policy of Theorem 1 depends on the number of items \( N \geq 1 \) and may sometimes be difficult to compute. However, assuming that the condition

\[
\int_0^\infty (F_X(x)(1 - F_X(x)))^{1/2} dx < \infty \tag{12}
\]

holds, it follows by Theorem 4 of Stigler (1974) that one may replace \( \mathbb{E}[\bar{V}_{1,N}(t^*)] \) by the simpler value \( \bar{V}(t^*) \) in the definition of \( \mathcal{P}(n) \) in Theorem 1. The integrability condition (12) 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).

Theorem 1 states that in the BSP problem with identical marginal costs, the firm may asymptotically achieve perfect price discrimination by offering only a single bundle of size \( [Nt^*] \) at the price \( N \mathbb{E}[\bar{V}(t^*)] - g(N) \), where \( g(N) \in \omega_+((\sqrt{N}) \cap o(N) \). We note here that theorem holds for any \( g(N) \in \omega_+((\sqrt{N}) \cap o(N) \). In fact, in Section 7 we show that in the case of zero marginal costs the optimal \( g(N) = \sigma \sqrt{N \log N} + o(\sqrt{N \log N}) \).

The intuition behind Theorem 1 is as follows. First, let \( \bar{\pi}(p) = N^{-1}\pi(p) \). It then follows by (7) that we may write

\[
\bar{\pi}(p) \leq \sum_{m=1}^{M} \sup_{0 \leq t \leq 1} (\bar{V}_{m,N}(t) - ct). \tag{13}
\]
Next by Lemma A1 in the Appendix, we have that $\bar{V}_{m,N}(\cdot)$ converges uniformly almost surely to the deterministic function $\bar{V}(\cdot)$. Therefore, by the continuous mapping theorem (Billingsley 1999) and the bound (13), we get that

$$\limsup_{N \to \infty} \bar{\pi}(p) \leq M \sup_{0 \leq t \leq 1} (\bar{V}(t) - ct).$$

Moreover, since the valuation curve for all of the customers converge to the same deterministic valuation curve $\bar{V}(t)$, it is reasonable for the firm to restrict its offering to only a bundle of size $\lceil Nt^* \rceil$ where $t^* = \arg\max(\bar{V}(t) - ct)$. Furthermore, since $F_X^{-1}(1 - \cdot)$ is a non-increasing function whose range is the support of $F_X$, which is a subset of the positive reals, it follows from (10) that $\bar{V}(t) - ct$ is a concave function on $[0, 1]$. Hence, its optimizer may be obtained by solving for the first order condition $F_X^{-1}(1 - t^*) = c$ and, since $1 - t^*$ is a continuity point of $F_X^{-1}$, then we obtain that $t^* = 1 - F_X(c)$.

In Figure 1, we provide 4 illustrations of the limiting valuation curve $\bar{V}(t)$ relative to the marginal cost curve $ct$. In the top left graph, we have that $t^* = 1$. This corresponds to the case in which the customers value each item at least as much as the constant marginal cost $c$ and so from the firm’s point-of-view, pure bundling turns out to be the optimal strategy despite the positive marginal costs. In both the top right and bottom left graphs, we have that $0 < t^* < 1$. Hence, the optimal
strategy for the firm is to offer a size restriction on the offered bundle. However, we note that in the top right graph pure bundling, i.e. \( t = 1 \), still results in a positive profit for the firm, while in the bottom left graph it leads to a loss. Hence, even though \( \bar{V}(1) = \mu > c \) in the top right graph, it is not necessarily the case that pure bundling is the preferred strategy. Meanwhile, even when pure bundling leads to negative profits, it is still possible for the firm to asymptotically extract all of the consumer surplus using bundle size pricing as shown in the bottom left graph. This is intuitive because allowing customers to self-select their most valued items eliminates items whose valuations are below the marginal cost. Finally, in the bottom right graph we have that \( t^* = 0 \). In this case, \( P(X_1, 1 \leq c) = 1 \) and so offering a bundle of any size will result in a negative profit for the firm with probability 1. Hence, the optimal strategy in this case is to offer a bundle of size zero.

4.2. Item-dependent Marginal Costs with I.I.D. Valuations

We now consider the BSP problem in the case of heterogeneous item-dependent marginal costs. Our basic setup is similar to that in Section 4.1. In particular, a risk-neutral firm is selling \( N \geq 1 \) items to \( M \geq 1 \) customers, where \( X_{m,n} \) is the valuation which customer \( m \) places on item \( n \), and \( \{X_{m,n}; m = 1, \ldots, M; n = 1, \ldots, N\} \) is i.i.d. with common distribution \( F_X \). However, we now assume that the items no longer have identical marginal costs. In particular, we now denote by \( c(n) \geq 0 \) the item-dependent marginal cost of item \( n = 1, \ldots, N \).

In the case of item dependent marginal costs, the cost for the firm to offer a bundle of a particular size depends on the items which the customer chooses to place in it. Hence, the BSP profit function (6) of Section 4.1 must be modified accordingly. For each \( m = 1, \ldots, M \), let \( \tau_m : (1, 2, \ldots, N) \mapsto (1, 2, \ldots, N) \) be a permutation function such that \( X_{m,\tau_m(n)} = X_{m,(n)} \) for \( n = 1, \ldots, N \). In other words, \( \tau_m(n) \) is the index of the item corresponding to the \( n \)th order statistic of the valuations of customer \( m \). In order to break ties in the case where a customer values multiple items identically, we assume that the customer prefers the item with the lowest index amongst all such items. In this case, \( \tau_m \) is uniquely defined and the cost for the firm to sell a bundle of size \( n \) to customer \( m \) is given by

\[
c_{m,N}(n) = \sum_{k=0}^{n-1} c_{\tau_m(N-k)}.
\]

Note in particular that \( c_{m,N}(n) \) is a random variable since it depends on which items customer \( m \) decides to place in the bundle.

Using the same framework as in Section 4.1, we now have that given a price vector \( p \in (\mathbb{R}_+ \cup \infty)^N \), the firm’s realized profit is

\[
\pi(p) = \sum_{m=1}^{M} \sum_{n=1}^{N} (p(n) - c_{m,N}(n))1\{\zeta(X_m, p) = n\}.
\]
Now note that for each \( m = 1, ..., M, \) and \( n = 1, ..., N, \) the random variable \( 1\{\zeta(X_m, p) = n\} \) is independent of the sequence \( \{\tau_m(n); n = 1, ..., N\} \), and hence independent of the random variable \( c_{m,N}(n) \). Moreover, since \( \{X_m,n; m = 1, ..., M; n = 1, ..., N\} \) is i.i.d., it follows that \( \mathbb{P}(\tau_m(n) = k) = 1/N \) for \( k = 1, ..., N \). Moreover, denoting by

\[
F_{C,N}(x) = \frac{1}{N} \sum_{n=1}^{N} 1\{c(n) \leq x\}, \quad x \geq 0,
\]

the empirical distribution function of \( \{c(n), n = 1, \ldots, N\} \) and taking expectations in (15), it follows that \( \mathbb{E}[c_{m,N}(n)] = n\bar{c}_N \), where

\[
\bar{c}_N = \int_{\mathbb{R}_+} cdF_{C,N}(c).
\]

Now taking expectations on both sides of the realized profit (16) and recalling the basic model setup, we obtain that the firm’s profit maximization problem now reduces to solving the optimization problem

\[
\max_{p \in (\mathbb{R}_+ \cup \infty)^N+1} \mathbb{E}[\pi(p)] = \max_{p \in (\mathbb{R}_+ \cup \infty)^N+1} \left( M \sum_{n=1}^{N} (p(n) - n\bar{c}_N) \mathbb{P}(\zeta(X_1, p) = n) \right).
\]

Before presenting the main result in this section, we make the following basic assumption regarding the empirical distribution function \( F_{C,N} \).

**Assumption 1.** There exists a distribution function \( F_C \) such that \( \mathbb{P}\text{-a.s.}, \)

\[
\sup_{x \geq 0} |F_{C,N}(x) - F_C(x)| \to 0 \quad \text{as} \quad N \to \infty.
\]

Moreover, \( \mathbb{P}\text{-a.s.} \) the sequence \( \{F_{C,N}, N \geq 1\} \) is uniformly integrable.

Note that Assumption 1 states that for each \( x \geq 0, \) the number of \( c(n)\)s that are less than or equal to \( x \) is equal to \( NF_C(x) + o(N), \) uniformly in \( x. \)

The following is our main result for the BSP problem in the presence of item-dependent marginal costs but i.i.d. customer valuations.

**Theorem 2.** In the presence of item-dependent marginal costs, let

\[
\bar{c} = \int_{\mathbb{R}_+} cdF_C(c) < \infty,
\]

and assume that \( 1 - t^* = F_X(\bar{c}) \) is a continuity point of \( F_X^{-1}. \) Then, setting

\[
\mathcal{P}(n) = \begin{cases} \mathbb{E}[V_{1,N}(t^*)] - g(N) & \text{if } n = \lfloor Nt^* \rfloor, \\ +\infty & \text{if } n \neq \lfloor Nt^* \rfloor, \end{cases}
\]

where \( g(N) \in \omega_+(\sqrt{N}) \cap o(N), \) we have that

\[
\lim_{N \to \infty} \frac{\mathbb{E}[\pi(\mathcal{P})]}{M \mathbb{E}[(X_{1,1} - \bar{c})^+]} = \lim_{M \to \infty} \frac{\mathbb{E}[\pi(p^*)]}{M \mathbb{E}[(X_{1,1} - \bar{c})^+]} = 1.
\]
As in Theorem 1, we also have that if the integrability condition (12) holds, then \( E[\bar{V}_1(t^*)] \) may be replaced by \( \bar{V}(t^*) \) in the pricing policy above.

Now note that Theorem 2 implies that in the case of BSP with non-identical marginal costs but i.i.d valuations, \( E[\bar{\pi}_N(p^*)] \to M E[(X_{1,1} - \bar{c})^+] \) as \( N \to \infty \), where \( \bar{\pi}_N(p^*) = N^{-1} \pi(p^*) \). On the other hand, the normalized expected profit under perfect price discrimination obeys

\[
\frac{1}{N} \sum_{m=1}^M \sum_{n=1}^N E[(X_{m,n} - c(n))^+] \to M \int_{\mathbb{R}^+} E[(X_{1,1} - c)^+] dF_C(c) \quad \text{as} \quad N \to \infty.
\]

However, by Jensen’s inequality, it follows that

\[
E[(X_{1,1} - \bar{c})^+] \leq \int_{\mathbb{R}^+} E[(X_{1,1} - c)^+] dF_C(c),
\]

and so the optimal expected profit for BSP with non-identical marginal costs does not asymptotically achieve the expected profit under perfect price discrimination. Moreover, notice that the limiting BSP problem with heterogeneous marginal costs is equivalent to the limiting BSP problem with equal marginal costs \( \bar{c} \). Hence, for the same \( \bar{c} \) the inefficiency of the BSP policy relative to perfect price discrimination increases as the heterogeneity in the marginal costs increases.

The main source of inefficiency of BSP when the items have heterogeneous marginal costs is the fact that the firm has no control over how customers pick their items for any particular size. Consequently, the firm cannot properly discriminate between customers who are interested in low cost items from those interested in high cost items. For this reason and in order to improve the performance of the BSP policy, we propose two modified BSP policies which we refer to as “clustered bundle size pricing” and “assorted bundle size pricing”.

We next analyze the two new policies when the customers’ valuations are i.i.d. but the marginal costs are different. We assume that valuation distribution \( F_X \) is uniformly continuous with density \( f_X \). The items may have different marginal costs, however the cost of each item belongs to a set \( \{c_k, k = 1, \ldots, K\} \) which consists of \( K > 1 \) non-negative distinct costs. We say that an item is of type \( k = 1, \ldots, K \), if its cost is equal to \( c_k \). Let \( N_k \) be the number of items of type \( k \) where \( \sum_{k=1}^K N_k = N \). Without loss of generality, we assume that \( 0 \leq c_1 < c_2 < \cdots < c_K < \infty \) and that the market consists of one customer, i.e. \( M = 1 \). We denote by \( \alpha_k = \lim_{N \to \infty} \frac{N_k}{N} \) the limiting fraction of items of type \( k \) which we assume to exist. We denote the asymptotic expected profit of BSP, clustered BSP, assorted BSP, and the expected profit under perfect price discrimination by \( \text{BSP} \), \( \text{K-BSP} \), \( \text{ABSP} \), and \( \text{PPD} \), respectively.

**4.2.1. Clustered Bundle Size Pricing.** In the clustered BSP policy, the firm clusters items together based on their marginal costs and then uses a BSP policy for each cluster separately. In particular, when there are \( K \) different marginal cost types, the firm clusters the items with equal
marginal costs $c_k$ into a cluster of size $N_k$, where $\sum_{k=1}^{K} N_k = N$, and allows the customers to pick a bundle of size $n_k = 0, \ldots, N_k$ from each of the $k$-th cluster. We denote this clustered BSP by $K$-BSP. It is easy to see that the clustered BSP problem now decouples into $K$ independent BSP problems each with identical marginal cost.

From an operational point of view, the firm can either advertise the clustered BSP for each cluster separately or as a “super” bundle that includes size restrictions on different clusters. The two policies are asymptotically equivalent since in the case when each cluster is advertised separately, then for the asymptotically optimal $K$-BSP policy, the customers will end up buying the bundle sizes of all the clusters. In other words, he will asymptotically buy the super bundle. It is worth mentioning that advertising a super bundle with size restrictions on different clusters of items is a common practice for mobile phone plans where each plan comes with different quantity restrictions on the total number of messages, total minutes, and overall data-package among others.

We now analyze the $K$-BSP policy. The analysis of the $K$-BSP and later the assorted BSP policies will be based only on the limiting optimization problem. Establishing the convergence results follow the same line of proof as in Section 4.2 and hence are omitted.

Let $t_k^*$ denote the asymptotically optimal bundle size for the items of type $k$, $1 \leq k \leq K$. Likewise, denote by $\pi_k(p)$ the corresponding profit of the $k$th cluster.

The optimal size for each cluster $k = 1, \ldots, K$, is given by

$$t_k^* = 1 - F(c_k).$$

Next, recall that the cost of the items in each cluster $k$ are identical. Hence, the respective limiting normalized expected profits for BSP restricted to items of type $k$ is given by

$$\lim_{N \to \infty} \frac{E[\pi_k(p^*)]}{N} = \alpha_k \left( \bar{V}(t_k^*) - c_k t_k^* \right) = \alpha_k \left( \int_0^{t_k^*} F_X^{-1}(1-s)ds - c_k t_k^* \right).$$

Therefore, the limiting normalized expected profit of the clustered BSP policy is given by

$$K\text{-BSP} = \sum_{k=1}^{K} \alpha_k \left( \int_0^{t_k^*} F_X^{-1}(1-s)ds - c_k t_k^* \right). \quad (17)$$

We note that a 1-BSP, where all the items are clustered together, is equivalent to the BSP problem studied in Section 4.2 which we refer to as the regular BSP.

In order to understand the benefits of the $K$-BSP policy relative to the regular BSP policy, we next analyze the performance of the regular BSP policy. Recall that the expected limiting marginal cost of an item selected uniformly at random is given by

$$\bar{c} = \sum_{k=1}^{K} \alpha_k c_k.$$
Following the discussion of Section 4.2, on the average cost of selling a bundle, and since the customer’s valuations are independent from the marginal costs, the limiting average cost for selling a bundle of size $n$ to the customer is given by $\bar{c}$ and its total cost is given by $n\bar{c}$. As a result, the asymptotically optimal BSP size is

$$t^* = 1 - F_X(\bar{c}).$$

Therefore, the limiting normalized expected profit of the optimal BSP is given by

$$BSP = \lim_{N \to \infty} \frac{E[\pi(p^*)]}{N} = \bar{V}(t^*) - \bar{c}t^* = \int_0^{t^*} F_X^{-1}(1-s)ds - \bar{c}t^*. \quad (18)$$

The next lemma shows that the $K$-BSP policy outperforms the regular BSP (i.e. 1-BSP).

**Lemma 1.** The asymptotic profit of the optimal $K$-BSP policy (weakly) dominates that of 1-BSP policy.

Note that when there are $K$ types of items, $K$-BSP helps the firm control the items picked by the customers from each type, and as a result, removes the main source of inefficiency of the regular BSP. In fact, it achieves asymptotic perfect price discrimination.

**Proposition 1.** When the costs of the items belong to $\{c_k : 1 \leq k \leq K\}$ and $N_k/N \to \alpha_k$ for $k = 1, \ldots, K$, $K$-BSP achieves asymptotic perfect price discrimination.

Proposition 1 is theoretically appealing as it highlights the efficiency of a clustered BSP policy. However, it assumes a zero perfect monitoring cost on behalf of the firm. That is, the firm can perfectly observe the items that the customers are choosing without any cost. While this may be a reasonable assumption in the case of digital goods such as cell phone plans or online streaming, in many cases monitoring is costly and is imperfect. Nevertheless, even if monitoring has zero cost, creating a large number of clusters could potentially induce a cognitive burden on the customers’ behalf, leading to a disutility. For this reason, we next analyze a different approach to reducing the inefficiency of the regular BSP policy that avoids any clustering but involves pre-selecting which items to offer to the customers.

**4.2.2. Assorted Bundle Size Pricing.** Assorted BSP is another approach to reduce the efficiency loss from regular BSP. In the assorted BSP problem, the firm offers only an assortment of the items (i.e. a pre-selected subset) to the customers along with a price for each bundle size of the offered assortment. The goal of this approach is to avoid selling high cost items under a low average price. A solution to the assorted BSP problem is characterized by the subset of items
to offer, the offered bundle sizes, and the corresponding prices. This approach can be particularly appealing when the firm incurs a high monitoring cost or does not want to offer a complicated menu of prices for different clusters of items as in clustered BSP.

In the asymptotic regime of the assorted BSP problem, the firm decides on the assortment vector \( y = (y_1, y_2, \cdots, y_K) \), where \( y_k \in [0, \alpha_k] \). Recall, that \( \alpha_k \) is the limiting fraction of items with marginal cost \( c_k \). Hence, \( y_k \) represents the fraction of the items of type \( k = 1, \ldots, K \), that will be offered to the customer. Notice that \( y_k = \alpha_k \) corresponds to the case where the firm offers all the items of type \( k = 1, \ldots, K \). Hence, for a given assortment vector \( y \), the limiting weighted average cost for selling an item in any bundle size of \( y \) is given by

\[
\bar{c}(y) = \frac{\sum_{k=1}^{K} y_k c_k}{\sum_{k=1}^{K} y_k}. \tag{19}
\]

Notice that for any offered assortment consisting of \( \alpha N \) items where \( 0 \leq \alpha \leq 1 \), the limiting normalized valuation of the bundle sizes is given by \( \{\alpha \tilde{V}(t), 0 \leq t \leq 1\} \), where now \( t \) is the fraction of the \( \alpha N \) offered items. Therefore, the asymptotic normalized expected profit of the optimal assorted BSP can be written as

\[
\text{ABSP} = \max_{(y_1, y_2, \cdots, y_K)} \left( \sum_{k=1}^{K} y_k \right) \left( \int_0^{t} F_X^{-1}(1-s) ds - \bar{c}(y) \cdot t \right),
\text{ s.t. } 0 \leq t \leq 1,
0 \leq y_k \leq \alpha_k, \quad 1 \leq k \leq K.
\]

Note that the objective function of this mathematical program is continuous and the feasible set is compact. Therefore, there exists a global maximum point \( (t^*, y_1^*, y_2^*, \cdots, y_K^*) \) where the average cost of the firm for selling any bundle is \( \bar{c}(y_1^*, y_2^*, \cdots, y_K^*) \). Moreover, since \( F_X \) is uniformly continuous, the asymptotically optimal bundle size is given by \( t^* = 1 - F_X(\bar{c}(y_1^*, y_2^*, \cdots, y_K^*)) \). Thus, the optimization problem can be simplified to the following,

\[
\text{ABSP} = \max_{(y_1, y_2, \cdots, y_K)} \left( \sum_{k=1}^{K} y_k \right) \left( \int_0^{1-F_X(\bar{c}(y))} F_X^{-1}(1-s) ds - \bar{c}(y) \cdot (1 - F_X(\bar{c}(y))) \right), \tag{20}
\text{ s.t. } 0 \leq y_k \leq \alpha_k, \quad 1 \leq k \leq K.
\]

The next proposition characterizes the structure of the optimal solution for the limiting assorted BSP problem. In particular, we show that the optimal solution is always a corner point of the feasible set and follows a nested-in-cost structure. We remind the reader that we have assumed that the costs are indexed such that \( 0 \leq c_1 < c_2 < \cdots < c_K < \infty \).

**Proposition 2.** There exists an optimal solution \( (y_1^*, y_2^*, \cdots, y_K^*) \) 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: an index \( 0 \leq K^* \leq K \) can be found such that \( y_k^* = \alpha_k \) for \( k \leq K^* \) and \( y_k^* = 0 \) for \( k > K^* \).
Proposition 2 implies that the asymptotically optimal assortment consists of all items whose costs are among the $K^*$ cheapest costs. Meanwhile, the asymptotically optimal assorted BSP is to do the regular BSP on this assortment. Hence, Proposition 2 provides a simple way to find the most profitable set of products to offer in the assorted BSP problem. In particular, finding the optimal assortment is equivalent to finding the optimal value of $K^*$ among $K + 1$ different clusters. In return, the optimal set of items will consist of all items whose costs are among $\{c_1, c_2, \ldots, c_{K^*}\}$.

Clearly the profit under assorted BSP weakly dominates that of regular BSP since regular BSP is simply a feasible solution to the assorted BSP problem. Hence, assorted BSP can reduce the inefficiency of regular BSP but has no guarantee on completely eliminating it as is the case for clustered BSP.

We now illustrate the difference between the three different policies, BSP, $K$-BSP, and Assorted BSP using a numerical example. Let $M = 1$ and assume that $X_{1,n} \sim \text{Unif } [0, 1]$. Assume that there are two types of items with either high or low marginal cost. Let $c_L = 0.2$ and $c_H = 0.8$ be the marginal cost of the low and high cost item types, respectively. Also let $\alpha_L = 0.6$ and $\alpha_H = 0.4$ be the respective fractions of low and high marginal cost items. In this case, using simple algebra we obtain that the limiting expected profit under perfect price discrimination is equal to 0.2.

