

# Price and Quality Decisions of a Service Provider Under Heterogeneous Demand

**Başak Altan\***  
Özyeğin University

**Pelin Atahan\*\***  
Özyeğin University

## Abstract

A monopolist service provider's quality and price decisions are analyzed in a vertically differentiated market where customers demand different quantities of a service. We find that depending on the relative sizes of the market segments and the difference in the valuations of different customers, the service provider may find it optimal to either offer a non-discriminating service or a discriminating service serving only high-valuation customers. The service provider never finds it optimal to serve the market segments that have low-valuation for quality when the discrimination strategy is optimal.

*Keywords: Vertical differentiation, service pricing, asymmetric information*

*JEL Classification: D82, L11*

## Hizmet Sağlayıcının Ayrışık Talep Durumunda Fiyat ve Kalite Kararları

### Özet

Tekelci bir hizmet sağlayıcısının kalite ve fiyat konusundaki kararları, müşterilerin farklı miktarlarda (veya hacimlerde) hizmet talep ettiği dikey bir farklılaştırma piyasasında analiz edilmiştir. Piyasa segmentlerinin nispi büyüklüğüne ve farklı müşterilerin değerlemelerindeki farklılıklarına bağlı olarak, hizmet sağlayıcı, ya ayrımcı olmayan bir hizmet, ya da hizmeti yüksek değerlendiren müşterisine ayrımcı bir hizmet sunmayı tercih etmektedir. Hizmet sağlayıcısı, ayrımcılık stratejisi optimal olduğu durumlarda, düşük kalite değerlemesine sahip pazar segmentlerine hizmet etmeyi hiçbir zaman tercih etmemektedir.

*Anahtar Kelimeler: Dikey farklılaştırma, hizmet fiyatlandırma, asimetrik bilgi*

*JEL Sınıflaması: D82, L11*

---

\* Başak Altan Economics, Ozyegin University, Istanbul, Turkey [basak.altan@ozyegin.edu.tr](mailto:basak.altan@ozyegin.edu.tr) Phone: +90 216 564 9506 Fax: +90 216 564 90 43

\*\* Pelin Atahan Management Information Systems, Ozyegin University, Istanbul, Turkey [pelin.atahan@ozyegin.edu.tr](mailto:pelin.atahan@ozyegin.edu.tr) Phone: +90 216 564 9506 Fax: +90 216 564 90 43

## Introduction

In this study we examine the quality and price decisions of a monopolist service provider. We consider a service, such as a cloud computing service that can support various processes of different firms (Armbust et al. 2010). Public cloud computing service providers offer on-demand computing services that are available over the internet for storing, managing, and processing data (Zhou et al. 2010). These services are offered to both individuals and firms. We focus on the business-to-business market, where the buyers are typically made up of a variety of firms that differ in their processing needs. Although we motivate our model using the cloud computing market, it can be applied to other settings that include service provisions, such as a content distribution setting, where customers are heterogeneous both in their demand and valuation for quality, and quantity demanded is not divisible.

A service provider can offer multiple services that are differentiated based on quality levels. Quality may refer to the level of some attribute or some scalar metric representing a vector of attributes (e.g., functionality, processing speed, reliability, security, etc.). Although all customers prefer higher quality service, they may value quality at different rates.

Although cloud computing services can be provisioned on a need basis to the customers, service providers typically offer discounts based on committed usage (Weins 2017). The quantity of services a buyer demands depends on the types of processes that will be supported. The service provider may differentiate the quality for each of the different versions of the services provided in order to price discriminate customers with heterogeneous valuations of quality.

An interesting aspect of this market is that valuation for quality is related to the quantity of service demanded. Buyers that have high demand are more likely to use the services for their main operations or for some customer-facing operations that require real-time processing. Hence, such customers will tend to have a higher valuation for quality. Consider, Expedia, which is a leading online travel company as an example. Expedia relies on cloud computing services for real-time processing of data streams coming from Expedia's global network of websites, primarily clickstream, user interaction, and supply data. It is natural that Expedia would value quality highly.

Unlike buyers that have high demand, buyers that have a low demand are likely to have a lower valuation for quality. Low-demand buyers may not be using cloud services for their main operations. Many buyers use cloud computing for smaller scale tasks such as hosting their website or social media monitoring or for other peripheral activities. Since the reliance of such companies on the services is limited, quality may not play a critical role. For instance, the flavor company McCormick provides a service, FlavorPrint, that generates individual flavor profiles for consumers based on culinary tastes. McCormick may not require a very high quality for such a peripheral service.

The quantity of the services demanded alone cannot fully determine the valuation of different buyers for quality. Valuation for quality also depends on the nature of the processes the received services support. For instance, a high-demand buyer using the services for non-urgent backend operations may not value quality highly. Consider as an example, the Kellogg Company, which uses cloud services to analyze its promotional costs, analyzing large volumes of data offline to predict the more effective promotional activities. Although the demand for services by both Expedia and Kellogg may be high, quality in terms of processing speed and availability will be much more important for Expedia compared to Kellogg. Alternatively, some low-demand buyers may be smaller firms, such as startups, that use the services for

mission-critical or financial operations or for some customer-facing applications that require real-time processing. In such cases, the buyer may have a higher valuation for quality despite a low demand for services.

This paper studies the price-quality menu of a monopolist service provider in such a business-to-business market where customers are heterogeneous in their valuation for quality and quantity they demand. In terms of quantity demanded, there are two types of customers: high-demand and low-demand. The service provider can design and provide different versions of the service for customers based on their demand levels. In a business-to-business setting, the services designed for low-demand customers will not meet the needs of a customer that has a high demand. Those high-demand customers will have to get the rest of the demanded services from other sources, for instance, by building their own computing infrastructure. However, this would increase the complexity of managing their processes and, hence, increase their costs significantly. Therefore, we assume that high-demand customers do not receive a positive utility from a service that does not meet their demand requirement.

Prior research has found that in order to extract more consumer-surplus, a service provider may offer several qualities of the same product or service (Mussa and Rosen 1978, Besanko et al. 1987). These studies have assumed that marginal costs are increasing and convex in quality. They have found that the optimal policy is a discrimination policy that separates markets. In the lower type consumers receive a product with less than socially optimal quality, and under certain circumstances, the lowest valuation consumers may be priced out of the market. Gabszewicz et al. (1986) find that depending on the dispersion of customers' income, the optimal policy for the service provider is either to segment the market completely or offer only the highest quality product. Itoh (1983) studies how a monopolistic price schedule changes when the degree of product differentiation becomes finer assuming a constant marginal cost of quality. He finds that when the number of the types of commodities supplied by the monopolistic producer changes, the price of the higher quality goods changes by the same amount, whereas the price of the lower quality goods is not affected at all. Some work that has studied quality differentiation in a competitive setting with fixed costs has also found that firms choose distinct qualities to differentiate the market (e.g., Shaked and Sutton 1982 and Motta 1993).

A number of papers from the marketing literature study how to optimally price a product line in order to price discriminate (e.g., Reibstein and Gatignon 1984, Dobson and Kalish 1988, Moorthy 1984). A more recent paper by Anderson and Dana (2009) study monopoly price discrimination and characterize the conditions under which price discrimination is profitable. They show that price discrimination may not be always profitable when there is an upper limit to quality. They show that for price discrimination to be profitable the percentage change in surplus (i.e., consumers' total willingness to pay, less the firm's costs) associated with a product upgrade should be increasing in consumers' willingness to pay.

Nonlinear pricing with respect to quantity demanded is a common practice among sellers. Buchanan (1952) studies the theory behind quantity discounts and its effect on the welfare of the customers. He establishes that quantity discounts are limited to certain types of consumers to encourage them to purchase greater quantities at the same average price. Dolan (1987) studies the motivations behind quantity discounts from a managerial perspective and shows that the pricing structure of a firm is important and the quantity discounts differ with respect to the competitiveness in an industry. Lu et. al. (2014) examine a dynamic quantity-based price differentiation and establish the comparative advantage of quantity-based price differentiation with respect to model parameters. The numerical study in Lu et. al. (2014) shows that when

the product has a low unit ordering cost and high utility, the manufacturer can experience a significant profit gain as a result of shifting from uniform pricing to quantity-based pricing. Gu and Yang (2018) empirically study the effect of quantity discounts on consumer buying behavior and show that some consumers perceive quantity discounts as gains, yet others perceive quantity discounts as losses. Hence, the empirical analysis established that manufacturers can benefit from consumers' quantity-discount-dependent preferences by incorporating such preferences in designing nonlinear pricing schemes.

In this paper, we identify the conditions in which price discrimination via sorting the customers based on valuation for quality is profitable for the monopolist service provider in a business-to-business market. We also evaluate the quality and price implications of the product-line choice of the service provider. This model applies to industries where customers are heterogeneous both in their valuation for quality and demand, and make purchasing agreements for the service in bulk quantities. In our model, similar to Ronnen (1991) and Lehman-Grube (1997), quality costs are incurred during the upfront investment and the costs of quality improvement are increasing and convex. When the monopolist decides to serve a market, the monopolist invests in the hardware and software configurations that will cover the market upfront.

Our results indicate that it is not always optimal for the service provider to sort between customers when the share of high-valuation customers in the market is sufficiently small. In this case, the service provider does not sort low-demand customers with respect to valuation. However, to discourage low-demand customers from buying a high-quantity service, the service provider reduces the price of the bundle offered to them depending on the ratio of the quantities of the bundles. For some level of high-valuation customers, the service provider finds it optimal to sort the customers. Under the full sorting strategy, the quality offered to the low-valuation customers is reduced to discourage high-valuation customers from buying the low-quality service. Whenever full sorting is the optimal strategy, the service provider finds it suboptimal to serve low-valuation customers.

The rest of the paper proceeds as follows. We introduce the model in Section 2. In Section 3, we provide the analysis considering the full sorting and partial sorting strategies and the optimal strategy of the service provider. Concluding remarks are provided in Section 4. All proofs are relegated to the Appendix.

## Model

We consider a service provider in a market with customers that demand high- or low-quantity of services. We refer to these customers as *high-demand* and *low-demand* customers, respectively. The quantity demanded by a customer is represented as  $q^j$  (superscripts are used to denote the quantity demanded by customers), where  $j \in \{h, l\}$ .

Let  $z$  be the quality of the service, where  $z \in [0, \infty)$ . For a given quality of service  $z$ , the benefit that customer  $n$  with demand  $q^j$  derives from the provided services is described by the following utility function:

$$U_{\theta}^j(z, q) = \begin{cases} \theta z \bar{q}^j & \text{if } q \geq \bar{q}^j \\ 0 & \text{otherwise} \end{cases}$$

where  $\bar{q}^j$  is the minimum demand of consumer of demand type  $j$  and  $\theta$  is the consumer's value of quality,  $\theta \in \{\theta_l, \theta_h\}$  and  $\theta_h > \theta_l$  (subscripts are used to denote the customer's valuation for quality).

The cost of the bundle with quality  $z$  and the quantity  $q$  to the service provider is:

$$C(z, q) = cz^2q^l, \text{ with } c > 0.$$

The service provider incurs a cost for each quality-quantity bundle  $(z, q)$  introduced to the market. We assume that the investment costs cover the market and the service provider does not incur additional unit costs from providing the service.

When we consider the customer heterogeneity in valuation of quality and demanded quantity, we end up with four types of customers: (1) *high-valuation – high-demand*, (2) *high-valuation – low-demand*, (3) *low-valuation – high-demand*, and (4) *low-valuation – low-demand*. The proportions of each of these types of customers are represented as (1)  $n_h^h$ , (2)  $n_h^l$ , (3)  $n_l^h$ , (4)  $n_l^l$ , respectively. The number of the customers in the market is normalized to 1. As a result,  $n_h^h + n_h^l + n_l^h + n_l^l = 1$ . In our setting, customers that have high demand are more likely to have a high valuation. Similarly, customers that have low demand are more likely to have a low valuation. This dependency is represented as follows:  $n_h^h > n_l^h$  and  $n_h^l > n_l^l$ . Throughout the paper, we represent the share of high-demand customers as  $n^h$  and the share of low-demand customers as  $n^l$ , i.e.,  $n^h = n_h^h + n_l^h$  and  $n^l = n_h^l + n_l^l$ .

The service provider may provide different versions of the service to *high-* and *low-demand* customers that may be further differentiated by the levels of quality offered. Customers will select the service with the quality and quantity that will provide the highest utility. The service provider does not know a given customer's valuation or demand. However, the service provider knows the distribution of customers in the market. The service provider can set different prices and qualities for customers that have different valuations and different levels of demand.

The timing of the interaction between the service provider and the customers is as follows. First, the service provider determines the quality of the services  $z$  and the quantity she plans to provide and makes the necessary investment. Then, the service provider announces the price, quality, and quantity of the services she is offering to the market.

## Analysis

The service provider may choose to produce different quality versions of the same product in an attempt to increase profits by sorting the customers with respect to their valuation for the service. First, we study two different pricing strategies for the service provider: (i) full sorting and (ii) partial sorting where the service provider follows a non-discriminatory strategy for at least one group of customers. Then, we identify the set of parameters supporting each strategy. When the difference in the customers' valuation for the quality of the service is sufficiently high, the optimal bundles naturally sort the customers with respect to demand. We identify the conditions where the service provider sorts with respect to demand when this difference is lower.