Figure 2 summarizes the comparison between the three proposed policies. Regular BSP is clearly inefficient as it asymptotically achieves about 78.4% of the expected profit under perfect price discrimination. Meanwhile, by creating two different clusters for the low and high marginal cost items, clustered BSP is fully efficient as expected. Meanwhile, assorted BSP dominates regular BSP as it asymptotically achieves 96% of the expected profit under perfect price discrimination. However, it does so by offering all the low cost items to the customers while refraining from offering any of the high cost items. In this case, the optimal bundle size for the low cost items is 0.8.
4.3. Item-dependent Marginal Costs with Cost-dependent Valuations

We now consider the case where items have heterogeneous marginal costs and each item’s valuation depends on its marginal cost. As in Section 4.2, we denote by $c(n) \geq 0$ the marginal cost of item $n = 1,...,N$, and we assume that the sequence of empirical distribution functions of the marginal costs satisfies Assumption 1. However, rather than assuming that the customers’ valuations are i.i.d. across items, we now assume that each item’s valuation is proportional to its marginal cost.

**Assumption 2.** For each $m = 1,...,M$, and $n = 1,...,N$, we assume that $X_{m,n} = c(n)Z_{m,n}$, where $\{Z_{m,n}; m = 1,...,M; n = 1,...,N\}$ is an i.i.d. sequence of random variables with finite mean and variance and common distribution $F_Z$, which we assume to be continuous. We further assume that $\{c(n), n = 1,...,N\}$ is uniformly bounded.

We note that assuming $\{c(n), n = 1,...,N\}$ is uniformly bounded implies that $\mathbb{P}$-a.s. the sequence $\{F_{C,N}, N \geq 1\}$ is uniformly integrable as required in Assumption 1.

Using the same notation as in Section 4.2, we have that the firm’s problem is to find the optimal price vector $p \in (\mathbb{R}_+ \cup \infty)^{N+1}$ according to the following optimization problem

$$
\max_{p \in (\mathbb{R}_+ \cup \infty)^{N+1}} \mathbb{E}[\pi(p)] = \max_{p \in (\mathbb{R}_+ \cup \infty)^{N+1}} \sum_{m=1}^{M} \sum_{n=1}^{N} \mathbb{E}[(p(n) - c_m(n))1\{\zeta(X_m, p) = n\}].
$$

(21)

However, unlike the setting in Section 4.2, we now have that $1\{\zeta(X_m, p) = n\}$ and $c_m(n)$ are no longer independent for $n = 1,...,N$. This is due to the fact that in the present setting an item with a higher marginal cost is more likely to be selected in a bundle.

Now let

$$
\bar{V}(t) = \int_0^t F_X^{-1}(1 - s)ds \text{ for } 0 \leq t \leq 1,
$$

where $F_X = \int_0^\infty F_Z(x/c) dF_C$. In fact, $\bar{V}(t)$ represents the limiting normalized valuation per customer for a bundle of size $t \in [0,1]$ (see Lemma A5 in the Appendix). Since the valuations are non-negative, it is straightforward to show that $\bar{V}(t)$ is a non-decreasing concave function.

Furthermore, let

$$
\bar{c}(t) = \int_0^\infty c(1 - F_Z(F_X^{-1}(1-t)/c)) dF_C(c) \text{ for } 0 \leq t \leq 1.
$$

(22)

We note that $\bar{c}(t)$ represents the limiting normalized cost for a bundle of size $t \in [0,1]$ (see Lemma A6 in Appendix).

Now denote by $\mathcal{T}^*$ the set of optimal solutions to the optimization problem

$$
\max_{t \in [0,1]} (\bar{V}(t) - \bar{c}(t)).
$$

(23)
In general, the set $T^*$ may not be a singleton. This is due to the fact that $\bar{c}(t)$ is neither concave nor convex in general. In particular, taking the derivative of $\bar{c}(t)$ in (22) with respect to $t \in (0, 1)$, we obtain that
\[
\frac{d\bar{c}(t)}{dt} = \left( \int_0^{\infty} f_Z(F_X^{-1}(1-t)/c)dF_C(c) \right) / \left( \int_0^{\infty} f_Z(F_X^{-1}(1-t)/c)\frac{dF_C(c)}{c} \right).
\]
Taking further the second order derivative reveals that $\bar{c}(t)$ is in general neither convex nor concave.

We now state our main result in this section.

**Theorem 3.** In the presence of item-dependent marginal costs with limiting empirical distribution function $F_C$, and cost-dependent valuations with limiting empirical distribution function $F_X$, setting
\[
P(n) = \begin{cases} N\mathbb{E}[\bar{V}_{1,N}(t^*)] - g(N) & \text{if } n = [Nt^*], \\ +\infty & \text{if } n \neq [Nt^*], \end{cases}
\]
where $t^* \in T^*$ and $g(N) \in \omega, (\sqrt{N}) \cap o(N)$, we have that
\[
\lim_{N \to \infty} \frac{\mathbb{E}[\pi(P)]}{MN(\bar{V}(t^*) - \bar{c}(t^*))} = \lim_{N \to \infty} \frac{\mathbb{E}[\pi(p^*)]}{MN(\bar{V}(t^*) - \bar{c}(t^*))} = 1.
\]
As with the previous theorems, if we assume that the tail condition (12) holds with respect to $F_Z$, then one may replace $\mathbb{E}[\bar{V}_{1,N}(t^*)]$ in the pricing policy of Theorem 3 with the simpler value $\bar{V}(t^*)$. The above theorem states that in the present setting the asymptotically optimal BSP policy is to offer one size $t^*$. However, $t^*$ does not have a simple closed form and may not be unique. We also point out that the above theorem does not guarantee that the optimal BSP can asymptotically achieve perfect price discrimination. In fact, in most cases it does not.

To elaborate on this further, in the present setting the limiting normalized profit under perfect price discrimination for a customer $m = 1, \ldots, M$, is given by
\[
\frac{1}{N} \sum_{m=1}^{M} \sum_{n=1}^{N} (X_{m,n} - c_n)^+ = \frac{1}{N} \sum_{m=1}^{M} \sum_{n=1}^{N} c_n(Z_{m,n} - 1)^+ \Rightarrow M \cdot \bar{c}(1)\mathbb{E}[(Z_{1,1} - 1)^+] \text{ as } N \to \infty.
\]
On the other hand, using an appropriate change-of-variables and letting $u = F_X^{-1}(1-t)$, the limiting normalized BSP profit can be written as
\[
M \cdot (\bar{V}(t) - \bar{c}(t)) \equiv M \cdot \int_0^{\infty} \int_{u/c}^{\infty} cf_Z(z) (z-1) dzdF_C(c).
\]
Notice that if $F_Z$ is strictly increasing then we have a one-to-one correspondence between the two representations in (24) where $t = 1 - F_X(u)$.

The previous two expressions suggest that the firm is only able to achieve perfect price discrimination if it can discern between those products for which $Z_{m,n} > 1$ for customer $m = 1, \ldots, M$, and those products for which $Z_{m,n} \leq 1$. 
Consider the case in which \( F_Z(1) = 0 \). In other words, \( \mathbb{P}(Z_{m,n} > 1) = 1 \). In this case, \( X_{m,n} - c_n = c_n(Z_{m,n} - 1) > 0 \) almost surely for each \( m = 1, \ldots, M \), and \( n = 1, \ldots, N \). Consequently, pure bundling turns out to be the asymptotically optimal strategy for the firm. Indeed, in this case one may rigorously verify that the optimal solution to the optimization problem (23) is given by \( t^* = 1 \) and that it achieves asymptotic perfect price discrimination. We omit the details.

More generally, it turns out that a necessary and sufficient condition can be provided for when asymptotic perfect price discrimination is achieved in the present setting. 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 \( Z_{m,n} \) may achieve, and \( z_l \) represents the largest value less than 1 that \( Z_{m,n} \) 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 3.** In the present setup, asymptotic perfect price discrimination is achieved if and only if \( z_l c_u \leq z_u c_l \). Moreover, if the previous inequality holds, then asymptotic perfect price discrimination is achieved by setting

\[
\mathcal{P}(n) = \begin{cases} 
N \mathbb{E}[\bar{V}_{1,N}(t^*)] - g(N) & \text{if } n = \lfloor N t^* \rfloor, \\
+\infty & \text{if } n \neq \lfloor N t^* \rfloor,
\end{cases}
\]

where \( t^* = 1 - F_Z(1) \) and \( g(N) \in \omega_+(\sqrt{N}) \cap o(N) \).

Note that one implication of Proposition 3 is that perfect price discrimination may only be achieved if \( Z_{m,n} \) “avoids 1” or \( F_C \) is degenerate. Meanwhile, if \( z_l = z_u = 1 \), then perfect price discrimination cannot be achieved asymptotically since \( c_l < c_u \) (with the exception of the degenerate case where \( c_l = c_u \)).

We now present an example where asymptotic perfect price discrimination cannot be achieved by BSP, but nevertheless the optimization problem (23) can be solved explicitly. Consider the case in which \( F_Z \) is the uniform distribution on \([1/2, 3/2]\). Loosely speaking, since the valuation of a customer \( m \) for item \( n \) is given by \( X_{m,n} = c_n Z_{m,n} \), then in the limit only half of the items will be valued at greater than their marginal costs (since \( \mathbb{P}(Z_{m,n} > 1) = 1/2 \)). Also assume there are only two types of marginal costs with \( c_L = 4 \) and \( c_H = 8 \), where \( F_C(c) = (3/4)1\{c \geq 4\} + (1/4)1\{c \geq 8\} \), \( c \geq 0 \). By a straightforward application of Lemmas A4 and A6, we obtain that

\[
F_X(x) = (3/4)F_Z(x/4) + (1/4)F_Z(x/8) \quad \text{for } x \geq 0.
\]
Since $F_Z$ is strictly increasing, we then have a one-to-one correspondence between $t$ and $u$ in the equivalent representation in (24). By considering the problem as a function of $u$, we then obtain

$$
\bar{V}(1 - F_X(u)) - \bar{c}(1 - F_X(u)) = \begin{cases} 
2 \int_{u/8}^{3/2} (z - 1)dz & \text{if } 6 \leq u \leq 12, \\
3 \int_{u/4}^{3/2} (z - 1)dz + 2 \int_{u/8}^{3/2} (z - 1)dz & \text{if } 4 \leq u \leq 6, \\
3 \int_{u/4}^{3/2} (z - 1)dz & \text{if } 2 \leq u \leq 4, \\
0 & \text{otherwise.}
\end{cases}
$$

(25)

In this case, the profit function is neither convex nor concave in $u$, in fact it is bi-modal. By working out the details, one can show that $u^* = 32/7$ and $t^* = 1/2$, where the limiting ratio of the BSP profit to perfect price discrimination is given by

$$
\frac{\bar{V}(1/2) - \bar{c}(1/2)}{\bar{c}(1/2) E[(Z_{1,1} - 1)^+]} = 23/35.
$$

Notice that in this example the asymptotically optimal size is still $t^* = 1 - F_Z(1)$ however BSP does not asymptotically achieve perfect price discrimination. The intuition behind this is that without the separation condition in Proposition 3, some customers will have items whose valuations are below the marginal cost in the top half of their most valued products, which leads to an efficiency loss. Also, some customers will have items whose valuations are above the marginal cost in the bottom half of their most valued products, which leads to an opportunity loss.

5. Bundle Size Pricing with Budgets

In this section, we consider the case where customers have budget constraints that limit their ability to pay for a bundle. As in Section 4.1, the firm is selling $N \geq 1$ items to a market size of $M \geq 1$ customers whose valuations $\{X_{m,n}, 1 \leq n \leq N, 1 \leq m \leq M\}$ are i.i.d. with a common distribution $F_X$ that has a finite mean $\mu$ and variance $\sigma^2$. We also assume that the items have an identical marginal cost $c > 0$. In Section 5.1, we consider the setting where each of the customers has an identical deterministic budget $b > 0$ that is known to the firm. Next, in Section 5.2, we extend the model to a setting in which customers have heterogeneous budgets that are private information.

5.1. Homogeneous Deterministic Budgets

We first consider the setting where each of the customers has the same budget $b > 0$ that is known to the firm. Regardless of the budget constraints and for a given bundle size $n = 1, \ldots, N$, each customer $m = 1, \ldots, M$, continues to select the items that maximize his utility. In particular, a customer $m$ selects the items to include in a bundle of size $n$ according to the following integer program

$$
\begin{align*}
\max & \sum_{r=1}^{N} I_{m,r} X_{m,r} \\
\text{subject to } & \sum_{r=1}^{N} I_{m,r} = n \\
& I_{m,r} \in \{0,1\} \text{ for } r = 1, \ldots, N.
\end{align*}
$$

(26)
Note that the above problem does not depend on the budget, since the budget does not affect which items the customer values the most. Nonetheless, the budget will affect his ability to pay for the bundle.

Using the same notation for order statistics as in Section 3, the optimal value of the optimization problem (26) is given by

\[ V_{m,N}(n) = \sum_{k=0}^{n-1} X_{m,(N-k)}, \]  

where \( V_{m,N}(0) = 0 \). Hence, the vector \((V_{m,N}(0), V_{m,N}(1), \ldots, V_{m,N}(N))\) represents the intrinsic valuation which customer \( m \) places on bundles of size 0 through \( N \), regardless of his budget. Now, for a given vector of bundle size prices \( p = (p(0) = 0, p(1), \ldots, p(N)) \), the customer takes into account his limited budget and purchases the bundle of an appropriate size in order to maximize his surplus subject to his budget constraint. That is, customer \( m \) will purchase a bundle of size

\[ \zeta(X_m, p, b) \in \arg \max_{\{n=0, \ldots, N; \ p(n) \leq b\}} (V_{m,N}(n) - p(n)). \]

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

\[ \pi(p, b) = \sum_{m=1}^{M} \sum_{n=1}^{N} (p(n) - nc) 1\{\zeta(X_m, p, b) = n\}, \]  

and the firms pricing problem is given by the optimization problem

\[ \max_{p \in (\mathbb{R}_+ \cup \infty)^{N+1}} M \cdot \sum_{n=1}^{N} (p(n) - nc) \mathbb{P}(\zeta(X_1, p, b) = n), \]

where we denote by \( \mathcal{P}^* \) its set of optimal solutions. Now letting \( p^* \in \mathcal{P}^* \), we have that \( E[\pi(p^*, b)] \) is the maximum expected profit that the firm may achieve from a bundle size pricing policy with budget-constrained customers.

We now proceed to obtain an upper bound on the firm’s expected profit in the presence of homogeneous deterministic budgets. To do so, we first introduce the indirect utility of a customer \( m \) for a bundle of size \( n \) which can be written as

\[ V_{m,N}(n, b) = V_{m,N}(n) \wedge b, \]

where \( a \wedge b = \min\{a, b\} \). The indirect utility for bundle size \( n = 1, \ldots, N \), is the maximum attainable utility by the customer for a given budget \( b \) (see for example Chapter 3 of Mas-Colell et al. (1995)).

For each \( m = 1, \ldots, M \), and \( n = 1, \ldots, N \), we have that \( \{\zeta(X_m, p, b) = n\} \subseteq \{p(n) \leq V_{m,N}(n, b)\} \). Hence, it follows from (28) that

\[ \pi(p, b) \leq \sum_{m=1}^{M} \sup_{n=0,1,\ldots,N} (V_{m,N}(n, b) - nc). \]
Now taking expectations on both sides of the above, we obtain

$$\frac{\mathbb{E}[\pi(p, b)]}{M \mathbb{E}[\sup_{n=0,1,...,N} (V_{m,N}(n,b) - nc)]} \leq 1 \text{ for } p \in (R_+ \cup \infty)^N. \quad (30)$$

First notice that if $b \leq c$, then clearly the firm should not offer any bundle size (or equivalently offer a bundle size 0). On the other hand, if we assume that the budget $c < b < \infty$ is strictly less than the upper bound of the support of $F_X$, then from extreme value theory, it is straightforward to show that offering only a bundle of size $n = 1$ at price $b$ asymptotically achieves the bound in (30) with equality. For this reason, we introduce an assumption that the budget scales with the number of items, however the normalized budget remains constant for any $N$. More specifically, we now denote the budget by $b_N$ and we assume that the normalized budget $b_N/N = \bar{b}$ for every $N \in \mathbb{N}$.

For each $m = 1, ..., M$, now define the normalized indirect utility as

$$\bar{V}_{m,N}(t, \bar{b}) = \frac{1}{N} V_{m,N}(\lfloor Nt \rfloor) \wedge b_N$$

$$= \bar{V}_{m,N}(t) \wedge \bar{b} \text{ for } 0 \leq t \leq 1.$$

Also define $t_\bar{b} = \inf \{ t \in [0,1] : \bar{V}(t) \geq \bar{b} \}$, where we set $\inf \{ \emptyset \} = 1$. Recall that as in Section 4.1, $\bar{V}(t) = \int_0^t F_X^{-1}(1-s)ds$ is the limiting process of $\bar{V}_{m,N}(t)$ as $N \to \infty$. Hence, $t_\bar{b}$ is the smallest bundle size for which the budget constraint is binding in the limit. Let $\bar{V}(t, \bar{b}) = \bar{V}(t) \wedge \bar{b}$. Then, by the continuity of $\bar{V}(t)$, we have that $\bar{V}(t, \bar{b}) = \bar{b}$ for all $t \geq t_\bar{b}$ and, since $\bar{V}(t)$ is non-decreasing, it follows that $\bar{V}(t, \bar{b}) = \bar{V}(t \wedge t_\bar{b})$ for $t \in [0,1]$.

We now show that it is relatively straightforward to incorporate budgets into the bundle size pricing problem.

**Proposition 4.** Given identical marginal costs $c > 0$ with a positive normalized budget $\bar{b} > 0$, assume that $1 - t^* = F_X(c)$ is a continuity point of $F_X^{-1}$. Then, setting

$$\mathcal{P}(n) = \begin{cases} N \mathbb{E}[\bar{V}_{1,N}(t^*) \wedge \bar{b}] - g(N) & \text{if } n = [N (t^* \wedge t_\bar{b})], \\ +\infty & \text{if } n \neq [N (t^* \wedge t_\bar{b})], \end{cases}$$

where $g(N) \in \omega_+(\sqrt{N}) \cap o(N)$, we have that

$$\lim_{N \to \infty} \frac{\mathbb{E}[\pi(\mathcal{P}, b_N)]}{M \mathbb{E}[\sup_{n=0,1,...,N} (V_{m,N}(n,b_N) - nc)]} = \lim_{N \to \infty} \frac{\mathbb{E}[\pi(p^*, b_N)]}{M \mathbb{E}[\sup_{n=0,1,...,N} (V_{m,N}(n,b_N) - nc)]} = 1.$$

Similar to the follow up remarks on Theorem 1, if we assume that the integrability condition (12) holds, one can replace $\mathbb{E}[\bar{V}_{1,N}(t^*) \wedge \bar{b}]$ in the suggested price by the simpler value $\bar{V}(t^* \wedge t_\bar{b})$.

The intuition behind Proposition 4 is simple and is illustrated graphically using Figure 3 for the case of customers with high and low budgets. In Figure 3(a), we plot the asymptotic indirect
utility of a population of customers with a high budget level. In this case, since the budget level is non-binding at optimality, then the optimal size remains $t^\star$. On the other hand, in Figure 3(b), the budget is binding at optimality where $t_b < t^\star$. In this case, due to the low budget, the customers can only pay their budget for any bundle size $t \geq t_b$. Therefore, due to the positive marginal cost, it is not optimal for the firm to offer any bundle size $t > t_b$ including $t^\star$. Meanwhile, for $0 \leq t \leq t_b$ the consumer surplus is increasing. Therefore, the optimal solution is to offer the bundle size $t_b$ which is the least size for which the budget is binding. In both cases, the firm asymptotically achieves perfect price discrimination with respect to the indirect utility.

![Diagram](a) High budget

![Diagram](b) Low budget

**Figure 3** Bundle size pricing under different budget levels

### 5.2. Unknown Heterogeneous Budgets

We now consider a general setting in which the customers have heterogeneous budgets that are private information. We use the same notation as before except now customer $m = 1, \ldots, M$, has a budget $b_{m,N}$, whose normalized value is denoted by $\bar{b}_m = b_{m,N}/N$ for all $N \in \mathbb{N}$. Unlike the deterministic case in which the pricing policy is a function of the budget, the pricing policy in the case of unknown budgets cannot depend on any private information. In the following theorem, we propose a bundle size pricing policy for which the firm can asymptotically extract all the consumer surplus subject to the budgets.
Theorem 4. Given identical marginal costs $c > 0$ with a customer specific normalized budget $b_m$ for $m = 1, \ldots, M$, that is private information, assume that $1 - t^* = F_X(c)$ is a continuity point of $F_X^{-1}$. Then, offering the following pricing curve,

$$
\mathcal{P}([Nt]) = \begin{cases} 
N \mathbb{E}[\tilde{V}_{1,N}(t)] - h(t)g(N) & \text{if } t \leq t^*, \\
+\infty & \text{if } t > t^*,
\end{cases}
$$

where $h(t) \in \mathbb{R}_+$ is a strictly increasing function and $g(N) \in \omega_+(N^{1/2}) \cap o(N)$, we have that

$$
\lim_{N \to \infty} \frac{\sum_{m=1}^M \mathbb{E}[\pi(\mathcal{P}, \{b_m,N\}_{m=1}^M)]}{\sum_{m=1}^M \mathbb{E}[\sup_{n=0,1,\ldots,N}(V_{m,N}(n,b_m,N) - nc)]} = 1.
$$

Notice that the asymptotically optimal bundle size pricing policy is no longer to offer one size $t^*$. Instead, the firm now offers a menu of prices for all bundle sizes $t$ where $0 \leq t \leq t^*$. Again, if we assume that condition (12) holds, one can replace $\mathbb{E}[\tilde{V}_{1,N}(t)]$ in the suggested price by the simpler function $\tilde{V}(t)$.

The intuition behind Theorem 4 may be illustrated graphically. In Figure 4, we plot the asymptotically normalized indirect utility for a population that consists of two types of customers with either high or low budget levels, i.e. $\tilde{V}(t, \bar{b}_H)$ and $\tilde{V}(t, \bar{b}_L)$. The firm will only offer bundle sizes up to $t^*$ which is reflected in the limiting pricing curve $\tilde{\mathcal{P}}$ where $\tilde{\mathcal{P}}(t) = \lim_{N \to \infty} \mathcal{P}([Nt])/N$ for $t \in [0,1]$. The difference between the indirect utility curves and the price curve represent the surplus of each customer type. Now, notice that the pricing curve is constructed in such a way...
that the surplus of any customer type is increasing for all the bundles that are affordable. As a result, the high-budget customers will end up choosing the bundle of size $t^*$ while the low-budget customers will exhaust all their budgets by choosing the bundle of size $t_b^L$.