### Full Sorting Strategy

We begin our analysis by considering a full sorting strategy, where the service provider targets different services to each market segment. The service provider announces four price-quantity-quality bundles,  $(p, q, z)$ . The first bundle  $(p_h^h, q_h^h, z_h^h)$  targets high-valuation – high-demand

customers ( $n_h^h$ ). The second bundle  $(p_h^l, q^l, z_h^l)$  targets high-valuation – low-demand customers ( $n_h^l$ ). The third bundle  $(p_l^h, q^h, z_l^h)$  targets low-valuation – high-demand customers ( $n_l^h$ ). And the final bundle  $(p_l^l, q^l, z_l^l)$  targets low-valuation – low-demand customers ( $n_l^l$ ).

The profit of the service provider becomes

$$\Pi = \sum_{i,j \in \{h,l\}} (n_i^j p_i^j - c(z_i^j)^2 q_i^j)$$

The objective of the service provider is to maximize the profit subject to the participation and incentive compatibility constraints of the customers.

$$\max\{\Pi\}$$

subject to

$$IR_i^j \quad U_i^j(z_i^j, q^j) - p_i^j \geq 0 \quad \text{for all } i, j$$

$$IC_i^j - V: U_i^j(z_i^j, q^j) - p_i^j \geq U_i^j(z_{i'}^j, q^j) - p_{i'}^j \text{ for all } i, j, i' \neq i$$

$$IC_i^l - D: U_i^l(z_i^l, q^l) - p_i^l \geq U_i^l(z_{i'}^h, q^h) - p_{i'}^h, \text{ for all } i, i'$$

where  $IC_i^j - V$  is the incentive compatibility constraint of high- and low-demand customers with respect to valuation and  $IC_i^l - D$  is the incentive compatibility constraint of low-demand customers with respect to demand. Note that, since high-demand customers do not receive positive utility from getting a low-quantity service, there is no need for their incentive compatibility constraint with respect to demand.

The service provider identifies a price-quality strategy so that the high-valuation customers do not buy the bundle designed for the low-valuation customers. That is, the surplus that a high-valuation customer gets from his own bundle is no less than the surplus he would get from the bundle designed for a low-valuation customer. Additionally, the price-quality strategy ensures that customers buy the bundles designed for their own demand levels. To sort high- and low-valuation customers, the service provider reduces the prices offered to high-valuation customers compared to what the prices would be if she did not serve the low-valuation customers, which affects the qualities of the offered bundles. To prevent high-valuation – low-demand customers purchase a bundle with high quantity, the service provider may need to reduce the price offered to them.

**Lemma 1.** *The low-valuation – low-demand customers never purchase a bundle with high quantity when at least one type of customer is sorted.*

We establish in Lemma 1 that even when the demand is not verifiable, the low-valuation – low-demand customers always purchase a low-quantity bundle when the service provider sorts at least one type of customer. This implies that the incentive compatibility constraint of low-valuation – low-demand customers with respect to demand never binds.

**Lemma 2.** *When the service provider price discriminates, the optimal quality levels are:*

- $z_h^h = \frac{n_h^h \theta_h}{2c}$
- $z_h^l = \frac{n_h^l \theta_h}{2c}$

$$\bullet \quad z_l^h = \begin{cases} \frac{n_l^h \theta_l - n_h^h (\theta_h - \theta_l)}{2c} & \text{if } \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h} \\ 0 & \text{otherwise} \end{cases}$$

$$\bullet \quad z_l^l = \begin{cases} \frac{n_l^l \theta_l - n_h^l (\theta_h - \theta_l)}{2c} & \text{if } \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^l} \\ 0 & \text{otherwise} \end{cases}$$

When we consider the quality levels offered to the high-valuation customers, we see that the quality levels are independent of the market size and the valuation level of the low-valuation customers for both high- and low-demand customers. However, for the low-valuation customers, the quality is reduced further than what it would be if high-valuation customers did not exist in the market, in order to discourage high-valuation customers from buying the bundle offered to the low-valuation customers. The magnitude of the reduction in quality offered to low-valuation customers depends on the relative size of the market segments and the difference between the valuation levels of the high- and low-valuation customers. For larger shares of high-valuation customers in the market, this reduction is larger. In fact, when the size of high-valuation customer segments is sufficiently large relative to the low-valuation segments, the service provider finds it suboptimal to serve one or both segments of the low-valuation market. As the relative share of high-valuation customers in the market increases, the conditions  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h}$  and  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^l}$  hold. From now on, we will refer to the relative difference between the valuation levels of the high- and low-valuation customers,  $\frac{\theta_h - \theta_l}{\theta_l}$ , as the *relative difference in valuations* for brevity.

**Proposition 1.** *The optimal full sorting strategy is as follows:*

- i. When  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h}$ , the service provider offers four different bundles to the market: the first bundle ( $s_1$ ) targets the high-valuation – high-demand customers, the second bundle ( $s_2$ ) targets the high-valuation – low-demand customers, the third bundle ( $s_3$ ) targets the low-valuation – high-demand customers, and the last bundle ( $s_4$ ) targets the low-valuation – low-demand customers,
- ii. When  $\frac{n_l^h}{n_h^h} < \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^l}$ , the service provider offers three different bundles to the market: the first bundle ( $s_5$ ) targets the high-valuation – high-demand customers, the second bundle ( $s_3$ ) targets the high-valuation – low-demand customers, and the last bundle ( $s_4$ ) targets the low-valuation – low-demand customers,
- iii. When  $\frac{\theta_h - \theta_l}{\theta_l} > \frac{n_l^l}{n_h^l}$ , the service provider offers two different bundles to the market: the first bundle ( $s_5$ ) targets the high-valuation – high-demand customers and the other bundle ( $s_6$ ) targets the high-valuation – low-demand customers.