We note here that the differences between the price curve and the valuation curves are exaggerated for illustration purposes. In fact, the pricing curve should coincide with $\bar{V}(t, \bar{b}_H)$ to reflect the fact that the suggested policy extracts all the surplus. Hence, while it is true that the surplus is higher for the high-budget customers, the surplus of both customer types goes to zero in the limit.

6. Bundle Size Pricing with Multiple Customer Types

In this section, we study the bundle size pricing problem when the valuations across customers are no longer identically distributed. More specifically, we assume that a customer’s valuations depend on his type, where he can either be a high valuation or a low valuation type. We assume that the customer type is private information, however from the firm’s perspective each customer $m = 1, \ldots, M$, is a high type with probability $\alpha \in (0, 1)$ and a low type with probability $1 - \alpha$. For $m = 1, \ldots, M$, the customer’s valuation vector is given by the random vector

$$X_m = 1\{m \text{ is high type}\}X_m^H + 1\{m \text{ is low type}\}X_m^L,$$

where $X_m^H$ is the random valuation vector if customer $m$ is a high type, and $X_m^L$ is the random valuation vector if customer $m$ is a low type. We assume that each of $\{X_{m,n}^H; m = 1, \ldots, M; n = 1, \ldots, N\}$ and $\{X_{m,n}^L; m = 1, \ldots, M; n = 1, \ldots, N\}$ are i.i.d with common distributions $F_{X^H}$ and $F_{X^L}$, respectively, where each has a finite mean ($\mu^L$ and $\mu^H$) and variance ($\sigma^2_L$ and $\sigma^2_H$).

In order to model the disparity in valuations among the high and low types, we assume that $F_{X^H}$ and $F_{X^L}$ satisfy strict first-order stochastic dominance as defined below.

**Assumption 3 (Strict First-Order Stochastic Dominance).** We assume that

$$F_{X^L}(x) > F_{X^H}(x) \quad \text{for all} \quad x > 0. \quad (31)$$

Using the same notation for order statistics as in Section 3 and if customer $m$ is of high type, we denote his valuation for a bundle of size $n = 1, \ldots, N$, by

$$V_{m,N}^H(n) = \sum_{k=0}^{n-1} X_{m,(N-k)}^H,$$

Meanwhile, if customer $m$ is a low type, we denote his valuation for a bundle of size $n$ by

$$V_{m,N}^L(n) = \sum_{k=0}^{n-1} X_{m,(N-k)}^H.$$
Hence, for \( m = 1, \ldots, M \), customer \( m \)'s valuation for a bundle of size \( n = 1, \ldots, N \), is given by the random variable

\[
V_{m,N}(n) = 1\{m \text{ is high type}\} V_{m,N}^H(n) + 1\{m \text{ is low type}\} V_{m,N}^L(n),
\]

where the valuation for the no-purchase option is given by \( V_{m,N}(0) = 0 \).

Given a price vector \( p = (p(0) = 0, p(1), \ldots, p(N)) \in (\mathbb{R}_+ \cup \infty)^{N+1} \), the customer chooses the bundle size that maximizes his surplus. That is, customer \( m \) chooses the bundle size

\[
\zeta(X_m, p) \in \arg \max_{n=0, \ldots, N} (V_{m,N}(n) - p(n)),
\]

where for convenience we assume that the customer breaks ties by choosing the smallest size. Equivalently, we have that

\[
\zeta(X_m, p) \equiv 1\{m \text{ is high type}\} \zeta(X^H_m, p) + 1\{m \text{ is low type}\} \zeta(X^H_m, p).
\]

We assume throughout this section that the firm incurs equal marginal cost \( c \geq 0 \) for every item sold. In this case, the firm’s profit given a price vector \( p \) is given by the random variable

\[
\pi(p) = \sum_{m=1}^{M} \sum_{n=1}^{N} (p(n) - nc) 1\{\zeta(X_m, p) = n\}. \tag{32}
\]

The firm is risk-neutral and hence is interested in maximizing its expected profit which can be written as

\[
\mathbb{E}[\pi(p)] = M \cdot \sum_{n=1}^{N} (p(n) - nc) \left[ \alpha H \mathbb{P}(\zeta(X^H_1, p) = n) + (1 - \alpha)\mathbb{P}(\zeta(X^L_1, p) = n) \right].
\]

Consequently, the problem faced by the firm reduces to solving the following optimization problem

\[
\sup_{p \in (\mathbb{R}_+ \cup \infty)^{N+1}} \sum_{n=1}^{N} (p(n) - nc) \left[ \alpha H \mathbb{P}(\zeta(X^H_1, p) = n) + (1 - \alpha)\mathbb{P}(\zeta(X^L_1, p) = n) \right].
\]

We now proceed to get an upper bound on \( \mathbb{E}[\pi(p^*)] \) where \( p^* \) is an optimal solution to the above optimization problem. For each \( m = 1, \ldots, M \), we define a new random variable \( Z_{m,N}(X^H_m, X^L_m) \) which is the objective function of the following optimization problem,

\[
Z_{m,N}(X^H_m, X^L_m) = \sup_{p \in (\mathbb{R}_+ \cup \infty)^{N+1}} \sum_{n=0}^{N} (p(n) - nc) \left[ \alpha H 1\{\zeta(X_m, p) = n\} + (1 - \alpha) 1\{\zeta(X_m, p) = n\} \right].
\]

Notice that \( Z_{m,N}(X^H_m, X^L_m) \) represents the maximum profit that a bundle size pricing firm can obtain if it can observe the valuation vectors \( X^H_m \) and \( X^L_m \) without observing the customer type. In fact, \( Z_{m,N}(X^H_m, X^L_m) \) is the realized profit under second-degree price discrimination.
Taking the expectation of \(Z_{m,N}(X^H_m,X^L_m)\) and noticing the interchange of the sup and the expectation operators relative to \(E[\pi(p^*)]\), it follows that

\[
\frac{E[\pi(p)]}{ME[Z_{1,N}(X^H_1,X^L_1)]} \leq 1 \quad \text{for} \quad p \in (\mathbb{R}_+ \cup \infty)^{N+1}.
\] (33)

Notice that from (33) a bundle size pricing firm may no longer achieve the expected profit under perfect price discrimination as its expected profit is bounded above by that under second-degree price discrimination. Nevertheless, in the next theorem we propose a bundle size pricing policy which asymptotically achieves this bound. Letting \(\bar{V}^L(t) = \int_0^t F^{-1}_X(1-s)ds\) and \(\bar{V}^H(t) = \int_0^t F^{-1}_X(1-s)ds\), we now state our main result in this section.

**Theorem 5.** In the presence of identical marginal costs \(c \geq 0\) and two customer types, assume that \(1 - \bar{t}^H = F_{X^H}(c)\) is a continuity point of \(F_{X^H}^{-1}\) and let

\[
t^L \in \arg \max_{t \in [0,1]} (\bar{V}^L(t^L) - \alpha \bar{V}^H(t^L) - (1 - \alpha)t^L c).
\] (34)

Then, setting

\[
\mathcal{P}(n) = \begin{cases} 
N\mathbb{E}[\bar{V}^L_{1,N}(t^{L,*})] - g(N) & \text{if} \quad n = [Nt^{L,*}], \\
N\left[\mathbb{E}[\bar{V}^H_{1,N}(\bar{t}^H)] - (\mathbb{E}[\bar{V}^H_{1,N}(t^{L,*})] - \mathbb{E}[\bar{V}^L_{1,N}(t^{L,*})])\right] - h(N) & \text{if} \quad n = [N\bar{t}^H], \\
+\infty & \text{otherwise,}
\end{cases}
\]

where \(g(N), h(N) \in \mathbb{R}_+ (\sqrt{N}) \cap o(N)\) and \(\limsup_{n \to \infty} \frac{g(N)}{h(N)} < 1\), we have that

\[
\lim_{N \to \infty} \frac{\mathbb{E}[\pi(\mathcal{P})]}{ME[Z_{1,N}(X^H_1,X^L_1)]} = \lim_{N \to \infty} \frac{\mathbb{E}[\pi(p^*)]}{ME[Z_{1,N}(X^H_1,X^L_1)]} = 1.
\]

Similar to the follow up remarks on Theorem 1, if we assume that condition (12) holds for \(X^H\) and \(X^L\), one can replace \(E[\bar{V}^L_{1,N}(t)]\) and \(E[\bar{V}^L_{1,N}(t)]\) in the suggested price by the simpler functions \(\bar{V}^H(t)\) and \(\bar{V}^L(t)\) respectively.

The result of Theorem 5 is better understood using a graphical illustration. In Figure 5, we plot the limiting normalized valuation of the high and the low types. By the stochastic dominance assumption, the valuation curve of the high valuation customers is strictly above that of the low valuation customers for any size \(t > 0\). Next, notice that \(\bar{t}^H = 1 - F_{X^H}(c)\) is the efficient bundle size that the firm would offer if there was only high valuation customers in the market. In the presence of two types, the firm designs a bundle for the low types that has a lower size than that designed for the high types and it extracts all the surplus from the low types. On the other hand, the high valuation customers are served with their efficient bundle size \(\bar{t}^H\) but they retain a positive surplus due to the information rent. In this case, the bundle designed for the high types is priced low enough to ensure that it is incentive compatible for them to choose the bundle size \(\bar{t}^H\).
Notice that the univariate optimization problem (34) is in general hard to solve. However, we now consider a special case under which we have a simple closed form for \( t^L,\star \). We make a classical assumption as in the screening model literature where we assume that for \( m = 1,\ldots,M, \) and \( n = 1,\ldots,N, \) \( X_{m,n}^H = X_{m,n} \) and \( X_{m,n}^L = \theta X_{m,n} \) where \( 0 < \theta < 1 \) and \( \{X_{m,n}; m = 1,\ldots,M; n = 1,\ldots,N\} \) is an i.i.d. collection of random variables with a common distribution \( F_X \) that has a finite mean \( \mu \) and variance \( \sigma^2 \). Usually \( \theta \) is interpreted as the income effect on the valuation of the low types (see for example Armstrong (1999) and the references therein). Clearly, Assumption 3 is satisfied by this model for all \( x \) in the support of \( F_X \).

Under this particular specification, it is straightforward to show that

\[
\tilde{V}^L(t) = \theta \tilde{V}^H(t) = \theta \int_0^t F_X^{-1}(1-s)ds \quad \text{for} \quad t \in [0,1].
\]

Hence, the univariate optimization problem (34) simply reduces to

\[
\max_{t^L \in [0,1]} \left[ (\theta - \alpha) \tilde{V}^H(t) - (1 - \alpha) tc \right],
\]

which is a concave maximization problem for \( \theta > \alpha \). Meanwhile, for \( \theta \leq \alpha \), then clearly \( t^L,\star = 0 \).

The next proposition provides a characterization for optimal solution of the above problem.

**Proposition 5.** Given that \( 1 - t^L = F_X(\frac{1 - \alpha}{\theta - \alpha} c) \) is a continuity point of \( F_X^{-1} \), we obtain that

\[
t^L,\star = \begin{cases} 
0 & \text{if } \theta \leq \alpha, \\
1 - F_X(\frac{1 - \alpha}{\theta - \alpha} c) & \text{if } \theta > \alpha.
\end{cases}
\]
We skip the proof since it is straightforward. The result follows from the concavity of the objective function in (35) and its respective first order condition. Now notice that for $0 \leq \theta \leq \alpha$, the firm prices out the low valuation customers as the information rent is too high to justify serving both types simultaneously. On the other hand, if $1 > \theta > \alpha$, then the firm should serve both customers with two different designed bundles.

A few comments are in order. Since $0 < \theta < 1$, then the low valuation customers are always served with an inefficient smaller size compared to the size that they would have been offered if there were no high valuation customers. Moreover, as $\theta$ decreases, i.e. the valuation disparity increases, then the surplus retained by the high valuation customers increases up to a point when it is no longer profitable to serve the low types. At that point, the firm only serves the high types, and extracts all their surplus. Finally, in the case of zero marginal cost, i.e. $c = 0$, then the asymptotically optimal solution is to either price out the low customers ($t_{L,*} = 0$ if $\theta \leq \alpha$), or to treat the high types as if they were low types and sell the pure bundle to both customer types ($t_{L,*} = 1$) at price $p(N) = \theta N \mu - g(N)$ for any $g(N) \in \omega_+ (\sqrt{N}) \cap o(N)$.

Our asymptotically optimal bundle size pricing policy is not unique since from Figure 5 it is easy to see that the firm can devise a continuum of pricing policies that are asymptotically optimal as long as they ensure that the high valuation customers end up buying the size $t^H$ while the low types buy the size $t_{L,*}$. One special example of such policies is the two-part tariff proposed by Armstrong (1999) that involves a size restriction on the maximum number of products that a low type can purchase which is equivalent to $t_{L,*}$. Meanwhile, the high types have no size restrictions but by the design of the tariff they will end up buying at $t^H$.

7. Convergence Rate

We now study the convergence rate of the ratio in (11) and provide a tighter characterization of the optimal $g^*(N)$ function that appears in the pricing policy. To keep the analysis simple, we consider the case of zero marginal costs and i.i.d valuations. The analysis can be extended to non-zero marginal costs settings. From Section 3 and in the case of zero marginal costs $c = 0$, we have that the asymptotically optimal bundle size pricing policy is to offer the full bundle, i.e. size $N$ (or $t^* = 1$ for normalized sizes). Moreover, there exists a continuum of asymptotically optimal prices given by $p(N) = N \mu - g(N)$ where $g(N) \in \omega_+ (N^{1/2}) \cap o(N)$. Letting, $p^*(N)$ be an optimal solution to the pure bundling problem, we denote by $g^*(N) = N \mu - p^*(N)$. We first have the following refinement regarding $g^*(N)$.

**Proposition 6.** For $c = 0$, we have that $g^*(N) \in \omega_+ (N^{1/2}) \cap o(N^\beta)$ for any $\beta > 1/2$.

From a practical standpoint, Proposition 6 states that the optimal $g^*(N)$ cannot grow much faster than $N^{1/2}$. More importantly, from a theoretical standpoint, it allows us to invoke Cramer’s
moderate deviation principle in order to get a more exact characterization of $g^\star(N)$ and hence the convergence rate of the ratio of profits.

Before stating the convergence rate result we make the following assumption which is usually referred to as Cramer’s condition.

**Assumption 4 (Cramer’s Condition).** There exists a constant $K > 0$ such that $\mathbb{E}[\exp(tX_1)] < \infty$ for $|t| < K$.

**Theorem 6.** Given Assumption 4 and $c = 0$, we have that

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

and

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

Theorem 6 provides a tighter characterization of the optimal pricing policy for BSP under zero marginal cost relative to (5). It states that the optimal pricing policy for the pure bundle is $p^\star(N) = N\mu - \sigma \sqrt{N \log N} + o(\sqrt{N \log N})$, and that the ratio of the profits relative to perfect price discrimination is $\mathbb{E}[\pi(p^\star)]/MN\mu = 1 - \sigma / \sqrt{\log N/N} + o(\sqrt{\log N/N})$.

8. **Conclusion**

Although bundle size pricing (BSP) is a simple form of bundling, it is a hard multi-dimensional 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 multi-product firm that sells a large number of products. We show that in the limit, the BSP problem transforms from a hard multi-dimensional problem to a simple multi-unit monopolistic pricing problem with concave non-decreasing 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, when customers draw their valuations from the same distribution, then, as the number of items grows large, we show that the asymptotically optimal BSP involves selling only one size. However, regarding the relative performance of BSP to mixed bundling, we show that it highly depends on the properties of the products’ marginal costs. A previous numerical study by Chu et al. (2011) has documented that the relative performance of BSP to mixed bundling deteriorates with larger marginal costs but, on average, BSP attains about 98% of the profit of mixed bundling. On the contrary, our theoretical analysis shows that, for equal marginal costs, an increase or
decrease in the marginal costs level has no impact on the asymptotic performance of BSP relative to mixed bundling. 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 fact, as the heterogeneity in marginal costs increases, the ratio of the asymptotic BSP profits to mixed bundling decreases and can be very low.

In order to overcome the limitations of BSP in the presence of heterogeneous marginal costs, we propose two new BSP policies that we call “clustered BSP” and “assorted BSP”. In clustered BSP, the items are clustered into groups of products with homogeneous marginal costs and then BSP is 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 relative performance of BSP 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. In this case, offering BSP with one size is no longer asymptotically optimal and the firm should offer a pricing curve for multiple bundle sizes. We also study the BSP problem in the case of multiple customer types where customers draw their valuations from different distributions. In the case of identical marginal costs, we show that BSP no longer achieves perfect price discrimination due to the information rent but it can asymptotically achieve second-degree price discrimination.

Finally, we highlight some limitations of our framework that 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 rate is very fast. In this regard, our framework, while not being exact, can be viewed as a reasonable approximation to a setting where a multi-product firm sells a 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 to 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.

References
Tarek Abdallah. On the benefit (or cost) of large-scale bundling. Available at SSRN 2581170, 2015.


Will Ma and David Simchi-Levi. Reaping the benefits of bundling under high production costs. *Available at SSRN 2700394*, 2015.


Vahab Mirrokni and Hamid Nazerzadeh. Deals or no deals: Contract design for online advertising. *Available at SSRN 2693037*, 2015.


Ben Popper. The great unbundling: cable tv as we know it is dying, April 2015. *The Verge*, [Online; posted April 22, 2015].


Chris Welch. Verizon giving FiOS customers more flexibility over which channels they pay for, April 2015. The Verge, [Online; posted April 17, 2015].

Large-scale Bundle Size Pricing: 
A Theoretical Analysis

APPENDIX

Tarek Abdallah, Arash Asadpour, Josh Reed 
Leonard N. Stern School of Business, New York University, New York, NY 10012

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

A. Proofs of Section 4.

A.1. Proofs of Section 4.1.

In order to prove Theorem 1, we first need a preliminary result which will be useful for the remainder of the paper.

Lemma A1. For each \( m = 1, \ldots, M \), we have that \( \mathbb{P}\text{-}a.s., \bar{V}_{m,N} \to \bar{V} \) in \( D([0,1],\mathbb{R}) \) as \( N \to \infty \).
Moreover, \( \mathbb{E}[\bar{V}_{m,N}(t)] \to \bar{V}(t) \) for each \( t \in [0,1] \).

In order to prove Lemma A1, first for each \( m = 1, \ldots, M \), let
\[
\mathbb{F}_{X,m,N}(x) = \frac{1}{N} \sum_{n=1}^{N} 1\{X_{m,n} \leq x\}, \ x \geq 0,
\]
denote the empirical distribution function of \( \{X_{m,n}, n = 1, \ldots, N\} \), and recall by the Glivenko-Cantelli theorem (see Theorem 19.1 of Van der Vaart (2000)) that
\[
\sup_{x \geq 0} |\mathbb{F}_{X,m,N}(x) - F_X(x)| \to 0 \ \mathbb{P}\text{-}a.s. \ \text{as} \ N \to \infty. \quad \text{(A1)}
\]

Proof of Lemma A1: Let \( m \in \{1, \ldots, M\} \) and \( N \geq 1 \) and note that since \( \{X_{m,n}, n = 1, \ldots, N\} \) is i.i.d. with common distribution \( F_X \) and finite mean \( \mu < \infty \), it follows by the strong law of large numbers that \( \mathbb{P}\text{-}a.s., \)
\[
\bar{V}_{m,N}(1) \to \mu = \int_{0}^{1} F_X^{-1}(1-s)ds \ \text{as} \ N \to \infty, \quad \text{(A2)}
\]
where the equality follows by an appropriate change-of-variables. Now recall (see, for instance, Section 21 of Van der Vaart (2000)) that for each \( n = 1, \ldots, N \), we have \( X_{m,(n)} = F_X^{-1}(\bar{V}_{m,N}(t)) \) for \( t \in ((n-1)/N, n/N] \). Hence, using the definition of \( \bar{V}_{m,N} \) it follows after some algebra that
\[
\bar{V}_{m,N}(t) = \int_{0}^{t} F_X^{-1}_{X,m,N}(1-s)ds, \ 0 \leq t \leq 1.
\]
However, by (A1) and Lemma 21.2 of Van der Vaart (2000), we have that $\mathbb{P}$-a.s. $\mathbb{F}_{X,m,N}^{-1}(t) \to F_{X}^{-1}(t)$ for every $t \in (0, 1)$ where $F_{X}^{-1}(t)$ is continuous. Moreover, since $F_{X}^{-1}$ 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 $\mathbb{P}$-a.s.,

$$V_{m,N}(1) - \bar{V}_{m,N}(t) = \int_{t}^{1} \mathbb{F}_{X,m,N}^{-1}(1-s)ds \to \int_{t}^{1} F_{X}^{-1}(1-s)ds \text{ as } N \to \infty. \quad (A3)$$

Placing (A2) and (A3) together, we now obtain that for each $t \in [0, 1], \tilde{V}_{m,N}(t) \to \bar{V}(t), \mathbb{P}$-a.s., as $N \to \infty$. Moreover, since $\tilde{V}_{m,N}$ is a monotonic function and $\bar{V}$ is continuous, this implies that $\mathbb{P}$-a.s., $\tilde{V}_{m,N} \to \bar{V}$ in $D([0, 1], \mathbb{R})$ as $N \to \infty$. The proof that for each $t \in [0, 1]$, $\mathbb{E}[\tilde{V}_{m,N}(t)] \to \bar{V}(t)$ as $N \to \infty$, follows from Theorem 3 of Stigler (1974).

We now present the proof of Theorem 1.