We observe that the high-valuation – low-demand customers do not purchase the bundle designed for high-valuation – high-demand customers when  $\frac{\theta_h - \theta_l}{\theta_l} > \frac{n_l^h}{n_h^h}$  holds (in cases *ii.* and *iii.*). Otherwise (in case *i.*), these customers do not purchase the bundle designed for high-valuation – high-demand customers if  $\frac{q^l}{q^h} \geq \frac{n_l^h \theta_l - n_h^h (\theta_h - \theta_l)}{n_l^l \theta_l - n_h^l (\theta_h - \theta_l)}$  holds. However, in this case if  $\frac{q^l}{q^h} \geq \frac{n_l^h \theta_l - n_h^h (\theta_h - \theta_l)}{n_l^l \theta_l - n_h^l (\theta_h - \theta_l)}$  is not satisfied, then the service provider must lower the price of  $s_2$  to make the high-valuation – low-demand customers not purchase the bundle for the high-valuation – high-

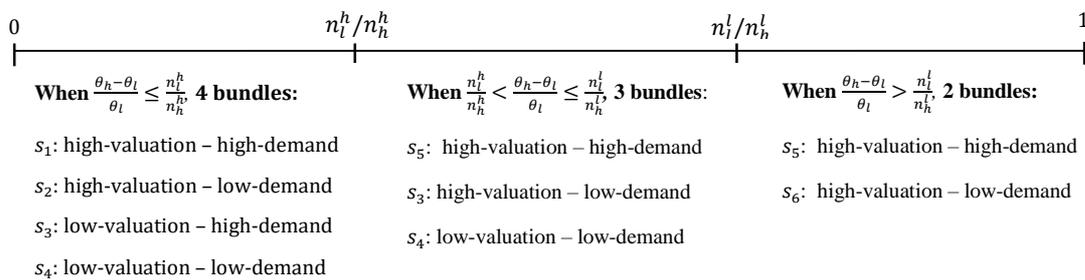
demand customers. We present the quality, price and quantity levels for each bundle in Table 1.

Table 1 – Full Sorting Strategy – Price, quality, and quantity levels

Bundle	Price (P)	Quality (z)	Quantity(q)
$S_1$	$\frac{n_h^h((\theta_h)^2+(\theta_h-\theta_l)^2)-n_l^h\theta_l(\theta_h-\theta_l)}{2c} q^h$	$\frac{n_h^h\theta_h}{2c}$	$q^h$
$S_2$	$\frac{n_h^l(\theta_h^2+(\theta_h-\theta_l)^2)-n_l^l\theta_l(\theta_h-\theta_l)}{2c} q^l$ , if $\frac{q^l}{q^h} \geq \frac{n_l^h\theta_l-n_h^l(\theta_h-\theta_l)}{n_l^l\theta_l-n_h^l(\theta_h-\theta_l)}$ $\frac{n_l^l\theta_h^2q^l-(\theta_h-\theta_l)(n_l^h\theta_l-n_h^h(\theta_h-\theta_l))q^h}{2c}$ , otherwise	$\frac{n_l^l\theta_h}{2c}$	$q^l$
$S_3$	$\frac{n_l^h(\theta_l)^2-n_h^h\theta_l(\theta_h-\theta_l)}{2c} q^h$	$\frac{n_l^h\theta_l-n_h^h(\theta_h-\theta_l)}{2c}$	$q^h$
$S_4$	$\frac{n_l^l(\theta_l)^2-n_h^l\theta_l(\theta_h-\theta_l)}{2c} q^l$	$\frac{n_l^l\theta_l-n_h^l(\theta_h-\theta_l)}{2c}$	$q^l$
$S_5$	$\frac{n_h^h(\theta_h)^2}{2c} q^h$	$\frac{n_h^h\theta_h}{2c}$	$q^h$
$S_6$	$\frac{n_h^l(\theta_h)^2}{2c} q^l$	$\frac{n_h^l\theta_h}{2c}$	$q^l$

Proposition 1 identifies the optimal bundles under the full sorting strategy of the service provider. These bundles offered to target different customer segments are summarized with respect to the proportion of high-valuation customers in Figure 2.

Figure 2. Full Sorting Strategy - Bundles



Proposition 1 highlights that under the full sorting strategy, when the relative difference in valuations is sufficiently low, the service provider’s optimal sorting strategy is to fully discriminate the market, i.e., offer four different price, quality, quantity bundles targeting each market segment. In this case, when this relative difference is even lower, the service provider has to reduce the price offered to high-valuation – low-demand customers further to prevent them from purchasing a bundle with high-quantity.

When the relative difference in valuations is higher, the service provider is better off serving both the high- and low-valuation segments of the low-demand market rather than serving only the high-valuation segment of the high-demand market. Finally, when the share of the high-valuation customers is sufficiently large, the service provider only serves the high-valuation customers, offering services designed for both high- and low-demand segments. As a result, the service provider can charge higher prices for the high-valuation customers, extracting a higher surplus compared to the case where she also serves the low-valuation customers. When the low-valuation segment is ignored, the service provider does not need to reduce the prices of the high-valuation services to ensure that the high-valuation customers do not buy the services targeted to the low-valuation customers. To summarize, in the full sorting strategy, as

the share of the high-valuation customers gets larger, the service provider does not find it optimal to serve some or all of the low-valuation customers.

**Corollary 1.** *The profit of the service provider is:*

$$\Pi = \begin{cases} \frac{(\theta_h)^2((n_h^h)^2q^h + (n_h^l)^2q^l)}{4c} + \frac{(n_h^l(\theta_h - \theta_l) - n_l^l\theta_l)^2q_l}{4c} + \frac{(n_h^h(\theta_h - \theta_l) - n_l^h\theta_l)^2q_l}{4c} - A & \text{if } \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h} \text{ and } \frac{q^l}{q^h} < \frac{n_l^h\theta_l - n_h^h(\theta_h - \theta_l)}{n_l^l\theta_l - n_h^l(\theta_h - \theta_l)} \\ \frac{(\theta_h)^2((n_h^h)^2q^h + (n_h^l)^2q^l)}{4c} + \frac{(n_h^l(\theta_h - \theta_l) - n_l^l\theta_l)^2q_l}{4c} + \frac{(n_h^h(\theta_h - \theta_l) - n_l^h\theta_l)^2q_l}{4c} & \text{if } \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h} \text{ and } \frac{q^l}{q^h} \geq \frac{n_l^h\theta_l - n_h^h(\theta_h - \theta_l)}{n_l^l\theta_l - n_h^l(\theta_h - \theta_l)} \\ \frac{(\theta_h)^2((n_h^h)^2q^h + (n_h^l)^2q^l)}{4c} + \frac{(n_h^l(\theta_h - \theta_l) - n_l^l\theta_l)^2q_l}{4c} & \text{if } \frac{n_l^h}{n_h^h} < \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^l} \\ \frac{(\theta_h)^2((n_h^h)^2q^h + (n_h^l)^2q^l)}{4c} & \text{if } \frac{\theta_h - \theta_l}{\theta_l} > \frac{n_l^l}{n_h^l} \end{cases}$$

where  $A = \frac{n_l^h}{2c}((\theta_h - \theta_l)^2(n_l^h q^l - n_h^h q^h) + \theta_l(\theta_h - \theta_l)(n_l^h q^h - n_l^l q^l))$ .

The optimal profit of the service provider for different shares of high-valuation customers in the market is provided in Corollary 1. As the share of high-valuation customers in the market increases, the service provider’s profits increase. This result holds even though the service provider may ignore the low-valuation customers under different market conditions.