**Proof of Theorem 1** Under the pricing policy described in the statement of the theorem, the firm’s normalized realized profit is given by

$$\bar{\pi}(p) = (\mathbb{E}[\tilde{V}_{1,N}(t^*)] - g(N)/N) \sum_{m=1}^{M} 1\{\tilde{V}_{m,N}(t^*) > \mathbb{E}[\tilde{V}_{1,N}(t^*)] - g(N)/N\}.$$

Since $g(N) \in o(N)$, it follows from Lemma A1 that $\mathbb{E}[\tilde{V}_{1,N}(t^*)] - g(N)/N \to \bar{V}(t^*)$ as $N \to \infty$. Moreover, using the fact that $t^* = 1 - F_X(c)$, it is straightforward to verify that $\bar{V}(t^*) = \mathbb{E}[(X_{1,1} - c)^+]$. Next, note that we may write

$$1\{\tilde{V}_{m,N}(t^*) > \mathbb{E}[\tilde{V}_{1,N}(t^*)] - g(N)/N\} = 1\{\tilde{V}_{m,N}(t^*) > -g(N)/\sqrt{N}\},$$

where $\tilde{V}_{m,N}(t^*) = \sqrt{N}(\tilde{V}_{m,N}(t^*) - \mathbb{E}[\tilde{V}_{1,N}(t^*)]).$ On the other hand, by Theorems 1 and 2 of Stigler (1974), we have that $\tilde{V}_{m,N}(t^*) \Rightarrow \bar{V}(t^*)$ as $N \to \infty$, where $\bar{V}(t^*)$ is normally distributed with a mean of zero and a finite variance, whose formula is given by (10) of Stigler (1974). Hence, since $g(N) \in \omega_+ (\sqrt{N})$, it follows that $1\{\tilde{V}_{m,N}(t^*) > \mathbb{E}[\tilde{V}_{1,N}(t^*)] - g(N)/N\} \Rightarrow 1$ as $N \to \infty$. Putting the above together, we obtain that $\mathbb{E}[\bar{\pi}(p)] \to M\mathbb{E}[(X_{1,1} - c)^+]$ as $N \to \infty$. The bound (9) now also implies that $\mathbb{E}[\bar{\pi}(p^*)] \to M\mathbb{E}[(X_{1,1} - c)^+]$ as $N \to \infty$, which completes the proof.

**A.2. Proofs of Section 4.2.**

In order to prove Theorem 2, we first need a preliminary result regarding the convergence of $\tilde{V}_{m,N}$.

For each $m = 1, \ldots, M$, let

$$\bar{c}_{m,N}(t) = \frac{1}{N} \cdot c_{m,N}(\lfloor Nt \rfloor) \text{ for } 0 \leq t \leq 1,$$

denote the normalized marginal cost for a bundle of size $\lfloor Nt \rfloor$ if chosen by customer $m$. 


Letting \( \bar{\pi}(p) = N^{-1}\pi(p) \), then it follows by the fact that \( \{\zeta(X_m, p) = n\} \subseteq \{p(n) < V_{m,N}(n)\} \) for each \( m = 1, \ldots, M \), and \( n = 1, \ldots, N \), together with (16), that we may write

\[
\bar{\pi}(p) \leq \sum_{m=0}^{M} \sup_{0 \leq t \leq 1} (\bar{V}_{m,N}(t) - \bar{c}_{m,N}(t)) \quad \text{for} \quad p \in (\mathbb{R}_+ \cup \infty)^N.
\]  

(A4)

Setting \( \bar{c}_{m,N} = (\bar{c}_{m,N}(t), 0 \leq t \leq 1) \), we now have the following result.

**Lemma A2.** Under Assumption 1 and for each \( m = 1, \ldots, M \), we have that \( \bar{c}_{m,N} \Rightarrow \bar{c} \) in \( D([0,1], \mathbb{R}) \) as \( N \to \infty \), where \( \bar{c}(t) = \bar{c}t \) for \( t \in [0,1] \). Moreover, \( \mathbb{E}[\bar{c}_{m,N}(t)] \to \bar{c}(t) \) for each \( t \in [0,1] \) as \( N \to \infty \).

**Proof of Lemma A2.** First note that for each \( m = 1, \ldots, M \), and \( n = 1, \ldots, N \), we have that \( \mathbb{E}[\bar{c}_{m,N}(n)] = n\bar{c}_N \). Therefore, in order to prove that \( \mathbb{E}[\bar{c}_{m,N}(t)] \to \bar{c}(t) \) for each \( 0 \leq t \leq 1 \), it suffices to prove that \( \bar{c}_N \to \bar{c} \). However, this follows by Assumption 1.

Next, we prove that for each \( m = 1, \ldots, M \), \( \bar{c}_{m,N} \Rightarrow \bar{c} \) in \( D([0,1], \mathbb{R}) \) as \( N \to \infty \). Since \( \bar{c}_{m,N} \) is a non-decreasing function and \( \bar{c}(t) \) is continuous, it suffices to show that \( \bar{c}_{m,N}(t) \Rightarrow \bar{c}(t) \) for each \( 0 \leq t \leq 1 \) as \( N \to \infty \). Moreover, since from the above we have that \( \mathbb{E}[\bar{c}_{m,N}(t)] \to \bar{c}(t) \) for each \( 0 \leq t \leq 1 \), it suffices to show that \( \text{Var}(\bar{c}_{m,N}(t)) \to 0 \).

Note that after some algebra we may write

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

where \( \tau_m \) is the permutation function that maps the order statistics to their corresponding item indices. Regarding the first term, recalling that \( \mathbb{P}(\tau_m(n) = k) = 1/N \) for \( k = 1, \ldots, N \), we have that

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

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

\[
\mathbb{P}(\tau_m(j) = \ell | \tau_m(i) = k) = 1/(N-1) \quad \text{for} \quad \ell = 1, \ldots, N \quad \text{with} \quad \ell \neq k,
\]

and so

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

Moreover, conditioning on \( \tau_m(1) \) we obtain that

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

Therefore,

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

Hence, \( \text{Var}(\bar{c}_{m,N}(t)) \to 0 \) as desired. \( \square \)
We now present the proof of Theorem 2.

Proof of Theorem 2. The proof is similar to the proof of Theorem 1 except now the denominator is no longer the asymptotic profit under perfect price discrimination.

Under the pricing policy described in the statement of the theorem, the firm’s normalized realized profit is

\[ \bar{\pi}(p) = \sum_{m=1}^{M} (E[\bar{V}_{1,N}(t^*)] - \bar{c}_{m,N}(t^*) - g(N)/N)1\{\bar{V}_{m,N}(t^*) > E[\bar{V}_{1,N}(t^*)] - g(N)/N\}. \]

Since \( g(N) \in o(N) \), it follows from Lemma A1 that \( E[\bar{V}_{1,N}(t^*)] - g(N)/N \rightarrow \tilde{V}(t^*) \) as \( N \rightarrow \infty \). Moreover, using the fact that \( t^* = 1 - F_X(\bar{c}) \), it is straightforward to verify that \( \tilde{V}(t^*) - t^*\bar{c} = E[(X_{1,1} - \bar{c})^+] \). Also, by Lemma A2, we have that \( E[\bar{c}_{m,N}(t^*)] \rightarrow t^*\bar{c} \) as \( N \rightarrow \infty \). Next, note that we may write

\[ 1\{\bar{V}_{m,N}(t^*) > E[\bar{V}_{1,N}(t^*)] - g(N)/N\} = 1\{\bar{V}_{m,N}(t^*) > -g(N)/\sqrt{N}\}, \]

where \( \bar{V}_{m,N}(t^*) = \sqrt{N}(\bar{V}_{m,N}(t^*) - E[\bar{V}_{1,N}(t^*)]) \). Moreover, by Theorems 1 and 2 of Stigler (1974), we have that \( \bar{V}_{m,N}(t^*) \Rightarrow \tilde{V}(t^*) \) as \( N \rightarrow \infty \), where \( \tilde{V}(t^*) \) is normally distributed with a mean of zero and a finite variance, whose formula is given by (10) of Stigler (1974). Hence, since \( g(N) \in \omega_+(\sqrt{N}) \), it follows that \( 1\{\bar{V}_{m,N}(t^*) > E[\bar{V}_{1,N}(t^*)] - g(N)/N\} \Rightarrow 1 \) as \( N \rightarrow \infty \). Putting the above together, we obtain that \( E[\bar{\pi}(p)] \rightarrow M\mathbb{E}[(X_{1,1} - \bar{c})^+] \) as \( N \rightarrow \infty \).

On the other hand, by Lemmas A1 and A2 and the continuous mapping theorem (see, for instance, Billingsley (1999)), it follows that for each \( m = 1, ..., M \),

\[ \sup_{0 \leq t \leq 1} (\bar{V}_{m,N}(t) - \bar{c}_{m,N}(t)) \Rightarrow \sup_{0 \leq t \leq 1} (\tilde{V}(t) - t\bar{c}) = \mathbb{E}[(X_{1,1} - \bar{c})^+], \]

where the final equality follows from the concavity of \( \tilde{V}(t) - t\bar{c} \) and solving for the first order condition. Hence, the bound (A4) now implies that \( \limsup E[\bar{\pi}(p^*)] \leq M\mathbb{E}[(X_{1,1} - \bar{c})^+] \), which completes the proof. \( \square \)

A.2.1. Proofs of Section 4.2.1.

Proof of Lemma 1. It follows from (17) that

\[ K\text{-BSP} = \max_{0 \leq t_1, t_2, ..., t_K \leq 1, 1 \leq k \leq K} \left[ \sum_{k=1}^{K} \alpha_k \left( \int_{0}^{t_k} F_X^{-1}(1-s)ds - c_k t_k \right) \right] \]

If we limit the feasibility set of this maximization problem by setting the constraints \( t_1 = t_2 = \cdots = t_K \), we have

\[ K\text{-BSP} \geq \max_{0 \leq t \leq 1} \left[ \sum_{k=1}^{K} \alpha_k \left( \int_{0}^{t} F_X^{-1}(1-s)ds - c_k t \right) \right] \]
\[
\begin{align*}
&= \max_{0 \leq t \leq 1} \left[ \left( \sum_{k=1}^{K} \alpha_k \right) \left( \int_0^t F_X^{-1}(1 - s)ds \right) - \left( \sum_{k=1}^{K} \alpha_k c_k \right) t \right] \\
&= \max_{0 \leq t \leq 1} \left[ \int_0^t F_X^{-1}(1 - s)ds - \bar{c}t \right],
\end{align*}
\]
where the latest equality follows from the definition of \( \bar{c} \) and the fact that \( \alpha_1 + \alpha_2 + \ldots + \alpha_K = 1 \).
Note that because of (18), the value of the final maximization problem is equal to \( \text{BSP} \), which completes the proof of the lemma. \( \square \)

**Proof of Proposition 1.** Note that the limiting normalized expected profit under perfect price discrimination is given by
\[
\text{PPD} = \lim_{N \to \infty} \frac{\sum_{n=1}^{N} \mathbb{E}[X_{1,n} - c(n)^+]}{N} = \sum_{k=1}^{K} \alpha_k \mathbb{E}[(X_{1,1} - c_k)^+] = \sum_{k=1}^{K} \left( \alpha_k \int_{c_k}^{1} (x - c_k) dF_X(x) \right).
\]
Using a change-of-variable \( x = F^{-1}(1 - s) \) and noting that \( t^*_k = 1 - F(c_k) \) for all \( 1 \leq k \leq K \), we then have
\[
\text{PPD} = \sum_{k=1}^{K} \alpha_k \left( \int_{1 - F_X(c_k)}^{0} - (F_X^{-1}(1 - s) - c_k) ds \right) = \sum_{k=1}^{K} \alpha_k \left( \int_{c_k}^{t^*_k} F_X^{-1}(1 - s)ds - c_k t^*_k \right),
\]
where the last equality follows from (17). \( \square \)

**A.2.2. Proofs of Section 4.2.2.** First, we provide a convexity result that is required to prove Proposition 2. We define a function \( H : \prod_{k=1}^{K} [0, \alpha_k] \to \mathbb{R} \) as follows,
\[
H(y_1, y_2, \ldots, y_K) = \left( \sum_{k=1}^{K} y_k \right) \left( \int_{0}^{1-F_X(c(y))} F_X^{-1}(1 - s)ds - \bar{c}(y) \cdot \left( 1 - F_X(\bar{c}(y)) \right) \right).
\]
Note that \( H(y_1, y_2, \ldots, y_K) \) is identical to the objective function of the optimization problem (20).

**Lemma A3.** For every vector \( (y_1, y_2, \ldots, y_K) \in \prod_{k=1}^{K} [0, \alpha_k] \), we have
\[
\frac{\partial^2}{\partial y^2_k} H(y_1, y_2, \ldots, y_K) > 0 \text{ for all } 1 \leq k \leq K.
\]

**Proof of Lemma A3.** By using the same change of variable as in the proof of Proposition 1, we can rewrite \( H(y_1, y_2, \ldots, y_K) \) as
\[
H(y_1, y_2, \ldots, y_K) = \left( \sum_{k=1}^{K} y_k \right) \left( \int_{\bar{c}(y)}^{1} (s - \bar{c}(y)) dF_X(s) \right).
\]

We remind the reader that for this section, we have assumed that $F_A$ exists. Let $\beta$ be such that $0 \leq \beta \leq 1$.

As a result, we obtain

$$\frac{\partial^2}{\partial y_k} H(y_1, y_2, \cdots, y_K) = \left( \sum_{k=1}^{K} y_k \right) \left( \frac{\partial}{\partial y_k} \tilde{c}(y) \right)^2 f_X(\tilde{c}(y)) \geq 0.$$

□

**Proof of Proposition 2.** 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 $\beta : \prod_{k=1}^{K}[0, \alpha_k] \rightarrow \mathbb{N}$ be function that maps any vector $y = (y_1, y_2, \cdots, y_K) \in \prod_{k=1}^{K}[0, \alpha_k]$ to a natural number where $\beta(y) = \{k : y_k \in \{0, \alpha_k\}\}$. 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. Consider the global optimal solution $y^* = (y_1^*, y_2^*, \cdots, y_K^*)$ with the highest binding number, i.e. with the largest $\beta(y^*)$. Since by our assumption $y^*$ is not a corner point, we have $\beta(y^*) < K$. Let $\hat{k}$ be such that $0 < y_{\hat{k}}^* < \alpha_{\hat{k}}$. Consider two solutions $\bar{y}^*$ and $\tilde{y}^*$ defined as follows:

$$y^* = (y_1^*, y_2^*, \cdots, y_{\hat{k}-1}^*, 0, y_{\hat{k}+1}^*, \cdots, y_K^*),$$

$$\bar{y}^* = (y_1^*, y_2^*, \cdots, y_{\hat{k}-1}^*, \alpha_{\hat{k}}, y_{\hat{k}+1}^*, \cdots, y_K^*).$$

Note that $y^* = (1 - y_{\hat{k}}^*/\alpha_{\hat{k}}) \bar{y}^* + (y_{\hat{k}}^*/\alpha_{\hat{k}}) \tilde{y}^*$. Due to the result of Lemma A3, we have

$$\frac{\partial^2}{\partial y_{\hat{k}}^2} H(y_1^*, y_2^*, \cdots, y_{\hat{k}-1}^*, y_{\hat{k}}, y_{\hat{k}+1}^*, \cdots, y_K^*) \geq 0.$$
Hence, $H(y^* ) \leq (1 - y^*_k / \alpha_k ) H(y^* ) + (y^*_k / \alpha_k ) H(\bar{y}^* )$, which means 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 is in contradiction 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 exist a global optimum solution with the nested-in-cost structure. The proof is again by contradiction. Suppose that no such solution exists. Then, for every corner point solution $y = (y_1, y_2, \cdots, y_K ) \in \prod_{k=1}^{K} \{0, \alpha_k \}$ that does not satisfy the nested property, we define the irregularity index of $y$ by the function $\gamma(y) : \prod_{k=1}^{K} [0, \alpha_k ] \mapsto \mathbb{N}$ which is defined by $\gamma(y) = \min \{ k : y_k = 0 \}$. Notice that since $y$ is 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 $\hat{k} = \gamma(y^* )$. We distinguish between two different cases.

Case I ($\alpha_l < \sum_{k=1}^{K} y^*_k$) Construct a new solution $\tilde{y}$ such that

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

ii. $\tilde{y}_k = \alpha_k$, for $\hat{k} < k \leq K$, and

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

iv. $\sum_{k=1}^{K} \tilde{y}_k = \left( \sum_{k=\hat{k}+1}^{K} y^*_k \right) - \alpha_l$.

Note that 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 $\bar{c}(\tilde{y}) \leq \bar{c}(y^*)$. Hence, by using (A6) 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 A3 similar to the first part of the proof to round all non-binding $y_k$’s for $k > \hat{k}$ to either 0 or $\alpha_k$, and obtain another global optimum solution $\bar{y}^*$ that is a corner point. However, the first $\hat{k}$ entries of $\bar{y}^*$ are non-zero, which means $\gamma(\bar{y}^* ) \geq \hat{k} + 1$. This is in contradiction with the assumption that $y^*$ is a global optimal solution with the highest irregularity index.

Case II ($\alpha_l \geq \sum_{k=1}^{K} y^*_k$) Consider a solution $\bar{y}$ such that

i. $\bar{y}_k = y^*_k = \alpha_k$, for $1 \leq k < \hat{k}$,

ii. $\bar{y}_k = \sum_{k=1}^{K} y^*_k$, and

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

Similar to the first case, we have $\sum_{k=1}^{K} \bar{y}_k = \sum_{k=0}^{K} y^*_k$ and $\bar{c}(\bar{y}) \leq \bar{c}(y^*)$. We use the convexity result of Lemma A3 to round $\bar{y}_k$ to either 0 or $\alpha_k$ and obtain a new global optimum solution $\tilde{y}$.

However, $\tilde{y}^*$ 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. □
A.3. Proofs of Section 4.3.

In order to prove Theorem 3, we need a number of technical results. First, recall from Section 4.1 the definition of the empirical distribution function of \( \{X_{m,n}; 1 \leq n \leq N\} \) for each \( m = 1, \ldots, M, \) that is given by

\[
F_{X,m,N}(x) = \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}\{X_{m,n} \leq x\}, \quad x \geq 0.
\]

Since \( \{X_{m,n}; 1 \leq n \leq N\} \) is no longer i.i.d., it does not immediately follow by the Glivenko-Cantelli theorem (see Theorem 19.1 of Van der Vaart (2000)) that (A1) holds. Nevertheless, we still have the following Lemma.

**Lemma A4.** Under Assumptions 1 and 2 and for each \( m = 1, \ldots, M, \) we have that \( \mathbb{P}\text{-a.s.,} \]

\[
\sup_{x \geq 0} |F_{X,m,N}(x) - F_X(x)| \to 0 \quad \text{as} \quad N \to \infty,
\]

where

\[
F_X(x) = \int_{\mathbb{R}_+} F_Z(x/c) dF_C(c), \quad x \geq 0.
\]

**Proof of Lemma A4.** Let \( m \in \{1, \ldots, M\} \) and 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 that \( \mathbb{P}\text{-a.s.,} \]

\[
|F_{X,m,N}(x) - F_X(x)| \to 0 \quad \text{as} \quad N \to \infty.
\]

Next, since by Assumption 2, \( X_{m,n} = c_n Z_{m,n} \) for each \( m = 1, \ldots, M, \) and \( n = 1, \ldots, N, \) it follows after some algebra that for each \( x \geq 0 \) we may write

\[
F_{X,m,N}(x) = \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}\{Z_{m,n} \leq x/c(n)\} = \frac{1}{N} \Phi(N,x) + \int_{0}^{\infty} F_Z(x/c) dF_C(c) \tag{A7}
\]

where

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

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

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

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

\[
\int_{0}^{\infty} F_Z(x/c)(1 - F_Z(x/c)) dF_C(c) \to \int_{0}^{\infty} F_Z(x/c)(1 - F_Z(x/c)) dF_C(c) \quad \text{as} \quad N \to \infty,
\]
and so by the strong law of large numbers for martingales (see Section 2.6 of Liptser and Shiryayev (2012)), it follows that \( P \)-a.s., \((1/N)\Phi(N,x) \to 0 \) as \( N \to \infty \).

In addition, it follows from Assumption 1 and the assumption that \( F_Z \) is continuous, that

\[
\int_0^\infty F_Z(x/c)d\mathbb{P}_{C,N}(c) \to \int_0^\infty F_Z(x/c)d\mathbb{P}_C(c) \quad \text{as} \quad N \to \infty.
\]

Now the result immediately follows from the decomposition (A7). \( \square \)

Next, we have the following result regarding the convergence of the normalized valuation function for each customer \( m = 1, \ldots, M \). Its statement is the same as Lemma A4, but its proof is slightly different due to the more general setup of the present section.

**Lemma A5.** Under Assumptions 1 and 2 and for each \( m = 1, \ldots, M \), we have that \( P \)-a.s.,

\[
\bar{V}_{m,N}(1) = \frac{1}{N} \sum_{n=1}^N X_{m,n} = \frac{1}{N} \sum_{n=1}^N c(n)Z_{m,n}.
\]

Moreover, we have the decomposition

\[
\frac{1}{N} \sum_{n=1}^N c(n)Z_{m,n} = \frac{1}{N} \Phi(N) + \mathbb{E}[Z_{1,1}] \int_0^\infty c^2d\mathbb{P}_{C,N}(c),
\]

where

\[
\Phi(N) = \sum_{n=1}^N c(n)(Z_{m,n} - \mathbb{E}[Z_{1,1}]).
\]

However, by Assumption 2 it is clear that the process \( \Phi = \{\Phi(N), N \geq 1\} \) is a square-integrable martingale with predictable quadratic variation

\[
\langle \Phi \rangle_N = \sigma^2 \sum_{n=1}^N c(n)^2.
\]

On the other hand, by Assumption 1 and since \( \{c(n), n \geq 1\} \) is uniformly bounded, it follows that

\[
\frac{\sigma^2}{N} \sum_{n=1}^N c(n)^2 = \sigma^2 \int_0^\infty c^2d\mathbb{P}_{C,N}(c) \to \sigma^2 \int_0^\infty c^2d\mathbb{P}_C(c) \quad \text{as} \quad N \to \infty,
\]

and so by the strong law of large numbers for martingales (see Section 2.6 of Liptser and Shiryayev (2012)), it follows that \( P \)-a.s., \((1/N)\Phi(N) \to 0 \) as \( N \to \infty \). It now follows by the decompositions (A8) and (A9) along with Assumption 1 that \( P \)-a.s.,

\[
\bar{V}_{m,N}(1) \to \mathbb{E}[Z_{1,1}]c(1) \quad \text{as} \quad N \to \infty.
\]
Now recall (see, for instance, Section 21 of Van der Vaart (2000)) that for each \( n = 1, \ldots, N \), we have \( X_{m,n} = \mathbb{F}^{-1}_{X,m,N}(t) \) for \( t \in ((n - 1)/N, n/N] \). Hence, using the definition of \( \bar{V}_{m,N} \), it follows after some algebra that

\[
\bar{V}_{m,N}(t) = \int_{0}^{t} \mathbb{F}^{-1}_{X,m,N}(1 - s) ds, \quad 0 \leq t \leq 1.
\]

However, by Lemma A4 above and Lemma 21.2 of Van der Vaart (2000), \( \mathbb{P}\)-a.s. we have that \( \mathbb{F}^{-1}_{X,m,N}(t) \to F^{-1}_{X}(t) \) for every \( t \in (0, 1) \). Moreover, by the bounded convergence theorem (see Lieb and Loss (2001)), for each \( t > 0 \), \( \mathbb{P}\)-a.s.,

\[
\bar{V}_{m,N}(1) - \bar{V}_{m,N}(t) = \int_{t}^{1} \mathbb{F}^{-1}_{X,m,N}(1 - s) ds \to \int_{t}^{1} F^{-1}_{X}(1 - s) ds \quad \text{as} \quad N \to \infty. \tag{A11}
\]

Placing (A10) and (A11) together, we now obtain that for each \( t \in [0, 1] \), \( \bar{V}_{m,N}(t) \to \bar{V}(t) \), \( \mathbb{P}\)-a.s., as \( N \to \infty \). However, since \( \bar{V}_{m,N} \) is monotonic and \( \bar{V} \) is continuous, this implies that \( \mathbb{P}\)-a.s., \( \bar{V}_{m,N} \to \bar{V} \) in \( D([0, 1], \mathbb{R}) \) as \( N \to \infty \). The proof that for each \( t \in [0, 1] \), \( \mathbb{E}[\bar{V}_{m,N}(t)] \to \bar{V}(t) \) as \( N \to \infty \), follows from a generalization of Theorem 3 of Stigler (1974) to the non-identical case (see also the discussion following Theorem 6 of Stigler (1974)).

We next we have the following generalization of Lemma A2 of Section 4.2.

**Lemma A6.** Under Assumptions 1 and 2, and for each \( m = 1, \ldots, M \), we have that \( \bar{c}_{m,N} \Rightarrow \bar{c} \) in \( D([0, 1], \mathbb{R}) \) as \( N \to \infty \), where

\[
\bar{c}(t) = \int_{0}^{\infty} c(1 - F_{Z}(F^{-1}_{X}(1 - t)/c)) dF_{C}(c) \quad \text{for} \quad 0 \leq t \leq 1. \tag{A12}
\]

**Proof of Lemma A6.** Let \( m \in \{1, \ldots, M\} \) and note that since \( \bar{c}_{m,N}(\cdot) \) is a non-decreasing function while \( \bar{c}(t) \) is continuous, it suffices to show that \( \bar{c}_{m,N}(t) \Rightarrow \bar{c}(t) \) for each \( t \in [0, 1] \) as \( N \to \infty \). Next, recall that as in (15) we may write

\[
\bar{c}_{m,N}(t) = \sum_{n=0}^{\lfloor Nt \rfloor - 1} c_{\tau_{m}(N-n)}, \quad \text{for} \quad t \geq 0. \tag{A13}
\]

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.,

\[
\sum_{n=0}^{\lfloor Nt \rfloor - 1} c_{\tau_{m}(N-n)} = \sum_{n=1}^{N} c(n) \{X_{m,n} > X_{m,(N-\lfloor Nt \rfloor + 1)}\},
\]

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

\[
\bar{\varepsilon}_{m}(t) = \frac{1}{N} \sum_{n=1}^{N} c(n) (1\{X_{m,n} > X_{m,(N-\lfloor Nt \rfloor + 1)}\} - 1\{X_{m,n} > F_{X}^{-1}(1 - t)\}), \quad t \in [0, 1],
\]
and

\[ \tilde{\delta}_m(t) = \frac{1}{N} \sum_{n=1}^{N} c(n)(1\{X_{m,n} > F_X^{-1}(1-t)\} - (1 - F_Z(F_X^{-1}(1-t)/c(n))))), \ t \in [0,1]. \]

It then follows that we may write

\[ \tilde{c}_{m,N}(t) = \frac{1}{N} \sum_{n=1}^{N} c(n)(1 - F_Z(F_X^{-1}(1-t)/c(n))) + \tilde{c}_{m}(t) + \tilde{\delta}_m(t), \ t \in [0,1]. \]

Moreover, by Assumption 1 and using similar techniques to those used in the proof of Lemma A4, one may show that

\[ \frac{1}{N} \sum_{n=1}^{[Nt]} c(n)(1 - F_Z(F_X^{-1}(1-t)/c(n))) \to \int_0^\infty c(1 - F_Z(F_X^{-1}(1-t)/c))dF_C(c) \text{ as } N \to \infty. \]

Hence, in order to complete the proof it suffices now to show that \( \tilde{c}_{m}(t) \Rightarrow 0 \) and \( \tilde{\delta}_m(t) \Rightarrow 0 \) as \( N \to \infty \).

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

\[ E[|\tilde{c}_{m}(t)|| \leq \frac{1}{N} \sum_{n=1}^{N} c(n)E[1\{X_{m,n} > X_{m,(N-[Nt]+1)}\} - 1\{X_{m,n} > F_X^{-1}(1-t)\}]. \tag{A14} \]

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

\[ |1\{X_{m,n} > X_{m,(N-[Nt]+1)}\} - 1\{X_{m,n} > F_X^{-1}(1-t)\}| \]

\[ = 1\{\min(X_{m,(N-[Nt]+1)}, F_X^{-1}(1-t)) < X_{m,n} < \max(X_{m,(N-[Nt]+1)}, F_X^{-1}(1-t))\} \]

Therefore, noting that \( X_{m,n} = c(n)Z_{m,n} \) and substituting into \( \text{(A14)} \), we obtain

\[ E[|\tilde{c}_{m}(t)|| \]

\[ \leq \frac{1}{N} \sum_{n=1}^{N} c(n)E[1\{\min(F_X^{-1}_{X,m,N}(1-t), F_X^{-1}(1-t)) < c(n)Z_{m,n} < \max(F_X^{-1}_{X,m,N}(1-t), F_X^{-1}(1-t))\}]] \]

\[ = \int_0^\infty c \left[ F_Z(\max(F_X^{-1}_{X,m,N}(1-t), F_X^{-1}(1-t))/c) - F_Z(\min(F_X^{-1}_{X,m,N}(1-t), F_X^{-1}(1-t))/c) \right]dF_C(c). \]

However, from Lemma A4 and the fact that \( F_X \) is continuous, we get that \( F_X^{-1}_{X,m,N}(1-t) \Rightarrow F_X^{-1}(1-t) \) for \( t \in (0,1) \) as \( N \to \infty \). Therefore, by the continuous mapping theorem and for every \( c > 0 \), we get

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

Regarding \( \bar{\delta}_m(t) \), note that since \( \{X_{m,n}, 1 \leq n \leq N\} \) is i.i.d. it is straightforward to verify that \( \mathbb{E}[\bar{\delta}_m(t)] = 0 \). Moreover, using Assumption 1 we have that
\[
\text{Var}(\bar{\delta}_m(t)) = \frac{1}{N^2} \sum_{n=1}^{N} c(n)^2 \to 0 \quad \text{as} \quad N \to \infty.
\]
Hence, it follows that \( \bar{\delta}_m(t) \to 0 \) as \( N \to \infty \). \( \Box \)

We are now ready to present the proof of Theorem 3.

**Proof of Theorem 3.** Using Lemmas A5 and A6, the proof is similar to that of Theorem 2. The one exception is that we need to show a central limit theorem convergence for \( \tilde{V}_{m,N}(t^*) \) for every \( m = 1, \ldots, M \), under the current general setting. However, from Assumption 2 and for \( m = 1, \ldots, M \), we have that \( \{X_{m,n}, n = 1, \ldots, N\} \) is a collection of independent random variables. Therefore, in order to show that \( \bar{V}_{m,N}(t^*) \) is normally distributed with mean zero and a finite variance, it suffices to verify that the four conditions of Theorem 6 of Stigler (1974) are satisfied.

First, we have that \( F_{X_m,n}(x) = F_Z(x/c(n)) \) for every \( x \in \mathbb{R} \). Now, define a new random variable \( Y \) such that \( Y = K Z_{1,1} \) where \( 0 < K < \infty \) is the upper bound on the marginal costs \( \{c(n), n \geq 1\} \). Then, we have that \( F_Y(x) = F_Z(x/K) \) for every \( x \in \mathbb{R} \). Therefore, since the \( X_{m,n}s^t \) are non-negative, we have that \( F_Y(x) = F_{X_m,n}(x) = 0 \) for \( x \leq 0 \) and \( n = 1, \ldots, N \). Meanwhile, for \( x \geq 0 \) and \( n = 1, \ldots, N \), we have \( F_Y(x) = F_Z(x/K) \leq F_{X_m,n}(x) \) where the last inequality follows from the fact that \( K \geq \sup\{c_n, n \geq 1\} \). Therefore, the first condition of Theorem 6 of Stigler (1974) is satisfied.

Next, we have that
\[
\frac{1}{N} \sum_{n=1}^{N} F_{X_m,n}(x) = \frac{1}{N} \sum_{n=1}^{N} F_Z(x/c(n)) = \int_{0}^{\infty} F_Z(x/c) d\mathbb{F}_{C,N}, \quad x \geq 0.
\]
Therefore, from Assumption 1 and the fact that \( F_Z \) is continuous, we obtain that
\[
\frac{1}{N} \sum_{n=1}^{N} F_{X_m,n}(x) \to \int_{0}^{\infty} F_Z(x/c) d\mathbb{F}_C \quad \text{as} \quad N \to \infty.
\]
Hence, the second condition of Theorem 6 of Stigler (1974) is satisfied.

Likewise, for any \( x, y \in \mathbb{R} \) we have that
\[
\frac{1}{N} \sum_{n=1}^{N} (F_{X_m,n}(\min(x,y)) - F_{X_m,n}(x) F_{X_m,n}(y))
\to \int_{0}^{\infty} (F_Z(\min(x,y)/c) - F_Z(x/c) F_Z(y/c)) d\mathbb{F}_C \quad \text{as} \quad N \to \infty,
\]
\[
= F_X(\min(x,y)) - F_X(\max(x,y)).
\]
where $F_X(\min(x,y)) - F_X(\max(x,y)) > 0$ for an appropriate range of $x, y \in \mathbb{R}_+$. Therefore, the third and fourth condition of Theorem 6 of Stigler (1974) are satisfied. As a result, the asymptotic normality for $\bar{V}(t^*)$ follows from Theorem 6 of Stigler (1974). The rest of the proof follows similarly to the proof of Theorem 2 and hence is omitted. □

Proof of Proposition 3. We first prove the if part. Consider the following bundle size $t = 1 - F_X(z,c_l)$. Since $z_c \leq z_c l$, then we have that $z_l \leq z_c l/c \leq z_c l/c \leq z_c$ for every $c \in [c_l, c_u]$ where $F_Z(z_l) = F_X(z_c l) = F_Z(1)$. Hence, it can be shown that $u = F_X^{-1}(F_X(z_c l)) = z_c$. It now follows from (24) that the limiting normalized profit under the proposed pricing policy is given by

$$\bar{V}(t) - c(t) = \int_{c_l}^{c_u} \int_{z_c l/c}^\infty c f_Z(z) (z - 1) dz dF(c) = \mathbb{E}[(Z_{1,1} - 1)^+ c(1)],$$

where the second equality is due to the fact that for $c \in [c_l, c_u]$ we have $z_l \leq z_c l/c \leq z_c$. Finally, since $z_l \leq z_c l/c \leq z_c$ for every $c \in [c_l, c_u]$, then it can be shown that $F_X(z_c l) = F_Z(1)$ and hence $t = 1 - F_Z(1)$ 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 $t^*$ but $z_c l > z_c l$. Hence, we get that $z_l < 1 < z_u$ where $F_Z(z_l) = F_z(z_u) = F_Z(1)$. Let $u = F_X^{-1}(1 - t^*)$. Next, we consider three different cases for the values of $u$.

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

$$\bar{V}(t) - c(t) \leq \mathbb{E}[(Z_{1,1} - 1)^+ \tilde{c}(1)] + \int_{\tilde{c}}^{c_u} \int_{z_c l/c}^{z_l} c f_Z(z) (z - 1) dz dF(c).$$

However, since $u/c < z_l$ for $c \in [\tilde{c}, c_u]$, we get that $\int_{\tilde{c}}^{c_u} \int_{z_c l/c}^{z_l} c f_Z(z) (z - 1) dz dF(c) < 0$ which is a contradiction.

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

$$\bar{V}(t) - c(t) \leq \mathbb{E}[(Z_{1,1} - 1)^+ \tilde{c}(1)] - \int_{\tilde{c}}^{c_u} \int_{z_c l/c}^{u/c} c f_Z(z) (z - 1) dz dF(c).$$

However, since $u/c > z_u$ for $c \in [c_l, \tilde{c})$, we get that $\int_{\tilde{c}}^{c_u} \int_{z_c l/c}^{u/c} c f_Z(z) (z - 1) dz dF(c) > 0$ which is a contradiction.

Case III ($z_c l > u > z_c l$) We have that $u/c > z_l$ and $u/c > z_u$. Consequently there exists $c_1, c_2 \in (c_l, c_u)$ where $c_1 < c_2$ such that

- $u/c > z_u$ for any $c \in [c_l, c_1)$,
- $z_l \leq u/c \leq z_u$ for any $c \in [c_1, c_2]$,
• $u_c/c < z_l$ for any $c \in (\tilde{c}_2, c_u]$.

After some algebra, the limiting BSP profit can be written as

$$V(t) - \bar{c}(t) = \mathbb{E}[\{Z_{1,1} - 1\}^+\bar{c}(1) - \int_{e_1}^{c_1} \int_{z_u}^{u/c} cf_Z(z)(z-1)dzdF_C(c) + \int_{e_2}^{c_u} \int_{u/c}^1 cf_Z(z)(z-1)dzdF_C(c).$$

However, since $u_c/c > z_u$ for any $c \in [e_1, e_1)$ and $u_c/c < z_l$ for any $c \in (\tilde{c}_2, c_u]$, then

$$\int_{e_1}^{c_1} \int_{z_u}^{u/c} cf_Z(z)(z-1)dzdF_C(c) + \int_{e_2}^{c_u} \int_{u/c}^1 cf_Z(z)(z-1)dzdF_C(c) < 0,$

which is a contradiction. □

B. Proofs of Section 5.

B.1. Proofs of Section 5.1.

Proof of Proposition 4. Under the pricing policy described in the statement of the theorem, the firm’s normalized realized profit, which is denoted by $\bar{\pi}(p, b_N) = N^{-1}\pi(p, b_N)$, is given by

$$\bar{\pi}(p, b_N) = \left(\mathbb{E}[\tilde{V}_{1,N}(t^*) \wedge \tilde{b}] - g(N)/N - (t^* \wedge t_b)c\right) \sum_{m=1}^{M} 1\{\tilde{V}_{m,N}(t^*, \tilde{b}) > \mathbb{E}[\tilde{V}_{1,N}(t^*) \wedge \tilde{b}] - g(N)/N\}.$$

Since $g(N) \in o(N)$, it follows from Lemma A1 that $\mathbb{E}[\tilde{V}_{1,N}(t^*) \wedge \tilde{b}] - g(N)/N - (t^* \wedge t_b)c \to \tilde{V}(t^* \wedge t_b) - (t^* \wedge t_b)c$ as $N \to \infty$.

Next, note that we may write

$$1\{\tilde{V}_{m,N}(t^*) \wedge \tilde{b} > \mathbb{E}[\tilde{V}_{1,N}(t^*) \wedge \tilde{b}] - g(N)/N\} = 1\{\tilde{V}_{m,N}(t^*, \tilde{b}) > -g(N)/\sqrt{N}\},$$

where $\tilde{V}_{m,N}(t^*, \tilde{b}) = \sqrt{N}(\tilde{V}_{m,N}(t^*) \wedge \tilde{b}) - \mathbb{E}[\tilde{V}_{1,N}(t^*) \wedge \tilde{b})].$ Moreover, by Theorems 1 and 2 of Stigler (1974) and the continuous mapping theorem, we have that $\tilde{V}_{m,N}(t^*, \tilde{b})$ converges in distribution as $N \to \infty$ to a truncated normal random variable with a mean of zero and a finite variance. Hence, since $g(N) \in \omega_+(\sqrt{N})$, it follows that $1\{\tilde{V}_{m,N}(t^*, \tilde{b}) > -g(N)/\sqrt{N}\} \Rightarrow 1$ as $N \to \infty$. Putting the above together, we obtain that $\mathbb{E}[\bar{\pi}(p, b_N)] \to M (\tilde{V}(t^* \wedge t_b) - (t^* \wedge t_b)c)$ as $N \to \infty$.

We are now left to show that

$$E\left[\sum_{m=1}^{M} \sup_{0 \leq t \leq 1} (\tilde{V}_{m,N}(t) \wedge \tilde{b} - tc)\right] \to M (\tilde{V}(t^* \wedge t_b) - (t^* \wedge t_b)c) \quad \text{as } N \to \infty.$$

However, by Lemma 1 and the continuous mapping theorem, we have that

$$\sum_{m=1}^{M} \sup_{0 \leq t \leq 1} (\tilde{V}_{m,N}(t) \wedge \tilde{b} - tc) \Rightarrow M \sup_{0 \leq t \leq 1} (\tilde{V}(t) \wedge \tilde{b} - tc) \quad \text{as } N \to \infty.$$
Clearly, if $\tilde{V}(t^*) \leq \tilde{b}$ or equivalently if $t^* \leq \tilde{b}$, then $\sup_{0 \leq t \leq 1} (\tilde{V}(t) \wedge \tilde{b} - tc) = (\tilde{V}(t^*) - t^*c) = \mathbb{E}[(X_{1,1} - c)^+]$. Meanwhile, if $\tilde{V}(t^*) > \tilde{b}$ or equivalently $t^* > \tilde{b}$ then, since $\tilde{V}(t)$ is a concave non-decreasing function, we have that $\sup_{0 \leq t \leq 1} (\tilde{V}(t) \wedge \tilde{b} - tc) = (\tilde{V}(\tilde{b}) - \tilde{b}c)$. Therefore, combining both cases, it follows from Lemma A1 that

$$\sum_{m=1}^{M} \sup_{0 \leq t \leq 1} (\tilde{V}_{m,N}(t) \wedge \tilde{b} - tc) \Rightarrow M (\tilde{V}(t^* \wedge \tilde{b}_m) - (t^* \wedge \tilde{b}_m)c) \quad \text{as } N \to \infty.$$ 

Again by Lemma A1 and the bounded convergence theorem, it follows that the above convergence also holds in expectation which concludes our proof. □

B.2. Proofs of Section 5.2.

Before presenting the proof of Theorem 4, we first need to show that, under the suggested pricing curve, it is asymptotically incentive compatible for each customer to purchase the highest affordable budget size i.e. $t^* \wedge \tilde{b}_m$. For this reason, we introduce a new notation regarding the consumer surplus. For $m = 1, \ldots, M$ and $0 \leq t \leq t^*$, we define the scaled normalized surplus under the proposed pricing policy by

$$Q_{m,N}(p,t) = \left(\tilde{V}_{m,N}(t) + N^{-1/2} h(t)g(N)\right) / \left(N^{-1/2} h(t^* \wedge \tilde{b}_m)g(N)\right),$$

where $\tilde{V}_{m,N}(t) = N^{1/2}(\tilde{V}_{m,N}(t) - \mathbb{E}[\tilde{V}_{1,N}(t)])$. In words, $Q_{m,N}(p,t)$ is the realized surplus of customer $m$ normalized by $N^{1/2}$ and scaled by $N^{-1/2} h(t^* \wedge \tilde{b}_m)g(N)$. The scaling is for a technical reason to ensure that the limit is well-defined.

We have that customer $m$ will purchase a bundle of size

$$\zeta(X_m,p,b_m) \in \arg \max_{\{n \in \Theta | Np(n) \leq b_m\}} (V_{m,N}(n) - p(n)),$$

where for convenience we assume that the customer breaks ties by choosing the smallest bundle size in the set of maximizers. Letting $\bar{p}(t) = N^{-1} p(\lfloor Nt \rfloor)$ and $\bar{\zeta}(X_m,\bar{p},\bar{b}_m) = N^{-1}\zeta(X_m,p,b_m)$, we get

$$\bar{\zeta}(X_m,\bar{p},\bar{b}_m) \in \arg \max_{\{t \in [0,1] \mid \bar{p}(t) \leq \bar{b}_m\}} (\tilde{V}_{m,N}(t) - \bar{p}(t)),$$

where now we assume that $\bar{\zeta}(X_m,\bar{p},\bar{b}_m)$ corresponds the infimum over the set of optimizers. Notice that since $\tilde{V}_{m,N}(t)$ is right-continuous then the infimum is attained. Now notice that we can equivalently write $\bar{\zeta}(X_m,\bar{p},\bar{b}_m)$ as the optimizer of the scaled normalized surplus $Q_{m,N}(p,t)$. In particular, we have

$$\bar{\zeta}(X_m,\bar{p},\bar{b}_m) \in \arg \max_{\{t \in [0,1] \mid \bar{p}(t) \leq \bar{b}_m\}} Q_{m,N}(p,t).$$
As an intermediate step, we want to show that it is asymptotically incentive compatible for each customer to purchase the highest affordable size. More specifically, we want to show that for each \( m = 1, \ldots, M \) and under the proposed pricing policy \( p \), we get that \( \tilde{\zeta}(X_m, \bar{p}, \bar{b}_m) \Rightarrow t^* \wedge t_{b_m} \) as \( N \to \infty \). However, before doing so we need the following preliminary technical lemma regarding the uniform convergence in probability of \( Q_{m,N}(p,t) \).

**Lemma B1.** Under the proposed policy, we have that

\[
\sup_{\{t \in [0,1]: \bar{p}(t) \leq \bar{b}_m\}} |Q_{m,N}(p,t) - Q_m(p,t)| \Rightarrow 0 \quad \text{as } N \to \infty,
\]

where \( Q_m(p,t) = h(t) / h(t^* \wedge t_{b_m}) \).