### Partial Sorting Strategy

When the service provider prefers a non-discriminatory strategy for at least one group of customers, then there are three different cases to consider: (1) sorts only the high-demand customers, (2) sorts only the low-demand customers, and (3) does not sort customers with respect to valuation. In the first case, there will be three different bundles offered to the market: one bundle targeting high-valuation – high-demand customers, a second bundle targeting low-valuation – high-demand customers and the third bundle targeting all low-demand customers. In the second case, there will be again three different bundles offered to the market: one bundle targeting high-valuation – low-demand customers, a second bundle targeting low-valuation – low-demand customers, and the third bundle targeting all high-demand customers. In the third case, there will be only two different bundles offered: one targeting all the high-demand customers and the second one targeting all the low-demand customers.

#### When only high-demand customers are sorted:

The service provider offers three bundles:  $(p_h^h, q^h, z_h^h)$ ,  $(p_l^h, q^h, z_l^h)$  and  $(p^l, q^l, z^l)$  targeting high-valuation – high-demand, low-valuation – high-demand, and all the low-demand customers, respectively. In this case, the profit of the service provider becomes:

$$\Pi = n_h^h p_h^h + n_l^h p_l^h + (n_h^l + n_l^l) p^l - \sum_{z,q} C(z, q),$$

where  $\sum_{z,q} C(z, q) = C(z_h^h, q^h) + C(z_l^h, q^h) + C(z^l, q^l)$ . Then the objective of the service provider becomes:

$$\max\{\Pi\}$$

subject to

$$IR_i^j \quad \text{for all } i, j$$

$$IC_i^h - V, IC_i^l - D \quad \text{for all } i$$

**When only low-demand customers are sorted:**

The service provider offers three bundles  $(p_h^l, q^l, z_h^l)$ ,  $(p_l^l, q^l, z_l^l)$  and  $(p^h, q^h, z^h)$ , targeting high-valuation – low-demand, low-valuation – low-demand, and all high-demand customers, respectively. In this case, the profit of the service provider becomes:

$$\Pi = n_h^l p_h^l + n_l^l p_l^l + (n_h^h + n_l^h) p^h - \sum_{z,q} C(z, q)$$

where  $\sum_{z,q} C(z, q) = C(z_h^l, q^l) + C(z_l^l, q^l) + C(z^h, q^h)$ . Then the objective of the service provider becomes

$$\max\{\Pi\}$$

subject to

$$IR_i^j \quad \text{for all } i, j$$

$$IC_i^l - V, IC_i^l - D \quad \text{for all } i$$

**When the customers are not sorted with respect to valuation:**

The service provider offers two bundles  $(p^h, q^h, z^h)$  and  $(p^l, q^l, z^l)$  targeting the high- and low-demand customers, respectively. As the first bundle targets all high-demand customers, the service provider must set a price sufficiently low, so that even low-valuation customers prefer purchasing it. The profit of the service provider becomes:

$$\Pi = (n_h^h + n_l^h) p^h + (n_h^l + n_l^l) p^l - \sum_{z,q} C(z, q)$$

where  $\sum_{z,q} C(z, q) = C(z^h, q^h) + C(z^l, q^l)$ . Then the objective of the service provider becomes:

$$\max\{\Pi\}$$

subject to

$$IR_i^j \quad \text{for all } i, j$$

$$IC_i^l - D \quad \text{for all } i$$

We have identified in Lemma 1 the optimal quality levels for different bundles when the service provider sorts the customer. We now establish the quality levels for the case where the service provider does not sort the customers.

**Lemma 3.** *The optimal quality levels when the customers are not sorted are  $z^h = \frac{n^h \theta_l}{2c}$  and  $z^l = \frac{n^l \theta_l}{2c}$ .*

When the service provider does not sort the customers for a service designed for a specific demand level, the high-valuation customers receive a lower quality service compared to the quality they would receive under the sorting strategy. However, the quality of the service low-valuation customers receive is higher under the partial sorting strategy.

**Proposition 2.** *The optimal bundles under the partial sorting strategy are as follows.*

- i. When  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h}$ , the service provider offers two different bundles to the market: the first bundle ( $s_7$ ) targets all the high-demand customers and the second bundle ( $s_8$ ) targets all the low-demand customers,
- ii. When  $\frac{\theta_h - \theta_l}{\theta_l} > \frac{n_l^h}{n_h^h}$ , the service provider offers two different bundles to the market: the first bundle ( $s_5$ ) targets the high-valuation – high-demand customers and the second bundle ( $s_8$ ) targets all the low-demand customers.

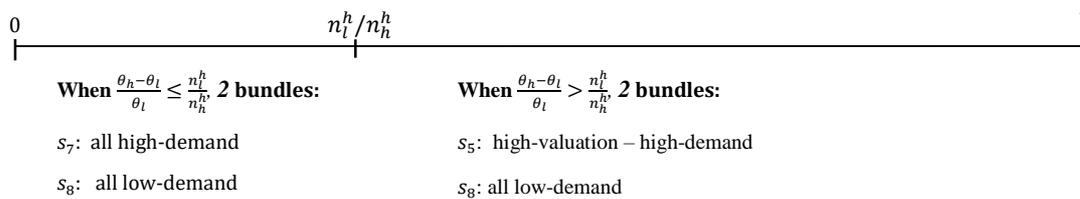
As Lemma 1 establishes, low-valuation customers always purchase the bundle designated for their own segment. We observe that the high-valuation – low-demand customers do not purchase the bundle designed for the high-demand customers when  $\frac{\theta_h - \theta_l}{\theta_l} > \frac{n_l^h}{n_h^h}$  holds (in case ii.). Otherwise (in case i.), these customers do not purchase the bundle designed for high-demand customers if  $\frac{q^l}{q^h} \geq \frac{n^l}{n^h}$  holds. However, in this case if the given condition is not satisfied, then the service provider must lower the price of  $s_8$  to make the high-valuation – low-demand customers not purchase the bundle for the high-demand customers. The price, quality and quantity levels for the different bundles offered under the optimal partial sorting strategy are presented in Table 2.

Table 2 – Partial Sorting Strategy – Price, quality, and quantity levels

Bundle	Price (P)	Quality (z)	Quantity(q)
$s_5$	$\frac{n_h^h(\theta_h)^2}{2c} q^h$	$\frac{n_h^h \theta_h}{2c}$	$q^h$
$s_7$	$\frac{n^h(\theta_l)^2}{2c} q^h$	$\frac{n^h \theta_l}{2c}$	$q^h$
$s_8$	$\frac{n^l(\theta_l)^2}{2c} q^l$ , if $\frac{q^l}{q^h} \geq \frac{n^l}{n^h}$ $\frac{n^l \theta_h \theta_l q^l - (\theta_h - \theta_l) \theta_l n^h q^h}{2c}$ , otherwise	$\frac{n^l \theta_l}{2c}$	$q^l$

The different bundles targeting different customer segments are presented in Figure 3.