**Proof of Lemma B1.** First notice that we have the following bound on the absolute difference

\[
\sup_{\{t \in [0,1]: \bar{p}(t) \leq \bar{b}_m\}} |Q_{m,N}(p,t) - Q_m(p,t)| \leq \sup_{\{t \in [0,1]: \bar{p}(t) \leq \bar{b}_m\}} \left| \frac{\check{V}_{m,N}(t)}{(N^{-1/2}h(t^* \wedge t_{b_m})g(N))} \right| + \sup_{\{t \in [0,1]: \bar{p}(t) \leq \bar{b}_m\}} \left| \frac{h(t)}{h(t^* \wedge t_{b_m})} - \frac{h(t)}{h(t^* \wedge t_{b_m})} \right|
\]

\[
= \sup_{\{t \in [0,1]: \bar{p}(t) \leq \bar{b}_m\}} \left| \frac{\check{V}_{m,N}(t)}{(N^{-1/2}h(t^* \wedge t_{b_m})g(N))} \right| \quad (B1)
\]

From Theorems 1 and 2 of Stigler (1974), we have that \( \check{V}_{m,N}(t) \) converges in distribution to a normal random variable with zero mean and finite variance. Moreover, since \( g(N) \in \omega_{+}(N^{1/2}) \) then \( N^{-1/2}h(t^* \wedge t_{b_m})g(N) \to \infty \). Hence, it follows that \( \check{V}_{m,N}(t) / (N^{-1/2}h(t^* \wedge t_{b_m})g(N)) \Rightarrow 0 \) as \( N \to \infty \). Finally, since \( \check{V}_{m,N}(t) \) is a non-decreasing function, it follows that the convergence is uniform which concludes the proof. \( \square \)

We now have the following lemma regarding the convergence of \( \tilde{\zeta}(X_m, \bar{p}, \bar{b}_m) \) for \( m = 1, \ldots, M \).

**Lemma B2.** Under the proposed pricing policy, we have that

\[
\tilde{\zeta}(X_m, \bar{p}, \bar{b}_m) \Rightarrow t^* \wedge t_{b_m} \quad \text{as } N \to \infty.
\]

**Proof of Lemma B2.** Notice that since the limiting consumer surplus \( Q_m(p,t) \) is strictly increasing in \( t \) for \( 0 \leq t \leq t^* \wedge t_{b_m} \), we have that

\[
t^* \wedge t_{b_m} = \arg\max_{\{t \in [0,t^*]: \bar{p}(t) \leq \bar{b}_m\}} Q_m(p,t).
\]

For any \( \epsilon > 0 \), let \( \epsilon = Q_m(p,t^* \wedge t_{b_m}) - \sup_{\{t \in [0,t^*]: |t - t^* \wedge t_{b_m}| > \delta\}} Q_m(p,t^* \wedge t_{b_m}) > 0 \). From Lemma B1, we have that there exists a sequence of events \( S_N \) such that \( \mathbb{P}(S_N) \to 1 \) as \( N \to \infty \) and the following holds on \( S_N \)

\[
Q_{m,N}(p,\tilde{\zeta}(X_m, \bar{p}, \bar{b}_m)) - Q_m(\tilde{\zeta}(X_m, \bar{p}, \bar{b}_m)) < \epsilon / 2 \quad \text{and} \quad Q_m(p,t^* \wedge t_{b_m}) - Q_{m,N}(p,t^* \wedge t_{b_m}) < \epsilon / 2.
\]
Hence, on $S_N$, we have the following chain of inequalities

$$Q_m(p, \tilde{\zeta}(X_m, \tilde{p}, \tilde{b}_m)) + \epsilon/2 > Q_{m,N}(p, \tilde{\zeta}(X_m, \tilde{p}, \tilde{b}_m)) \geq Q_{m,N}(p, t^* \wedge t_{b_m}) > Q_m(p, t^* \wedge t_{b_m}) - \epsilon/2.$$  

After rearranging the above, we get

$$Q_m(p, t^* \wedge t_{b_m}) - Q_m(p, \tilde{\zeta}(X_m, \tilde{p}, \tilde{b}_m)) < \epsilon. \quad (B2)$$

Now recalling the definition of $\epsilon = Q_m(p, t^* \wedge t_{b_m}) - \sup_{t \in [0, t^*]} \{ |\tilde{\zeta}(X_m, \tilde{p}, \tilde{b}_m) - t^* \wedge t_{b_m}| \} Q_m(p, t^* \wedge t_{b_m})$, it follows that whenever (B2) holds then $|\tilde{\zeta}(X_m, \tilde{p}, \tilde{b}_m) - t^* \wedge t_{b_m}| \leq \delta$. Finally, since (B2) holds on $S_N$, then, for any $\delta > 0$, we have that $\mathbb{P}(|\tilde{\zeta}(X_m, \tilde{p}, \tilde{b}_m) - t^* \wedge t_{b_m}| \leq \delta) \leq \mathbb{P}(S_N) \to 1$ as $N \to \infty$ which concludes the proof. □

We are now ready to present the proof of the theorem.

**Proof of Theorem 4.** Under the pricing policy described in the statement of the Proposition, the firm’s normalized realized profit, which is denoted by $\tilde{\pi}(p, \{b_m\}_{m=1}^M) = N^{-1} \pi(p, \{b_m\}_{m=1}^M)$, is given by

$$\tilde{\pi}(p, \{b_m\}_{m=1}^M) = \sum_{m=1}^M \sum_{\{t: Nt \in \{0, 1, \ldots, \lceil Nt^* \rceil \}} 1\{\tilde{\zeta}_m(X_m, \tilde{p}, \tilde{b}_m) = t\} \left( \mathbb{E}[\tilde{V}_{1,N}(t)] - h(t)g(N)/N - tc \right).$$

Since $g(N) \in o(N)$ and $h(t) \in \mathbb{R}_+$, it follows from Lemma A1 that $\mathbb{E}[\tilde{V}_{1,N}(t)] - h(t)g(N)/N - tc \to \tilde{V}(t) - tc$ as $N \to \infty$. Therefore, using Lemma B2 and the dominated convergence theorem, we get that

$$\mathbb{E}[\tilde{\pi}(p, \{b_m\}_{m=1}^M)] \to \sum_{m=1}^M \left( \tilde{V}(t^* \wedge t_{b_m}) + (t^* \wedge t_{b_m})c \right).$$

Finally, similar to the proof in Proposition 4, we have that

$$N^{-1} \sum_{m=1}^M \mathbb{E}[\sup_{n=0, 1, \ldots, N} (V_{m,N}(n, b_m) - nc)] \to \sum_{m=1}^M \left( \tilde{V}(t^* \wedge t_{b_m}) + (t^* \wedge t_{b_m})c \right) \text{ as } N \to \infty,$$

which concludes the proof. □

**C. Proofs of Section 6.**

Before presenting the proof Theorem 5, we need to prove a number of intermediate results. We will split the proof into two parts. In the first part, we show that the normalized denominator converges to some positive real number as $N \to \infty$. Then, we show that the normalized expected BSP profit under the proposed policy also converges to the same number.
C.1. Asymptotic Analysis of $\mathbb{E}[\bar{Z}_{1,N}(X^H_1, X^L_1)]$.

First let $\bar{\zeta}(X_m, p) = N^{-1}\zeta(X_m, p)$ and $\bar{p}(t) = N^{-1}p([Nt])$, we have that

$$
\bar{\zeta}(X_m, p) \in \arg \max_{t \in [0, 1]} (\bar{V}_{m,N}(t) - p(t)) ,
$$

where for convenience we assume that the customer break ties by choosing the infimum of all maximizers.

Now, let $\bar{Z}_{m,N}(X^H_m, X^L_m) = N^{-1}Z_{m,N}(X^H_m, X^L_m)$ denote the normalized profit under second-degree price discrimination. For $m = 1, \ldots, M$, we have that

$$
\bar{Z}_{m,N}(X^H_m, X^L_m) = \sup_{\bar{p} \in (\mathbb{R}_+ \cup \{\infty\})^N} \left\{ \sum_{t : N t \in \{0, 1, \ldots, N\}} (\bar{p}(t) - tc) \left[ \alpha^H 1\{\bar{\zeta}(X^H_m, p) = t\} + (1 - \alpha^H) 1\{\bar{\zeta}(X^L_m, p) = t\} \right] \right\}.
$$

Notice that when the firm can observe the valuation vectors $X^H_m$ and $X^L_m$ for each customer $m$ without his type, then we can invoke the revelation principal of the principal agent models to write $\bar{Z}_{m,N}(X^H_m, X^L_m)$ as an optimization problem with incentive compatibility and individual rationality constraints for each type (see Chapter 14 of Mas-Colell et al. (1995)). Hence for $m = 1, \ldots, M$, we get that

$$
\bar{Z}_{m,N}(X^H_m, X^L_m) = \sup_{\bar{p} \in (\mathbb{R}_+ \cup \{\infty\})^N} \alpha (\bar{p}_m(t^H_m) - t^H_m c) + (1 - \alpha) (\bar{p}_m(t^L_m) - t^L_m c)
$$

s.t. $\bar{V}^H_{m,N}(t^H_m) - \bar{p}_m(t^H_m) \geq \bar{V}^H_{m,N}(t^L_m) - \bar{p}_m(t^L_m)$ (IC$_{m-H}$)

$\bar{V}^L_{m,N}(t^L_m) - \bar{p}_m(t^L_m) \geq \bar{V}^L_{m,N}(t^H_m) - \bar{p}_m(t^H_m)$ (IC$_{m-L}$)

$\bar{V}^H_{m,N}(t^H_m) - \bar{p}_m(t^H_m) \geq 0$ (IR$_{m-H}$)

$\bar{V}^L_{m,N}(t^L_m) - \bar{p}_m(t^L_m) \geq 0$ (IR$_{m-L}$)

$t^H_m, t^L_m \in [0, 1]

\bar{p}_m(t^H_m), \bar{p}_m(t^L_m) \in \mathbb{R}_+$,

where (IC$_{m-H}$) and (IC$_{m-L}$) are the incentive compatibility constraints for the high and low types respectively and (IR$_{m-H}$) and (IR$_{m-L}$) are the respective individual rationality constraints. The revelation principal implies that when firm observes the valuation vectors but not the customers’ types, then its optimal policy is to offer each customer $m$ two tailored bundle sizes denoted by $t^H_m$ and $t^L_m$ and allow them to self-select the appropriate bundle size according to their type.

We note that the exact characterization of $\mathbb{E}[\bar{Z}_{1,N}(X^H_1, X^L_1)]$ is not possible since Assumption 3 does not guarantee that $\bar{V}^H_{m,N}(t) \geq \bar{V}^L_{m,N}(t)$ for all $t \in [0, 1]$ and $m = 1, \ldots, M$. In fact, the two curves can be intersecting or even $\bar{V}^L_{m,N}(t)$ can be below $\bar{V}^H_{m,N}(t)$. Moreover, the supremum might not be attained since the feasible set is not necessarily closed. Hence, the problem does not satisfy the usual “single-crossing property” in the classical monopolistic screening problems. This makes
the problem complicated. However, we note that if we replace $\tilde{V}^{H}_{m,N}$ and $\tilde{V}^{L}_{m,N}$ with their limiting process $\tilde{V}^{H}$ and $\tilde{V}^{L}$ where $\tilde{V}^{H}(t) = \int_{0}^{t} F_{X}^{-1}(1-s)ds$ and $\tilde{V}^{L}(t) = \int_{0}^{t} F_{X}^{-1}(1-s)ds$ for $t \in [0,1]$, then it is possible to show that the obtained limiting problem satisfy the single-crossing property.

Therefore, the problem belongs to the same overarching framework of this paper, that is, the problem is in general complicated but its limiting problem is easy to solve. However, in this case, establishing weak convergence of $\tilde{Z}_{m,N}(X_{m}^{H},X_{m}^{L})$ (a constrained optimization problem) is more difficult. In addition, there are technical issues with the convergence analysis regarding the difference between the normalized curves near $t = 0$. For this reason, instead of directly proving the convergence of the optimization problem, we focus on analyzing $\tilde{Z}_{m,N}(X_{m}^{H},X_{m}^{L})$ on the event where the valuation curve of the high type is above that of the low type on some closed interval away from zero. In particular for any $\delta \in (0,1)$, we define the following event

$$A_{m,N}(\delta) = \{\tilde{V}^{H}_{m,N}(t) - \tilde{V}^{L}_{m,N}(t) \geq 0 \text{ for all } t \in [\delta,1]\},$$

where we denote its complement by $A^{c}_{m,N}(\delta)$.

We now show that for the sake of the asymptotic analysis $\mathbb{E}[\tilde{Z}_{1,N}(X_{1}^{H},X_{1}^{L})]$, we can restrict ourselves to analyzing $\tilde{Z}_{m,N}(X_{m}^{H},X_{m}^{L})$ on the event $A_{m,N}(\delta)$ for some $\delta \in (0,1)$. We first have the following Lemma.

**Lemma C1.** Given Assumption 3 and for any $\delta \in (0,1)$, we have that $1\{A_{m,N}(\delta)\} \overset{L_{1}}{\rightarrow} 1$ and hence $1\{A_{m,N}(\delta)\} \Rightarrow 1$ as $N \to \infty$.

**Proof of Lemma C1.** Fix $\delta \in (0,1)$, we have that $\mathbb{E}[1\{A_{m,N}(\delta)\} - 1] = 1 - \mathbb{E}[1\{A_{m,N}(\delta)\}] = 1 - \mathbb{P}(A_{m,N}(\delta))$. Therefore, the $L_{1}$ convergence of $1\{A_{m,N}(\delta)\}$ is equivalent to showing that $\mathbb{P}(A_{m,N}(\delta)) \to 1$ as $N \to \infty$. Now notice that

$$A_{m,N}(\delta) = \left\{ \inf_{t \in [\delta,1]} [\tilde{V}^{H}_{m,N}(t) - \tilde{V}^{L}_{m,N}(t)] \geq 0 \right\}.$$

Using Lemma A1, we have that $\mathbb{P}$-a.s. $\tilde{V}^{H}_{m,N} \to \tilde{V}^{H}$ and $\tilde{V}^{L}_{m,N} \to \tilde{V}^{L}$ in $D([0,1],\mathbb{R})$ as $N \to \infty$ and hence we get have that $\mathbb{P}$-a.s. $\tilde{V}^{H}_{m,N} - \tilde{V}^{L}_{m,N} \to \tilde{V}^{H} - \tilde{V}^{L}$ in $D([0,1],\mathbb{R})$ as $N \to \infty$. This implies that $\mathbb{P}$-a.s. $\inf_{t \in [\delta,1]} [\tilde{V}^{H}_{m,N}(t) - \tilde{V}^{L}_{m,N}(t)] \to \inf_{t \in [\delta,1]} [\tilde{V}^{H}(t) - \tilde{V}^{L}(t)]$ as $N \to \infty$.

However, it follows from Assumption 3 that $\tilde{V}^{H}(t) - \tilde{V}^{L}(t) > 0$ for $0 < t \leq 1$ where $\tilde{V}^{H}(t) - \tilde{V}^{L}(t)$ is increasing in $t$. Therefore, we get that $\inf_{t \in [\delta,1]} [\tilde{V}^{H}(t) - \tilde{V}^{L}(t)] = \tilde{V}^{H}(\delta) - \tilde{V}^{L}(\delta) > 0$. Now noting that $\mathbb{P}$-a.s. convergence implies convergence in distribution on continuity points, we get that

$$\mathbb{P}(A_{m,N}(\delta)) \to \mathbb{P}(\tilde{V}^{H}_{m,N}(\delta) - \tilde{V}^{L}_{m,N}(\delta) > 0) = 1 \text{ as } N \to \infty.$$

This establishes the $L_{1}$ convergence of $1\{A_{m,N}(\delta)\}$ which in return implies convergence in probability. □
Now denote by $t_m^{H,*}$ and $t_m^{L,*}$ as an optimal solution to $Z_{m,N}(X_m^H, X_m^L)$. Moreover, for a given $\epsilon > 0$, we define another event

$$B_{m,N}(\epsilon) = \{t_m^{L,*}, t_m^{H,*} > \epsilon\},$$

where we denote its complement by $B_{m,N}^c(\epsilon)$. We now have the following result on the asymptotic value $\mathbb{E}[\bar{Z}_{1,N}(X_1^H, X_1^L)]$.

**Lemma C2.** For a given $\delta \in (0, 1)$, if there exists $\epsilon > 0$ such that $1\{A_{m,N}(\delta) \cap B_{m,N}(\epsilon)\} \Rightarrow 1$ and $1\{A_{m,N}(\delta) \cap B_{m,N}(\epsilon)\} \bar{Z}_{m,N}(X_m^H, X_m^L) \Rightarrow a$ as $N \to \infty$ for some $a \in \mathbb{R}_+$, then we get that

$$\mathbb{E}[\bar{Z}_{1,N}(X_1^H, X_1^L)] \to a \text{ as } N \to \infty.$$  

**Proof of Lemma C2.** Notice that for any $\delta, \epsilon \in (0, 1)$, we can write

$$\mathbb{E}[\bar{Z}_{1,N}(X_1^H, X_1^L)] = \int 1\{A_{1,N}(\delta) \cap B_1(N, \epsilon)\} \bar{Z}_{1,N}(X_1^H, X_1^L)d\mathbb{P}$$

$$+ \int 1\{A_{1,N}(\delta) \cup B_1(N, \epsilon)\} \bar{Z}_{1,N}(X_1^H, X_1^L)d\mathbb{P}.$$  

However, we have that $\bar{Z}_{1,N}(X_1^H, X_1^L) \leq N^{-1} \sum_{n=1}^N (X_{1,n}^H + X_{1,n}^L)$, where the right hand side is the profit under (first-degree) perfect price discrimination with zero marginal costs. As a result, we get

$$\int 1\{A_{1,N}(\delta) \cup B_1(N, \epsilon)\} \bar{Z}_{1,N}(X_1^H, X_1^L)d\mathbb{P} \leq \int 1\{A_{1,N}(\delta) \cup B_1(N, \epsilon)\} N^{-1} \sum_{n=1}^N (X_{1,n}^H + X_{1,n}^L)d\mathbb{P}.$$  

(C1)

By the weak law of large numbers, we have that $N^{-1} \sum_{n=1}^N (X_{1,n}^H + X_{1,n}^L) \Rightarrow \mu^H + \mu^L < \infty$ as $N \to \infty$ and from the statement of the lemma we have that $1\{A_{1,N}(\delta) \cup B_{1,N}(\epsilon)\} \Rightarrow 0$ as $N \to \infty$. Therefore, by the continuous mapping theorem, we get that $1\{A_{1,N}(\delta) \cup B_{1,N}(\epsilon)\} N^{-1} \sum_{n=1}^N (X_{1,n}^H + X_{1,n}^L) \Rightarrow 0$. Moreover, by the generalized dominated convergence theorem we get that $\int 1\{A_{1,N}(\delta) \cup B_{1,N}(\epsilon)\} N^{-1} \sum_{n=1}^N (X_{1,n}^H + X_{1,n}^L) d\mathbb{P} \to 0$ as $N \to \infty$ and it follows from (C1) that $\int 1\{A_{1,N}(\delta) \cup B_{1,N}(\epsilon)\} \bar{Z}_{1,N}(X_1^H, X_1^L)d\mathbb{P} \to 0$ as $N \to \infty$. Consequently, we get that $\lim_{N \to \infty} \mathbb{E}[\bar{Z}_{1,N}(X_1^H, X_1^L)] = \lim_{N \to \infty} \int 1\{A_{1,N}(\delta) \cap B_1(N, \epsilon)\} \bar{Z}_{1,N}(X_1^H, X_1^L)d\mathbb{P}.$

Again from the statement of the Lemma and using the generalized dominated convergence, we get that $\lim_{N \to \infty} \int 1\{A_{1,N}(\delta) \cap B_1(N, \epsilon)\} \bar{Z}_{1,N}(X_1^H, X_1^L)d\mathbb{P} = a$ which concludes the proof. \hfill \Box

We now have the following refinement.

**Lemma C3.** Given Assumption 3 assume that $t_m^{L,*} \Rightarrow d$ and $t_m^{H,*} \Rightarrow e$ as $N \to \infty$ for some $d, e > 0$ a. Then there exists $\epsilon > 0$ such that for any $\delta \in (0, 1)$ we have that $1\{A_{1,N}(\delta) \cap B_{1,N}(\epsilon)\} \xrightarrow{L} 1$ and hence $1\{A_{1,N}(\delta) \cap B_{1,N}(\epsilon)\} \Rightarrow 1$ as $N \to \infty$. 

Proof of Lemma C3. Similar to the proof of C1, it suffices to show that there exists $\epsilon > 0$ such that for any $\delta \in (0, 1)$ we get that $\mathbb{P}(A_{m,N}(\delta) \cap B_{m,N}(\epsilon)) \to 1$ as $N \to \infty$.

We have that $\mathbb{P}(A_{m,N}(\delta) \cap B_{m,N}(\epsilon)) \geq \mathbb{P}(A_{m,N}(\delta)) + \mathbb{P}(B_{m,N}(\epsilon)) - 1$. Now from Lemma C1 and for any $\delta \in (0, 1)$, we have that $\mathbb{P}(A_{m,N}(\delta)) \to 1$ as $N \to \infty$. Therefore, it suffices to show that there exists $\epsilon > 0$ such that $\mathbb{P}(B_{m,N}(\epsilon)) \to 1$ as $N \to \infty$.

First, notice that for any $\epsilon > 0$, we have that

$$B_{m,N}(\epsilon) \equiv \{\min\{t_{m,\ast}^L, t_{m,\ast}^H\} > \epsilon\}.$$ 

Now, let $\epsilon = \min\{d, e\}/2$. Therefore, we get that

$$\mathbb{P}(B_{m,N}(\min\{d, e\}/2)) = \mathbb{P}(\min\{t_{m,\ast}^L, t_{m,\ast}^H\} - \min\{d, e\} > -\min\{d, e\}/2) \geq \mathbb{P}(\min\{t_{m,\ast}^L, t_{m,\ast}^H\} - \min\{d, e\} \leq \min\{d, e\}/2).$$

However, by the continuous mapping theorem we get that $\min\{t_{m,\ast}^L, t_{m,\ast}^H\} \Rightarrow \min\{d, e\}$ as $N \to \infty$. Consequently, by the definition of convergence in probability we get that $\mathbb{P}(\min\{t_{m,\ast}^L, t_{m,\ast}^H\} - \min\{d, e\} \leq \min\{d, e\}/2) \to 1$ as $N \to \infty$ which concludes the proof. \(\square\)

Let $\tilde{t}^H = 1 - F_{\hat{X}^H}(c)$ where $\tilde{t}^H$ is the asymptotically optimal size if the firm serves only the high types. We now have the following result.

**Lemma C4.** Given Assumption 3 and assuming that $t_{m,\ast}^L \Rightarrow d$ and $t_{m,\ast}^H \Rightarrow e$ as $N \to \infty$ for some $d, e > 0$, then $t_{m,\ast}^H \Rightarrow \tilde{t}^H$ as $N \to \infty$. Furthermore, we have that $\mathbb{E}[\bar{Z}_{1,N}(X_{1}^H, X_{1}^L)] \to \bar{Z}$ as $N \to \infty$ where

$$\bar{Z} = \alpha(\bar{V}(\tilde{t}^H) - \tilde{t}^H c) + \max_{t^L \in [0, 1]} [\bar{V}^L(t^L) - \alpha \bar{V}^H(t^L) - (1 - \alpha) t^L c].$$

(C2)

**Proof of Lemma C4.** Following from Lemma C3, there exists $\epsilon > 0$ such that $1\{A_{m,N}(\delta) \cap B_{m,N}(\epsilon) \} \to 1$ as $N \to \infty$. Therefore, from Lemma C2 we can restrict our analysis to any event $A_{m,N}(\delta) \cap B_{m,N}(\epsilon)$ where $\delta \in (0, 1)$.

The realized value of $\bar{Z}_{m,N}(X_{m}^H, X_{m}^L)$ on the event $A_{m,N}(\delta) \cap B_{m,N}(\epsilon)$ reduces to the following optimization problem

$$\bar{Z}_{m,N}(X_{m}^H, X_{m}^L) = \sup \alpha (\bar{p}_m(t_{m}^H) - t_{m}^H c) + (1 - \alpha) (\bar{p}_m(t_{m}^L) - t_{m}^L c)$$

s.t. $V_{m,N}(t_{m}^H) - \bar{V}_m(t_{m}^H) \geq V_{m,N}(t_{m}^L) - \bar{V}_m(t_{m}^L)$ \hspace{1cm} (IC$_{m}$-H)

$V_{m,N}(t_{m}^L) - \bar{V}_m(t_{m}^L) = 0$ \hspace{1cm} (IR$_{m}$-L)

$t_{m}^H, t_{m}^L \in [\epsilon, 1]$

$\bar{p}_m(t_{m}^H), \bar{p}_m(t_{m}^L) \in \mathbb{R}_+$,
where IC\(_m\text{-}L\) and IR\(_m\text{-}H\) are automatically satisfied at optimality since the high valuation curve is above the low valuation curve on the interval \([\epsilon, 1]\) given the event \(A_{m,N}(d)\).

From the classical monopolistic screening problems (see section 14C of Mas-Colell et al. (1995)), both IR\(_m\text{-}L\) and IC\(_m\text{-}H\) will be binding at optimality. Moreover, the high valuation types will be served with their efficient bundle size had they been alone in the market. In particular, we have that \(t_{m}^{H*} = 1 - \inf \{t : F_{XU,m,N}^{-1}(t) \geq c\}\) where \(\inf \{t : F_{XU,m,N}^{-1}(t) \geq c\}\) is the normalized rank of the last order statistic that is larger than the marginal cost.

Therefore, setting \(t_{m}^{H*} = 1 - \inf \{t : F_{XU,m,N}^{-1}(t) \geq c\}\) and \(\bar{p}_{m}(t_{m}^{H}) = \bar{V}_{m,N}(t_{m}^{H*}) - (\bar{V}_{m,N}(t_{m}^{L}) - \bar{V}_{m,N}(t_{m}^{L})),\) then \(Z_{m,N}(X_{m},X_{m}^{L})\) given the event \(A_{m,N}(\delta) \cap B_{m,N}(\epsilon)\) reduces to the following univariate optimization problem

\[
\bar{Z}_{m,N}(X_{m},X_{m}^{L}) = \sup_{t_{m}^{L}} \alpha \left(\bar{V}_{m,N}(t_{m}^{H*}) - (\bar{V}_{m,N}(t_{m}^{L}) - \bar{V}_{m,N}(t_{m}^{L}))) - t_{m}^{L} c\right) + (1 - \alpha) \left(\bar{V}_{m,N}(t_{m}^{L}) - t_{m}^{L} c\right)
\]

However, since on the event \(B_{m,N}(\epsilon)\) we have that \(t_{m}^{L*} \geq \epsilon,\) then we can use the relaxed problem on \([0, 1]\) and \(t_{m}^{L*} \in \arg\sup_{t \in [0, 1]} \bar{V}_{m,N}(t_{m}^{L}) - \alpha \bar{V}_{m,N}(t_{m}^{L}) - (1 - \alpha) t_{m}^{L} c.\)

From Lemma A1, we get that \(\bar{V}_{m,N}(t_{m}^{L}) - \alpha \bar{V}_{m,N}(t_{m}^{L}) - (1 - \alpha) t_{m}^{L} c\) converges uniformly in probability to \(\bar{V}_{m,N}(t_{m}^{L}) - \alpha \bar{V}_{m,N}(t_{m}^{L}) - (1 - \alpha) t_{m}^{L} c\) on \([0, 1]\) as \(N \to \infty.\) Therefore,

\[
\sup_{t \in [0, 1]} \left(\bar{V}_{m,N}(t_{m}^{L}) - \alpha \bar{V}_{m,N}(t_{m}^{L}) - (1 - \alpha) t_{m}^{L} c\right) \Rightarrow \sup_{t \in [0, 1]} \left(\bar{V}_{m,N}(t_{m}^{L}) - \alpha \bar{V}_{m,N}(t_{m}^{L}) - (1 - \alpha) t_{m}^{L} c\right) \text{ as } N \to \infty.
\]

In addition, by the Glivenko-Cantelli lemma and the fact that \(F_{XU}^{-1}\) is non-decreasing, we have that \(P\text{-a.s } F_{XU,m}^{-1}(t) \Rightarrow F_{XU}^{-1}(t)\) for \(t(0, 1)\) where \(F_{XU}^{-1}(t)\) is continuous. Since \(F_{XU}(c)\) is a continuity point of \(F_{XU}(t),\) then \(t_{m}^{H*,t} \Rightarrow 1 - \inf \{t : F_{XU}(t) \geq c\} = 1 - F_{XU}(c) \equiv \tilde{t}_{H}.\) Therefore, by the continuous mapping theorem we have that

\[
1\{A_{m,N}(\delta) \cap B_{m,N}(\epsilon)\} \bar{Z}_{m,N}(X_{m},X_{m}^{L}) \Rightarrow Z \text{ as } N \to \infty.
\]

Finally, the convergence in expectations follows from Lemma C2 which concludes the proof. \(\square\)

**Lemma C5.** Given Assumption 3, then \(t_{m}^{H*} \Rightarrow 0\) as \(N \to \infty\) if and only if \(\tilde{t}_{H} = 0.\) Moreover, if \(\tilde{t}_{H} = 0\) then \(t_{m}^{L*} \Rightarrow t_{L*} = 0\) as \(N \to \infty.\)

**Proof of Lemma C5.** First notice that the limit of any weakly convergent subsequence of \(t_{m}^{H*}\) should be less than \(\tilde{t}_{H}.\) Hence, if \(\tilde{t}_{H} = 0\) then \(t_{m}^{H*} \Rightarrow 0.\) In addition, due to Assumption 3, if \(\tilde{t}_{H} = 0,\) then \(\tilde{t}_{L} = 0\) where \(\tilde{t}_{L} = 1 - F_{XU}(c)\) is the asymptotically optimal size that would have been offered to the low types had they been alone in the market. It can also be shown that \(t_{L*} \leq \tilde{t}_{L}\) hence also \(t_{L*} = 0.\) As a result, it has to be the case that \(t_{m}^{L*} \Rightarrow t_{L*} = 0\) as \(N \to \infty.\)

On the other hand, if \(t_{m}^{H*} \Rightarrow 0\) as \(N \to \infty,\) then by the continuous mapping theorem we have that \(\bar{V}_{H}(t_{m}^{H*}) \Rightarrow 0,\) i.e. the firm should only sell to the low types asymptotically. But in this case,
we know that the asymptotically optimal size for the low customers is \( \hat{t}^L = 1 - F_{X_L}(c) \). But if \( \hat{t}^L > 0 \), then the firm can get asymptotically higher profits by selling a bundle for the high valuation customers of size \( t^H = \hat{t}^L \). Therefore, it has to be case that \( \hat{t}^L = 0 \) and \( t^L_m = t^H_m = 0 \) as \( N \to \infty \). Therefore, if \( t^L_m \to 0 \) then the asymptotic profits is zero which implies that \( \hat{t}^H = 0 \) since otherwise, by letting \( t^H_m \to \hat{t}^H \) as \( N \to \infty \), the firm can asymptotically achieve strictly positive profits. \( \square \)

The implications are intuitive: if it is not (asymptotically) optimal to sell bundles for high valuation customers when they are alone in the market, then the firm should not sell any bundles when there are both high and low types. We now have the following result.

**Lemma C6.** Given Assumption 3, we have that \( \mathbb{E}[Z_{1,N}(X^H_1,X^L_1)] \to \bar{Z} \) as \( N \to \infty \).

**Proof of Lemma C6.** Let \( T^* = \arg \max_{t \in [0,1]} W(t^L) \) where \( W(t^L) = \bar{V}^L(t^L) - \alpha \bar{V}^H(t^L) - (1 - \alpha)t^Lc \). Notice that \( \bar{Z} \) is attainable in the limit by setting \( t^H = \hat{t}^L \) and \( t^L_m \to \hat{t}^L \) for any \( t^L \in T^* \) as \( N \to \infty \).

First if \( \hat{t}^H = 0 \), then it follows from Lemma C5 that \( t^H_m, t^L_m \to 0 \) as \( N \to \infty \) and \( T^* = \{0\} \). Hence, the result is trivially established.

We now consider the case when \( \hat{t}^H > 0 \). Since \( (t^H_m) \) and \( (t^L_m) \) are bounded sequences then each has at least one weakly convergent subsequence. We examine two cases.

**Case I** \( (0 \notin T^*) \). We have that \( W(t^L_m) > 0 \) for all \( t^L_m \in T^* \). Now assume that \( (t^L_m) \) has a subsequence that converges in probability to 0. Then, the corresponding weakly convergent subsequence of \( (t^H_m) \) converges in probability to \( \hat{t}^L \) and \( \mathbb{E}[Z^*_{m,N}(X^H_m,X^L_m)] \to \alpha(\bar{V}^H(\hat{t}^L) - \hat{t}^Lc) < \bar{Z} \) as \( N \to \infty \) which is a contradiction. Therefore, every weakly convergent subsequence of \( (t^L_m) \) converges in probability to some \( d \in (0,1] \) (\( d \) can potentially be different for different subsequences).

Moreover, by Lemma C5 every corresponding weakly convergent subsequence of \( (t^H_m) \) converges in probability to some \( e \in (0,1] \) (\( e \) can potentially be different for different subsequences).

**Case II** \( (0 \in T^*) \). We have that \( W(t^L_m) > 0 \) for all \( t^L_m \in T^* \). In this case, if \( (t^L_m) \) has a subsequence that converges in probability to 0 then \( Z^*_{m,N}(X^H_m,X^L_m) \to \alpha(\bar{V}^H(\hat{t}^L) - \hat{t}^Hc) = \bar{Z} \) therefore the result is established on that subsequence. Now assume there exists a weakly convergent subsequence of \( (t^L_m) \) converges in probability to some \( d \in (0,1] \) (\( d \) can potentially be different for different subsequences) and every corresponding weakly convergent subsequence of \( (t^H_m) \) converges in probability to some \( e \in (0,1] \) (\( e \) can potentially be different for different subsequences).

Therefore in both cases, we are left to examine what happens on every weakly convergent subsequence of \( (t^L_m) \) and \( (t^H_m) \) that converge in probability to some \( d \in (0,1] \) and \( e \in (0,1) \). However, by Lemma C4, it follows that on each of these subsequences we have that \( \mathbb{E}[Z_{m,N}(X^H_m,X^L_m)] \to \bar{Z} \) as \( N \to \infty \). Therefore, since these subsequences were chosen arbitrarily, then our result is established. \( \square \)
C.2. Asymptotically optimal bundle size policy

We are now ready to present the proof of the theorem where we show how the the optimal solution to $\bar{Z}$ can be used to construct an asymptotically optimal bundle size policy that achieves the expected profit under second degree price discrimination.

Proof of Theorem 5. First, notice that the univariate optimization problem in (C2) is not necessarily concave. Hence, in general, it is not easy to obtain $t_{L,\ast}$ and is not necessarily unique. For the sake of the proof, we assume that we can obtain at least one maximizer $t_{L,\ast}$.

Under the pricing policy described in the statement of the theorem, and since $\limsup_{N \to \infty} \frac{g(N)}{\kappa(N)} < 1$, then it follows that the probability that the high valuation customer retains a higher surplus at $\bar{t}^H$ than $t_{L,\ast}$ goes to 1 as $N$ goes to infinity. Likewise, the surplus of the low valuation customer is higher at $t_{L,\ast}$. Hence, we are left to show that both customer types retain positive surplus at their respective bundle sizes in the limit.

For the high valuation customers, letting $V_{1,N}^H(\bar{t}^H) = \left(V_{1,N}^H(\bar{t}^H) - \mathbb{E}[V_{1,N}^H(\bar{t}^H)]\right)/\sqrt{N}$, we obtain

$$P\left(V_{1,N}^H(\bar{t}^H) > \mathbb{E}[V_{1,N}^H(\bar{t}^H)] - (\mathbb{E}[\hat{V}_{1,N}^H(t_{L,\ast})] - \mathbb{E}[\hat{V}_{1,N}^L(t_{L,\ast})]) - h(N)/N\right)$$

$$= P\left(\hat{V}_{1,N}^H(\bar{t}^H) > -\sqrt{N} (\mathbb{E}[\hat{V}_{1,N}^H(t_{L,\ast})] - \mathbb{E}[\hat{V}_{1,N}^L(t_{L,\ast})]) - h(N)/\sqrt{N}\right) \to 1 \quad \text{as } N \to \infty,$$

where the convergence result follows from the fact that $\hat{V}_{1,N}^H(\bar{t}^H)$ is asymptotically normal, $\lim_{N \to \infty} (\mathbb{E}[\hat{V}_{1,N}^H(t_{L,\ast})] - \mathbb{E}[\hat{V}_{1,N}^L(t_{L,\ast})]) \geq 0$, and $h(N) \in \omega(\sqrt{N})$.

For the low valuation customers, we have

$$P\left(V_{1,N}^L(t_{L,\ast}) > \mathbb{E}[\hat{V}_{1,N}^L(t_{L,\ast})] - g(N)/N\right) = P\left(\hat{V}_{1,N}^L(t_{L,\ast}) > -g(N)/\sqrt{N}\right) \to 1 \quad \text{as } N \to \infty,$$

where the convergence result follows from the fact that $\hat{V}_{1,N}^L(t_{L,\ast})$ is asymptotically normal and $g(N) \in \omega(\sqrt{N})$.

The expected normalized profit is given by

$$\mathbb{E}[\bar{\pi}(p)]/M = \alpha^H \left(\mathbb{E}[\hat{V}_{1,N}^H(\bar{t}^H)] - (\mathbb{E}[\hat{V}_{1,N}^H(t_{L,\ast})] - \mathbb{E}[\hat{V}_{1,N}^L(t_{L,\ast})]) - h(N)/N - \lceil N\bar{t}^H \rceil \cdot c/N\right)$$

$$\cdot P\left(\hat{V}_{1,N}^H(\bar{t}^H) > -\sqrt{N} (\mathbb{E}[\hat{V}_{1,N}^H(t_{L,\ast})] - \mathbb{E}[\hat{V}_{1,N}^L(t_{L,\ast})]) - h(N)/\sqrt{N}\right)$$

$$+ (1 - \alpha^H) \left(\mathbb{E}[\hat{V}_{1,N}^L(t_{L,\ast})] - g(N)/N - \lceil Nt_{L,\ast} \rceil \cdot c/N\right) P\left(\hat{V}_{1,N}^L(t_{L,\ast}) > -g(N)/\sqrt{N}\right) + o(1)$$

$$\to \alpha^H \left[\hat{V}_{1,N}^H(\bar{t}^H) - (\hat{V}_{1,N}^H(t_{L,\ast}) - \hat{V}_{1,N}^L(t_{L,\ast}) - \bar{t}^H c) \right] + (1 - \alpha^H) \left[\hat{V}_{1,N}^L(t_{L,\ast}) - t_{L,\ast} c\right].$$

Therefore as $N \to \infty$, we get

$$\mathbb{E}[\bar{\pi}(p)] \to M \left[\alpha^H \left(\hat{V}_{1,N}^H(\bar{t}^H) - \hat{V}_{1,N}^L(t_{L,\ast}) - \bar{t}^H c\right) + \hat{V}_{1,N}^L(t_{L,\ast}) - (1 - \alpha^H) t_{L,\ast} c\right]$$

$$= M \bar{Z}.$$

Finally, it follows from Lemma C6 that also $M \mathbb{E}[\bar{Z}_{1,N}(X^H_1, X^L_1)] \to M \bar{Z}$ as $N \to \infty$ which concludes the proof. \qed
D. Proofs of Section 7.

Before presenting the proofs of this section, we introduce some convergence rate definitions that will be useful to establish the results.

**Definition 1 (Modified Landau notation).** We use the following definitions to represent the limiting behavior of a function $f_n : \mathbb{N} \to \mathbb{R}$:

- $O(h_n) = \{ f_n : \limsup_{n \to \infty} \left| \frac{f_n}{h_n} \right| < +\infty \}$.
- $\Omega(h_n) = \{ f_n : \liminf_{n \to \infty} \left| \frac{f_n}{h_n} \right| > 0 \}$.
- $\Omega_+(h_n) = \{ f_n : \liminf_{n \to \infty} \frac{f_n}{h_n} > 0 \}$.
- $\Omega_-(h_n) = \{ f_n : \limsup_{n \to \infty} \frac{f_n}{h_n} < 0 \}$.
- $\Theta(h_n) = O(h_n) \cap \Omega(h_n)$.
- $\Theta_+(h_n) = O(h_n) \cap \Omega_+(h_n)$.
- $o(h_n) = \{ f_n : \lim_{n \to \infty} \left| \frac{f_n}{h_n} \right| = 0 \}$.
- $\omega_+(h_n) = \{ f_n : \liminf_{n \to \infty} \frac{f_n}{h_n} = +\infty \}$.
- $\omega_-(h_n) = \{ f_n : \limsup_{n \to \infty} \frac{f_n}{h_n} = -\infty \}$.

**Proof of Proposition 6.** We start the analysis by showing that $g^*(N) \in \omega_+(\sqrt{N}) \cap o(N)$. To do so, we first argue by contradiction that $g^*(N) \notin \omega_-(\sqrt{N}) \cup O(\sqrt{N}) \cup \Omega_+(N)$. However, we note that $\omega_-(\sqrt{N}) \cup O(\sqrt{N}) \cup \Omega_+(N) \subset \left( \omega_+(\sqrt{N}) \cap o(N) \right)^c \dagger$. For this reason, we further examine each boundary case separately to show that $g^*(N) \notin \left( \omega_+(\sqrt{N}) \cap o(N) \right)^c$.

In the case of offering size $N$, the valuation for the full bundle is given by the partial sum $V_{m,N} = \sum_{n=1}^{N} X_{m,n}$. We do a case-by-case analysis to show that $g^*(N) \notin \omega_-(\sqrt{N}) \cup O(\sqrt{N}) \cup \Omega_+(N)$.

**Case I** ($g(N)^* \in \omega_-(\sqrt{N})$)

We consider two subcases. First, assume $g^*(N) \in \omega_-(\sqrt{N}) \cap o(N)$, then we have that

$$
\limsup_{N \to \infty} \frac{\mathbb{E}[\pi(p)]}{MN\mu} = \limsup_{n \to \infty} \left( 1 - \frac{g^*(N)}{N\mu} \right) \mathbb{P}(V_{1,N}(N) > N\mu - g^*(N))
$$

$$
= \limsup_{n \to \infty} \mathbb{P} \left( \frac{V_{1,N}(N) - N\mu}{\sigma\sqrt{N}} > \frac{g^*(N)}{\sigma\sqrt{N}} \right).
$$

By the central limit theorem, we have that $\frac{V_{1,N}(N) - N\mu}{\sigma\sqrt{N}} \xrightarrow{d} N(0,1)$, i.e. $\frac{V_{1,N}(N) - N\mu}{\sigma\sqrt{N}}$ converges in distribution to a standard normal distribution. However, $\frac{g^*(N)}{\sigma\sqrt{N}} \to +\infty$ as $N \to \infty$, therefore

$$
\limsup_{N \to \infty} \frac{\mathbb{E}[\pi(p)]}{MN\mu} < 1.
$$

\dagger The reason that it is a proper subset is the existence of boundary cases which do not belong to any of the sets. For example, consider $g(N)$ such that $\limsup_{N \to \infty} \frac{g(N)}{\sqrt{N}} = +\infty$ and $0 < \liminf_{N \to \infty} \frac{g(N)}{\sqrt{N}} < +\infty$. Then $g(N) \in \left( \omega_+(\sqrt{N}) \cap o(N) \right)^c$ but $g(N) \notin \omega_-(\sqrt{N}) \cup O(\sqrt{N}) \cup \Omega_+(N)$. 

Therefore, \( g^*(N) \in \omega_-(\sqrt{N}) \cap o(N) \) is not optimal.

Now we consider the subcase where \( g^*(N) \in \Omega_-(N) \). We have the following inequality on the ratio of the profits

\[
\limsup_{N \to \infty} \frac{E[\pi(p)]}{MN\mu} = \limsup_{N \to \infty} \left( 1 - \frac{g^*(N)}{N\mu} \right) \mathbb{P}(V_{1,N}(N) > N\mu - g^*(N))
\]

\[
\leq \limsup_{N \to \infty} \left( 1 - \frac{g^*(N)}{N\mu} \right) \mathbb{P}([V_{1,N}(N) - N\mu| \geq -g^*(N))
\]

\[
\leq \limsup_{N \to \infty} \left( 1 - \frac{g^*(N)}{N\mu} \right) \frac{N\sigma^2}{g^*(N)^2} \quad (\text{using Chebychev's Inequality for } -g^*(N) \in \Omega_+(N))
\]

\[
= \limsup_{N \to \infty} \left[ \frac{N\sigma^2}{g^*(N)^2} - \frac{\sigma^2}{\mu g^*(N)} \right] = 0, \quad (\text{since } g^*(N) \in \Omega_-(n))
\]

which implies that \( g^*(N) \in \Omega_-(N) \) is not optimal.

Notice that the proof of this case is not complete yet since there remains two boundary cases where \( \liminf_{N \to \infty} \frac{g^*(N)}{\sqrt{N}} < 0 \) but \( \limsup_{N \to \infty} \frac{g^*(N)}{\sqrt{N}} = 0 \) in addition to the case where \( \liminf_{N \to \infty} |\frac{g^*(N)}{\sqrt{N}}| = 0 \) but \( \limsup_{N \to \infty} \frac{g^*(N)}{\sqrt{N}} \neq 0 \). This is handled in Boundary Case 2 later in the proof.