Figure 3. Partial Sorting Strategy - Bundles



When the service provider follows a partial sorting strategy, there are only two optimal strategies depending on the distribution of the customers. When the proportion of high-valuation customers is sufficiently high, the service provider does not find it profitable to sell a bundle to low-valuation – high-demand customers. There will be two bundles in the market: one targeting high-valuation – high-demand customers and one targeting all low-demand customers. When the proportion of high-valuation customers is low, it is not optimal for the service provider to offer a specific product with a higher quality for this customer group as the additional revenue will not cover the production cost. Hence, it is not optimal for the service provider to sort the low-demand customers, when no-sorting is the optimal strategy for the high-demand customers. As the low-demand customers are more likely to be low-valuation

compared to the high-demand customers, the additional profit margin generated from a higher quality product is even lower for low-demand customers. In both of these cases, the low-demand customers are sorted with respect to valuation. To prevent low-demand customers from buying a high-demand bundle the price should be reduced when the ratio of low- to high-quantity is less than the ratio of low- to high-demand customers in the market.

**Corollary 2.** *The profit of the service provider is:*

$$\Pi = \begin{cases} \frac{(n^h \theta_l)^2 - 2n^h n^l (\theta_h - \theta_l) \theta_l}{4c} q^h + \frac{(n^l)^2 \theta_l (2\theta_h - \theta_l)}{4c} q^l & \text{if } \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h} \text{ and } \frac{q^l}{q^h} < \frac{n^l}{n^h} \\ \frac{(n^h \theta_l)^2}{4c} q^h + \frac{(n^l \theta_l)^2}{4c} q^l & \text{if } \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h} \text{ and } \frac{q^l}{q^h} \geq \frac{n^l}{n^h} \\ \frac{(n_h^h \theta_h)^2}{4c} q^h + \frac{(n^l \theta_l)^2}{4c} q^l & \text{if } \frac{\theta_h - \theta_l}{\theta_l} > \frac{n_l^h}{n_h^h} \end{cases}$$

**Optimal Strategy**

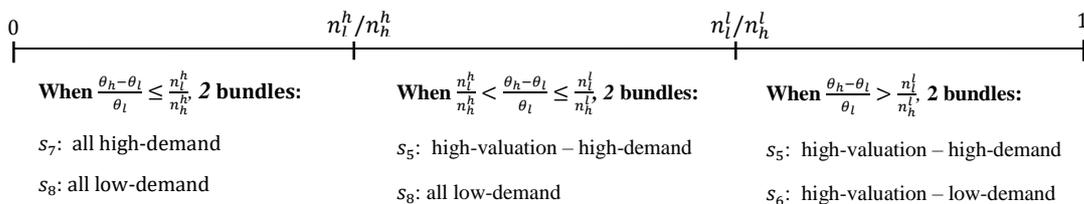
Finally, we compare the profits of the service provider under the full and partial sorting strategies to determine the optimal one for the service provider. Theorem 1 presents the bundles offered under the optimal strategy.

**Theorem 1.** *The optimal bundles that the service provider offers to the customers are:*

- i. When  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h}$ , the service provider offers two different bundles to the market: the first bundle ( $s_7$ ) targets all the high-demand customers and the second bundle ( $s_8$ ) targets all the low-demand customers,
- ii. When  $\frac{n_l^h}{n_h^h} < \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^l}$ , the service provider offers two different bundles to the market: the first bundle ( $s_5$ ) targets the high-valuation – high-demand customers and the second bundle ( $s_8$ ) targets all the low-demand customers,
- iii. When  $\frac{\theta_h - \theta_l}{\theta_l} > \frac{n_l^l}{n_h^l}$ , the service provider offers two different bundles to the market: the first bundle ( $s_5$ ) targets the high-valuation – high-demand customers and the second bundle ( $s_6$ ) targets the high-valuation – low-demand customers.

Figure 4 presents the different bundles offered to different customer segments under the optimal strategy. The price, quality and quantity levels of these bundles can be found in Tables 1 and 2.

Figure 4. Optimal Strategy - Bundles



In the optimal strategy, we observe three different scenarios depending on the relative sizes of the market segments and the difference between the valuation levels. When the share of the high-valuation customers is sufficiently small, the service provider offers a single quality service for each demand level and serves both the high- and the low-valuation segments. When

the share of the high-valuation customers is higher, the service provider does not (to sort)? serve? the low-demand market, and in the high-demand market only serves the high-valuation customers. In this case, the share of the high-valuation customers in the high-demand market is large enough to make it beneficial to only serve them. In these two cases, to ensure that low-demand customers buy the bundle designed for them, its price should be reduced when the ratio of low- to high-quantity is less than the ratio of low- to high-demand customers in the market. When the share of the high-demand customers is larger, then the service provider finds it optimal to only serve the high-valuation customers in both the high- and low-demand segments. The share of the high-demand customers in this case indicates that the share of the high-valuation customers in the low-demand market is also large enough to justify serving only the high-valuation customers in that market.

**Corollary 3.** *The profit of the service provider is:*

$$\Pi = \begin{cases} \frac{(n^h \theta_l)^2 - 2n^h n^l (\theta_h - \theta_l) \theta_l}{4c} q^h + \frac{(n^l)^2 \theta_l (2\theta_h - \theta_l)}{4c} q^l & \text{if } \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h} \text{ and } \frac{q^l}{q^h} < \frac{n^l}{n^h} \\ \frac{((n_h^h + n_l^h) \theta_l)^2}{4c} q^h + \frac{((n_h^l + n_l^l) \theta_l)^2}{4c} q^l & \text{if } \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h} \text{ and } \frac{q^l}{q^h} \geq \frac{n^l}{n^h} \\ \frac{(n_h^h \theta_h)^2}{4c} q^h + \frac{((n_h^l + n_l^l) \theta_l)^2}{4c} q^l & \text{if } \frac{n_l^h}{n_h^h} < \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^l} \\ \frac{(\theta_h)^2 ((n_h^h)^2 q^h + (n_h^l)^2 q^l)}{4c} & \text{if } \frac{\theta_h - \theta_l}{\theta_l} > \frac{n_l^l}{n_h^l} \end{cases}$$

The profit of the service provider is presented in Corollary 3. We observe that the optimal profit is non-decreasing with respect to the size of the high-valuation customers. Additionally, as quantity demand(ed)? increases so does the profit. When the low-valuation customers are more likely to have a high demand,  $n_l^h/n_h^h$  increases and  $n_l^l/n_h^l$  decreases. When  $\frac{n_l^h}{n_h^h} < \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^l}$ , even though profit of the service provider decreases, as the low-valuation customers are more likely to have a high demand, the range of the relative difference in valuations supporting this outcome shrinks. When the high-valuation customers are more likely to have a high demand,  $n_l^h/n_h^h$  decreases and  $n_l^l/n_h^l$  increases. Hence, the service provider offers two bundles, where one targets high-valuation – high-demand customers and the other targets all the low-valuation customers.