**Case 2** \((g^*(N) \in O(\sqrt{N}))\)

Assume that \( g^*(N) \in O(\sqrt{N}) \), then we have that

\[
\limsup_{N \to \infty} \frac{E[\pi(p)]}{MN\mu} = \left( \lim_{N \to \infty} \frac{N\mu - g^*(N)}{N\mu} \right) \left( \limsup_{N \to \infty} \mathbb{P}(V_{1,N}(N) > N\mu - g^*(N)) \right)
\]

\[
= \limsup_{N \to \infty} \mathbb{P} \left( \frac{V_{1,N}(N) - N\mu}{\sigma \sqrt{N}} > \frac{-g^*(N)}{\sigma \sqrt{N}} \right).
\]

Again by the central limit theorem and since \( \frac{g^*(N)}{\sigma \sqrt{N}} \in O(1) \), we get

\[
\limsup_{N \to \infty} \frac{E[\pi(p)]}{MN\mu} < 1,
\]

which implies that \( g^*(N) \in O(\sqrt{N}) \) is not optimal.

**Case 3** \((g^*(N) \in \Omega_+(N))\)

Assume that \( g^*(N) \in \Omega_+(N) \), we have the following inequality on the ratio of the profits \( \frac{E[\pi(p)]}{MN\mu} \leq 1 - \frac{g^*(N)}{N\mu} \). However, since \( g^*(N) \in \Omega_+(N) \), we have \( \liminf_{N \to \infty} \frac{g^*(N)}{N\mu} > 0 \). Therefore,

\[
\limsup_{N \to \infty} \frac{E[\pi(p)]}{MN\mu} < 1,
\]

which implies that \( g^*(N) \in \Omega_+(N) \) is not optimal. This concludes all the possible cases which shows that \( g^*(N) \notin \omega_-(\sqrt{N}) \cup O(\sqrt{N}) \cup \Omega_+(n) \).

As mentioned earlier, to show that \( g^*(N) \notin \left( \omega_-(\sqrt{N}) \cap o(N) \right)^c \), we are left to show that the boundary cases are also not optimal. The proof for all these cases will be based on the following
argument: assume a sequence \( \{g^*(N)\} \) that satisfies one of the outlined boundary cases below, then there exists a subsequence that belongs to \( \omega_-(\sqrt{N}) \cup O(\sqrt{N}) \cup \Omega_+(N) \) where the limiting ratio of profits is strictly less than 1 which is a contradiction.

Since this argument holds for all the boundary cases, we only enumerate them and state the relative subsequences that are not optimal.

**Boundary Case 1** ("around" \( \omega_-(\sqrt{N}) \))

This case is represented by functions \( g_\alpha \) such that \( \liminf_{N \to \infty} \frac{g(N)}{\sqrt{N}} = -\infty \) and \( \limsup_{N \to \infty} \frac{g(N)}{\sqrt{N}} > -\infty \). For such functions, there exists a subsequence \( g(N_k) \) such that \( \lim_{k \to \infty} \frac{g(N_k)}{N_k^{1/2}} = -\infty \) i.e. \( g(N_k) \in \omega_-(N_k^{1/2}) \). Hence, it follows from Case 1 such functions are not optimal.

**Boundary Case 2** ("around" \( \Omega_-(N) \))

This case is represented by functions \( g(N) \) such that \( \liminf_{N \to \infty} \frac{g(N)}{N} < 0 \) and \( \limsup_{N \to \infty} \frac{g(N)}{N} \geq 0 \). For such functions, there exists a subsequence \( \{g(N_k)\} \) such that \( \lim_{k \to \infty} \frac{g(N_k)}{N_k} < 0 \) i.e. \( g(N_k) \in O(N_k^{1/2}) \). Hence, it follows from Case 1 that it is not optimal.

**Boundary Case 3** ("around" \( O(\sqrt{N}) \))

This case is represented by functions \( g(N) \) such that \( \liminf_{N \to \infty} \frac{g(N)}{\sqrt{N}} < +\infty \) and \( \limsup_{N \to \infty} \frac{g(N)}{\sqrt{N}} = +\infty \). For such functions, there exists a subsequence \( \{g(N_k)\} \) such that \( \lim_{k \to \infty} \frac{g(N_k)}{N_k^{1/2}} < \infty \) i.e. \( g(N_k) \in O(N_k^{1/2}) \). Hence, it follows from Case 2 such functions are not optimal.

**Boundary Case 4** ("around" \( \Omega_+(N) \))

These cases are represented by functions \( g(N) \) such that \( \limsup_{N \to \infty} \frac{g(N)}{N} > 0 \) and \( \liminf_{N \to \infty} \frac{g(N)}{N} \leq 0 \). For such functions, there exists a subsequence \( \{g(N_k)\} \) such that \( \lim_{k \to \infty} \frac{g(N_k)}{N_k} > 0 \) i.e. \( \{g(N_k)\} \in \Omega_+(N_k) \). Hence, it follows from Case 3 that such functions are not optimal.

This concludes the first part of the proof which shows that \( g^*(N) \in \omega_+(\sqrt{N}) \cap o(N) \). We now tighten the result further, by showing that \( g^*(N) \in \omega_+(\sqrt{N}) \cap o(N^2) \) for any \( \beta > 1/2 \). Clearly, the statement holds for \( \beta \geq 1 \). Now, assume for a contradiction that there exists \( \frac{1}{2} < \beta < 1 \) such that \( \limsup_{N \to \infty} \frac{g^*(N)}{N^{1/2}} > 0 \) (possibly equal to \( +\infty \)). Then there exists a subsequence \( \{g^*(N_k)\} \) such that \( \lim_{k \to \infty} \frac{g^*(N_k)}{N_k^{1/2}} > 0 \). In order to get a contradiction, we will show that pricing at \( p(N_k) = N_k \mu - 2 \sigma \sqrt{N_k} \log N_k \) asymptotically leads to higher revenues than pricing at \( p^*(N_k) = N_k \mu - g^*(N_k) \). To keep the notation simple, we will replace \( N_k \) by \( k \).

First, notice that for any \( g(k) \omega_+(\sqrt{k}) \cap o(k) \) we have

\[
\frac{\mathbb{E}[\pi(p)]}{Mk\mu} = 1 - \frac{g(k)}{k\mu} - \frac{(k\mu - g(k)) \mathbb{P}(V_{1,k}(k) \leq k\mu - g(k))}{k\mu}.
\]

Since \( g^*(k) \in \omega_+(\sqrt{k}) \cap o(k) \), then \( -\frac{g^*(k)}{k\mu} - \frac{(k\mu - g^*(k)) \mathbb{P}(V_{1,k}(k) \leq k\mu - g^*(k))}{k\mu} \to 0 \) as \( k \to \infty \). Therefore, using Taylor’s expansion for \( \log(1+x) \), we get

\[
\log \left( \frac{\mathbb{E}[\pi(p^*)]}{k\mu} \right) = -\frac{g^*(k)}{k\mu} - \frac{(k\mu - g^*(k)) \mathbb{P}(V_{1,k}(k) \leq k\mu - g^*(k))}{k\mu}.
\]
Taking the lim inf of the above ratio, we get

\[
\frac{-g^*(k)}{k\mu} - \frac{(k\mu - g^*(k)) \mathbb{P}(V_{1,k}(k) \leq k\mu - g^*(k))}{k\mu} = -\frac{g^*(k)}{k\mu} - \frac{(k\mu - g^*(k)) (1 - \mathbb{P}(V_{1,k}(k) > k\mu - g^*(k)))}{k\mu} + o\left(\frac{-g^*(k)}{k\mu} - \frac{(k\mu - g^*(k)) (1 - \mathbb{P}(V_{1,k}(k) > k\mu - g^*(k)))}{k\mu}\right).
\]

Using the above expression to compare the ratios of the expected revenues under price \( p(k) = k\mu - 2\sigma \sqrt{k \log k} \) relative to \( p^*(k) = k\mu - g^*(k) \), we get

\[
\log \left( \frac{\mathbb{E}[\pi(p)]}{Mk\mu} \right) = \log \left( \frac{\mathbb{E}[\pi(p^*)]}{Mk\mu} \right) = \frac{Mk\mu - \mathbb{E}[\pi(p)] + o(Mk\mu - \mathbb{E}[\pi(p^*)])}{Mk\mu - \mathbb{E}[\pi(p^*)] + o(Mk\mu - \mathbb{E}[\pi(p^*)])}.
\]

Taking the lim inf of the above ratio, we get

\[
\liminf_{k \to \infty} \frac{\log \left( \frac{\mathbb{E}[\pi(p)]}{Mk\mu} \right)}{\log \left( \frac{\mathbb{E}[\pi(p^*)]}{Mk\mu} \right)} = \liminf_{k \to \infty} \frac{Mk\mu - \mathbb{E}[\pi(p)]}{Mk\mu - \mathbb{E}[\pi(p^*)]} \geq 1,
\]

(D1)

where the inequality is due to the assumed optimality of \( p^* \).

We now show that given our assumption where \( \limsup_{n \to \infty} g^*(N)/n^\beta > 0 \), then inequality (D1) does not hold which is a contradiction. First notice that

\[
\frac{\mathbb{E}[\pi(p)]}{Mk\mu} = \frac{(k\mu - 2\sigma \sqrt{k \log k}) \mathbb{P}(V_{1,k}(k) - k\mu > -2\sigma \sqrt{k \log k})}{k\mu} = \frac{(k\mu - 2\sigma \sqrt{k \log k}) (1 - \mathbb{P}(V_{1,k}(k) - k\mu \leq -2\sigma \sqrt{k \log k}))}{k\mu}.
\]

Taking the log, we get

\[
\log \left( \frac{\mathbb{E}[\pi(p)]}{Mk\mu} \right) = \log \left( 1 - \frac{2\sigma \sqrt{k \log k}}{k\mu} \right) + \log \left( 1 - \mathbb{P}(V_{1,k}(k) - k\mu \leq -2\sigma \sqrt{k \log k}) \right).
\]

We have that \( 2\sigma \sqrt{k \log k}/k \to 0 \) and \( \mathbb{P}(V_{1,k}(k) - k\mu \leq -2\sigma \sqrt{k \log k}) \to 0 \) as \( k \to \infty \). Therefore, using Taylor’s expansion for \( \log(1 + x) \) around zero, we get

\[
\log \left( \frac{\mathbb{E}[\pi(p)]}{Mk\mu} \right) = -\frac{2\sigma \sqrt{k \log k}}{k\mu} - \mathbb{P}(V_{1,k}(k) - k\mu \leq -2\sigma \sqrt{k \log k}) + o\left(\frac{\sqrt{k \log k}}{k}\right) + o\left(\mathbb{P}(V_{1,k}(k) - k\mu \leq -2\sigma \sqrt{k \log k})\right).
\]

Likewise, we have

\[
\log \left( \frac{\mathbb{E}[\pi(p^*)]}{Mk\mu} \right) = -\frac{g^*(k)}{k\mu} - \mathbb{P}(V_{1,k}(k) - k\mu \leq -g^*(k)) + o\left(\frac{g^*(k)}{k}\right) + o\left(\mathbb{P}(V_{1,k}(k) - k\mu \leq -g^*(k))\right).
\]
Taking the lim sup of the ratio as \( k \to \infty \), we get

\[
\limsup_{k \to \infty} \frac{\log\left( \frac{E[\pi(p)]}{M k \mu} \right)}{\log\left( \frac{E[\pi(p^*)]}{M k \mu} \right)} = \limsup_{k \to \infty} \frac{2\sigma \sqrt{\log k}}{\sqrt{k} \mu} + \mathbb{P} \left( V_{1,k}(k) - k\mu \leq -2\sigma \sqrt{k \log k} \right)
\]

(D2)

Before proceeding further, we now state the moderate deviation principle (see for example Theorem 11.2 of Rassoul-Agha and Seppäläinen (2015).) that will be useful to finish the proof of this proposition.

**Lemma D1 (Moderate Deviation Principle).** For \( g(k) \in \omega_+(k^{1/2}) \cap o(k) \), we have

\[
\lim_{n \to \infty} \frac{k}{(g(k))^2} \log \left( \mathbb{P} \left( V_{1,k}(k) - k\mu \leq -g(k) \right) \right) = -\frac{1}{2\sigma^2}.
\]

A simpler interpretation of the moderate deviation principle is that

\[
\mathbb{P} \left( V_{1,k}(k) - k\mu \leq -g(k) \right) = \exp \left( -\frac{(g(k))^2}{2k\sigma^2} + o\left( \frac{(g(k))^2}{k} \right) \right) \quad \text{for } g(k) \in \omega_+(k^{1/2}) \cap o(k).
\]

(D3)

Using (D3), the limiting ratio in (D2) becomes

\[
\limsup_{k \to \infty} \frac{\log\left( \frac{E[\pi(p)]}{M k \mu} \right)}{\log\left( \frac{E[\pi(p^*)]}{M k \mu} \right)} = \limsup_{k \to \infty} \frac{2\sigma \sqrt{\log k}}{\sqrt{k} \mu} + \exp \left( -\frac{(g^*(k))^2}{2k\sigma^2} + o\left( \frac{(g^*(k))^2}{k} \right) \right)
\]

\[
= \limsup_{k \to \infty} \frac{g^*(k)}{\sqrt{k} \mu} \left[ 1 + \frac{\exp \left( -\frac{(g^*(k))^2}{2k\sigma^2} + o\left( \frac{(g^*(k))^2}{k} \right) \right)}{g^*(k)} \right].
\]

We first claim that \( \lim_{k \to \infty} \left[ 1 + \frac{\exp \left( -\frac{(g^*(k))^2}{2k\sigma^2} + o\left( \frac{(g^*(k))^2}{k} \right) \right)}{g^*(k)} \right] = 1 \). Notice that for this to hold, it suffices to show that \( \lim_{k \to \infty} \log \left( \frac{\exp \left( -\frac{(g^*(k))^2}{2k\sigma^2} + o\left( \frac{(g^*(k))^2}{k} \right) \right)}{g^*(k)} \right) = -\infty \). In order to see that the preceding holds, note that

\[
\lim_{k \to \infty} \log \left( \frac{\exp \left( -\frac{(g^*(k))^2}{2k\sigma^2} + o\left( \frac{(g^*(k))^2}{k} \right) \right)}{g^*(k)} \right)
\]

\[
= \lim_{k \to \infty} \left[ \frac{(g^*(k))^2}{2k\sigma^2} - \log \left( \frac{g^*(k)}{k \mu} \right) + o\left( \frac{(g^*(k))^2}{k} \right) \right]
\]

\[
= \lim_{k \to \infty} -\frac{k^{2\beta - 1}}{2\sigma^2} \lim_{k \to \infty} \frac{g^*(k)^2}{k^{2\beta}} \quad \text{(since } 2\beta > 1)\]

\[
= -\infty.
\]
As a result, (D4) reduces to the following
\[
\limsup_{k \to \infty} \frac{\log \left( \frac{E[\pi(p)]}{Mk\mu} \right)}{\log \left( \frac{E[\pi(p^*)]}{Mk\mu} \right)} = \limsup_{k \to \infty} \frac{2\sigma \sqrt{k \log k} + \exp(-2\log k + o(\log k))}{g^*(k)}
\]
\[
= \lim_{k \to \infty} \frac{2\sigma \sqrt{k \log k}}{g^*(k)} + \limsup_{k \to \infty} \frac{\exp(-2\log k + o(\log k))}{g^*(k)}
\]
\[
= \lim_{k \to \infty} \frac{2\sigma \sqrt{k \log k}}{g^*(k)} + \lim_{k \to \infty} \exp(-2\log k + o(\log k))
\]
\[
= \limsup_{k \to \infty} \frac{\exp(-2\log k + o(\log k))}{g^*(k)}.
\]
We now claim that \(\limsup_{k \to \infty} \frac{\exp(-2\log k + o(\log k))}{g^*(k)} = 0\). Again, it suffices to show that its logarithm goes to \(-\infty\). Proceeding, we obtain that
\[
\lim_{k \to \infty} \log \left( \frac{\exp(-2\log k + o(\log k))}{g^*(k)} \right) = \lim_{k \to \infty} (-2\log k + o(\log k) - \log g^*(k) + \log k + \log \mu)
\]
\[
= \lim_{k \to \infty} (-\log k - \log g^*(k) = -\infty).
\]
As a result, we get
\[
\limsup_{k \to \infty} \frac{\log \left( \frac{E[\pi(p)]}{Mk\mu} \right)}{\log \left( \frac{E[\pi(p^*)]}{Mk\mu} \right)} = 0.
\]
However, this contradicts inequality (D1), which concludes the proof. \(\square\)

**Proof of Theorem 6.** In order to establish (37), we show that pricing at \(p(N) = N\mu - \sigma \sqrt{N \log N}\) leads to the following limiting ratio
\[
\lim_{N \to \infty} \frac{MN\mu - E[\pi(p)]}{M\sigma \sqrt{N \log N}} = 1.
\]
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 \(h(N) \in \omega_+ (\frac{1}{\sqrt{\log N}}) \cap o\left( \frac{N^\epsilon}{\sqrt{\log N}} \right)\) for any \(\epsilon > 0\). Therefore, for any \(p(N) = N\mu - h(N)\sigma \sqrt{N \log N}\), we have
\[
E[\pi(p)]/M = N\mu - h(N)\sigma \sqrt{N \log N}
\]
\[
- \left[ N\mu - h(N)\sigma \sqrt{N \log N} \right] \mathbb{P} \left( V_{1,N}(N) - N\mu \leq -h(N)\sigma \sqrt{N \log N} \right) \tag{D5}
\]
\[
= N\mu - h(N)\sigma \sqrt{N \log N}
\]
\[
- R(g(N)) \left[ N\mu - h(N)\sigma \sqrt{N \log N} \right] \Phi(-h(N)/\sqrt{\log N}), \tag{D6}
\]
where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution and

$$R(g(N)) = \frac{\mathbb{P} \left( \frac{V_{1,N}(N) - N\mu}{\sigma \sqrt{N}} \leq -\frac{g(N)}{\sigma \sqrt{N}} \right)}{\Phi \left( -\frac{g(N)}{\sigma \sqrt{N}} \right)}.$$

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

**Lemma D2 (Asymptotic Expansion of the Tail of a Normal Distribution).**

$$\mathbb{P}(N(0,1) \geq x) = \frac{1}{\sqrt{2\pi}} \frac{\exp(-\frac{x^2}{2})}{x} \sum_{l=0}^{k-1} \frac{(-1)^l(2l - 1)!!}{x^{2l}} + (1)^k D_k,$$

where $D_k \in \Theta_+ \left( \left( \frac{\exp(-\frac{x^2}{2})}{x} \right) \right)$ as $x \to \infty$.

Using the asymptotic expansion in Lemma D2 for $k = 1$, we get

$$\Phi \left( -h(N)\sqrt{\log N} \right) = \frac{1}{\sqrt{2\pi h(N)\sqrt{\log N}}} \exp \left( -\frac{(h(N))^2 \log N}{2} \right) - \Theta \left( \frac{\exp \left( -h(N)\sqrt{\log N} \right)}{(h(N)\sqrt{\log N})^3} \right) = \frac{N^{-\frac{(h(N))^2}{2}}}{\sqrt{2\pi h(N)\sqrt{\log N}}} (1 + o(1)). \quad (D7)$$

Substituting (D7) in (D6) and after some algebra, we get that

$$\frac{MN\mu - \mathbb{E}[\pi(p)]}{M\sigma \sqrt{\log N}} = h(N) + \frac{R(g(N))\mu}{\sqrt{2\pi}\sigma} \sqrt{\frac{N}{\log N}} \frac{N^{-\frac{(h(N))^2}{2}}}{h(N)\sqrt{\log N}} (1 + o(1))$$

$$= h(N) + Q(g(N)) (1 + o(1)),$$

where $Q(g(N)) = \frac{R(g(N))\mu}{\sqrt{2\pi}\sigma} \sqrt{\frac{N}{\log N}} \frac{N^{-\frac{(h(N))^2}{2}}}{h(N)\sqrt{\log N}}$. Notice that we have

$$\log(Q(g(N))) = \log \left( \frac{R(g(N))\mu}{\sqrt{2\pi}\sigma} \right) + \frac{1}{2} \log N - \frac{1}{2} \log \log N - \frac{(h(N))^2}{2} \log N - \log h(N) - \frac{1}{2} \log \log N + o(1). \quad (D8)$$

We now state Cramer’s moderate deviation principle.

**Lemma D3.** Given Assumption 4, then for any $g(N) \in \omega_+ (\sqrt{N}) \cap o(N^{2/3})$, we have $R(g(N)) \to 1$ as $N \to \infty$.

This result dates back to Cramér (1938) (see Petrov (1975), Chapter VIII, Theorem 1). From proposition 6, we have that $g^*(N)$ belongs to Cramer’s moderate deviation regime and clearly so does $g(N) = \sigma \sqrt{N \log N}$. Hence, we have that $R(\sigma \sqrt{N \log N})$, $R(g^*(N)) \to \infty$ as $N \to \infty$. 
It now follows, that for \( p(N) = N\mu - \sigma \sqrt{N \log N} \) we get that \( \lim_{N \to \infty} \log(Q(\sigma \sqrt{N \log N})) = -\infty \).
This in return implies that \( \lim_{n \to \infty} Q(\sigma \sqrt{N \log N}) = 0 \) and

\[
\lim_{N \to \infty} \frac{MN\mu - \mathbb{E}[\pi(p)]}{M\sigma \sqrt{N \log N}} = 1.
\]

Let \( h^*(N) = g^*(N)/(\sigma \sqrt{N \log N}) \). In order to establish (37), we show by contradiction that if \( \lim_{n \to \infty} h^*(N) \neq 1 \) then the ratio

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

First, assume that \( \bar{h} = \liminf_{N \to \infty} h^*(N) > 1 \). It follows from (D8) and Lemma D3, that \( \limsup_{N \to \infty} \log Q(g^*(N)) = -\infty \) which implies that \( \limsup_{n \to \infty} Q(g^*(N)) = 0 \). Therefore, we get that

\[
\limsup_{N \to \infty} \frac{MN\mu - \mathbb{E}[\pi(p^*)]}{M\sigma \sqrt{N \log N}} = \bar{h} > 1.
\]

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 \( \limsup_{N \to \infty} h^*(N) < 1 \) then from (D8) and Lemma D3, we get that \( \liminf_{N \to \infty} \log(Q(g^*(N))) = +\infty \) which implies that \( \liminf_{N \to \infty} Q(g^*(N)) = +\infty \). Therefore, we get that

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

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 (36) and (37). \( \square \)