## Conclusion

We studied a monopolist service provider's quality and pricing decisions in the existence of heterogeneous customers in a business-to-business market. The market consists of high- or low-demand customers, who are also heterogeneous in their valuations for quality of service. We identified the conditions where price discrimination via sorting the customers is profitable for the service provider. We find that the service provider finds it optimal to sell either the same bundle to both high- and low-valuation customers or only to the high-valuation customers. When the share of the high-valuation customers is low, the service provider prefers to sell to both types of customers.

There are two main reasons why the profits obtained from the full sorting strategy are lower compared to the profits obtained from the partial sorting strategy. First, the reduction in the

price of the bundle offered to high-valuation customers to prevent them from purchasing the bundle offered to the low-valuation customers lowers the profits. Second, the reduction in the quality offered to the low-valuation customers under the full sorting strategy lowers profits. Therefore, the service provider does not sort customers based on valuation. However, in this case, if the ratio of low to high quantity is larger than the ratio of low-demand to high-demand customers, i.e.,  $\frac{q^l}{q^h} \geq \frac{n^l}{n^h}$ , the high-valuation–low-demand customers may obtain higher utility from the bundle targeted to the high-demand customers. When that is the case, the service provider needs to reduce the prices of the low-demand bundle to prevent the high-valuation–low-demand customers from buying the high-demand bundle. When the share of the high-valuation customers is sufficiently large, the service provider can extract a larger surplus by offering a service targeted only to the high-valuation customers with a higher quality and higher price. In this case, offering a service targeting low-valuation customers would require lowering the prices for the high valuation customers, and hence lead to a revenue loss. Furthermore, the service provider would incur additional fixed costs. Therefore, the service provider never finds it optimal to offer a bundle targeted to the low-valuation customers and serves only the high-valuation customers when the full sorting strategy is optimal.

There are numerous directions for extending our analysis. Competition is an essential part of industrial organizations. Duopolistic competition and how its results differ from the monopoly case are worth investigating. In our model, the number of customers in the market is assumed to be fixed. Another interesting extension would be to study how the results would be impacted if we relax this assumption. In our model, the utility function of the consumers is discreet with respect to not only valuation but also quantity. Another direction one can take is to consider continuum of types with respect to valuation as well as quantity. Finally, horizontal product differentiation under competition would be an interesting extension.

## References

- Anderson, E. Dana, J. D., (2009). "When Is Price Discrimination Profitable?", *Management Science*, 55(6), 980-989.
- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D, Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and M. Zaharia (2010). "A View of Cloud Computing", *Communications of the ACM*, 53(4), 50-58.
- Besanko, D., Donnenfeld, S. and L. J. White (1988). "The Multiproduct Firm, Quality Choice, and Regulation", *The Journal of Industrial Economics*, 36(4), 411-429.
- Buchanan, J. (1952). *The Theory of Monopolistic Quantity Discounts*. *The Review of Economic Studies*, 20(3), 199-208.
- Dobson, G., Kalish, S. (1988). "Positioning and Pricing A Product Line", *Marketing Science*, 7(2), 107-125.
- Dolan, R.J. (1987). "Quantity Discounts: Managerial Issues and Research Opportunities" *Marketing Science*, 6(1), 1-27.
- Gabszewicz, J.J., Shaked, a. and J. Sutton(1986)." Segmenting the Market: The Monopolist's Optimal Product Mix", *Journal of Economic Theory*, 39, 273-289.
- Gu, Z., S. Yang (2018). "Quantity-Discount-Dependent Consumer Preferences and Competitive Nonlinear Pricing", *Journal of Marketing Research*, 47(6), 1100-1113.
- Lehmann-Grube, U. (1997). "Strategic choice of quality when quality is costly", *RAND Journal of Economics*, 28(2), 372-384.
- Lu, Y., F. Chen, M. Song, and X. Yan (2014). "Optimal Pricing and Inventory Control Policy with Quantity-Based Price Differentiation", *Operations Research*, 62(3), 512-523.
- Moorthy, K. S. (1984). "Market Segmentation, Self-Selection, and Product Line Design", *Marketing Science*, Vol. 3, No. 4, 288-307. Motta M. (1993) "Endogenous Quality Choice: Price vs. Quantity Competition", *The Journal of Industrial Economics*, 41(2), 113-131.
- Mussa M., and S. Rosen (1978) "Monopoly and Product Quality", *Journal of Economic Theory*, 18, 301-317.
- Reibstein, D. J., Gatignon, H. (1984). "Optimal Product Line Pricing: The Influence of Elasticities and Cross-Elasticities", *Journal of Marketing Research*, 21(3), 259-267.
- Ronnen, U. (1991). "Minimum quality standards, fixed costs, and competition", *RAND Journal of Economics*, 22, 490-504.
- Shaked, A. and J. Sutton (1982), 'Relaxing Price Competition Through Product Differentiation', *Review of Economic Studies*, 49, 3-14.
- Weins, K.(2017) "Cloud pricing comparison: AWS vs. Microsoft Azure vs. Google Cloud vs. IBM Cloud", *InfoWorld*. Available online at

<https://www.infoworld.com/article/3237566/cloud-computing/cloud-pricing-comparison-aws-vs-azure-vs-google-vs-ibm.html>. Last accessed May 2018.

## Appendix

*Proof of Lemma 1.* Suppose that the service provider offers four bundles, one for each type of customers: (1)  $(z_1, p_1, q^h)$  for the high-valuation – high-demand customers; (2)  $(z_2, p_2, q^l)$  for the high-valuation – low-demand customers; (3)  $(z_3, p_3, q^h)$  for the low-valuation – high-demand customers; (4)  $(z_4, p_4, q^l)$  for the low-valuation – low-demand customers.

When the demand of a customer is verifiable, we only need to make sure that bundles are incentive compatible within a demand group. Then, at the optimal bundle we have the followings:

(i) The incentive compatibility constraint of the high-valuation – high-demand customers binds. That is,  $\theta_h z_1 q^h - p_1 = \theta_h z_3 q^h - p_3$  holds.

(ii) The participation constraint of the low-valuation – high-demand customers binds. That is,  $\theta_l z_3 q^h - p_3 = 0$  holds.

(iii) The incentive compatibility constraint of the high-valuation – low-demand customers binds. That is,  $\theta_h z_2 q^l - p_2 = \theta_h z_4 q^l - p_4$  holds.

(iv) The participation constraint of the low-valuation – low-demand customers binds. That is,  $\theta_l z_4 q^l - p_4 = 0$  holds.

The first two constraints imply that  $p_1 = \theta_h z_1 q^h - \theta_h z_3 q^h + \theta_l z_3 q^h$  and  $p_3 = \theta_l z_3 q^h$  hold. The rest implies that  $p_2 = \theta_h z_2 q^l - \theta_h z_4 q^l + \theta_l z_4 q^l$  and  $p_4 = \theta_l z_4 q^l$  hold.

First, consider a low-valuation – high-demand customer. This customer purchases bundle 3 only if her utility from this bundle is no less than her utility from other bundles. We already establish that she does not buy  $(z_1, p_1, q^h)$  which is designed for the high-valuation – high-demand customers. The utility of a low-valuation – high-demand customer from  $(z_2, p_2, q^l)$  which is designed for the high-valuation – low-demand customers is  $\theta_l z_2 q^l - p_2$ . As this utility is negative, a low-valuation – high-demand customer never purchases bundle 2. The utility of a low-valuation – high-demand customer from  $(z_4, p_4, q^l)$  which is designed for the low-valuation – low-demand customers is  $\theta_l z_4 q^l - p_4$ . As this utility is zero, a low-valuation, high-demand customer is indifferent between purchasing bundle 3 and bundle 4.

Second, consider a low-valuation – low-demand customer. This customer purchases  $(z_4, p_4, q^l)$  only if her utility from this bundle is no less than her utility from other bundles. We already establish that she does not buy  $(z_2, p_2, q^l)$ . The utility of a low-valuation – low-demand customer from  $(z_1, p_1, q^h)$  is  $\theta_l z_1 q^l - p_1$ . As this utility is negative, a low-valuation, – high-demand customer never purchases bundle 2. The utility of a low-valuation – high-demand customer from  $(z_3, p_3, q^h)$  is  $\theta_l z_3 q^l - p_3$ . As this utility is zero, a low-valuation – high-demand customer is indifferent between purchasing bundle 3 and bundle 4. Hence, a low-valuation customer always chooses her own bundle even if her demand is not identifiable.

*Proof of Lemma 2.* The service provider sets the quality levels,  $z_j^i$  with  $i, j \in \{h, l\}$ , in order to maximize the expected profit,  $\Pi = n_h^h p_h^h + n_h^l p_h^l + n_l^h p_l^h + n_l^l p_l^l - \sum_{z, q} C(z, q)$ , where

$\sum_{z,q} C(z, q) = c((z_h^l)^2 + (z_l^l)^2)q^h + c((z_h^h)^2 + (z_l^h)^2)q^l$ , subject to  $p_l^i = \theta_l z_l^i q^i$  and  $p_h^i = \theta_h z_h^i q^i - (\theta_h - \theta_l) z_l^i q^i$  for all  $i$ . Then, it is trivial to show that  $\frac{\partial^2 \Pi}{\partial z_l^i{}^2} < 0$  for all  $i, j$ . Then, we can identify the optimal quality levels from the first order conditions. It is trivial to show that the optimal quality of the service targeting the high-valuation – high-demand customers is,  $z_h^h = \frac{n_h^h \theta_h}{2c}$  and that the optimal quality of the service targeting the high-valuation – low-demand customers is  $z_h^l = \frac{n_l^l \theta_h}{2c}$ . The optimal quality of service for the low-valuation – high-demand customers depends on the parameter values. We have  $\frac{\partial \Pi}{\partial z_l^h} = -n_h^h(\theta_h - \theta_l)q^h + n_l^h \theta_l q^h - 2cz_l^h q^h$ . Then,  $z_l^h = \frac{n_l^h \theta_l - n_h^h(\theta_h - \theta_l)}{2c}$ , if  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h}$ . Otherwise,  $z_l^h = 0$ . By the same logic,  $\frac{\partial \Pi}{\partial z_l^l} = -n_l^l(\theta_h - \theta_l)q^h + n_l^l \theta_l q^h - 2cz_l^l q^h$ . Then,  $z_l^l = \frac{n_l^l \theta_l - n_h^l(\theta_h - \theta_l)}{2c}$  if  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^l}$ . Otherwise,  $z_l^l = 0$ .

*Proof of Proposition 1.* First, we observe that  $\frac{n_l^h}{n_h^h} < \frac{n_l^l}{n_h^l}$  holds valid as  $n_h^h > n_l^h$  and  $n_l^l > n_h^l$ . Then, we identify the optimal qualities under different values of  $\frac{\theta_h - \theta_l}{\theta_l}$ . The optimal quality for the high-valuation customers does not change with respect to  $\frac{\theta_h - \theta_l}{\theta_l}$ . For all values of  $\frac{\theta_h - \theta_l}{\theta_l}$ , we have  $z_h^h = \frac{n_h^h \theta_h}{2c}$  and  $z_h^l = \frac{n_l^l \theta_h}{2c}$ . When  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h}$ , we have  $z_l^h = \frac{n_l^h \theta_l - n_h^h(\theta_h - \theta_l)}{2c}$  and  $z_l^l = \frac{n_l^l \theta_l - n_h^l(\theta_h - \theta_l)}{2c}$ . That is, it is optimal to sell all four types of customers. When  $\frac{n_l^h}{n_h^h} \leq \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^l}$ , we have  $z_l^h = 0$  and  $z_l^l = \frac{(n_l^l \theta_l - n_h^l(\theta_h - \theta_l))}{2c}$ . In this case, the service provider prefers to service only the high-valuation segment of the high-demand customers by setting a price sufficiently high for the low-valuation – high-demand customers. When,  $\frac{\theta_h - \theta_l}{\theta_l} \geq \frac{n_l^l}{n_h^l}$ , we have  $z_l^h = 0$  and  $z_l^l = 0$ . In this case, the service provider prefers to service only the high-valuation customers by setting a price sufficiently high for all the low-valuation customers. The prices are determined by the incentive compatibility and participation constraints where  $p_l^i = \theta_l z_l^i q^i$  and  $p_h^i = \theta_h z_h^i q^i - (\theta_h - \theta_l) z_l^i q^i$  for all  $i \in \{h, l\}$ .

When the service provider offers four bundles, one for each type of customers: (1)  $(z_1, p_1, q^h)$  for the high-valuation – high-demand customers; (2)  $(z_2, p_2, q^l)$  for high-valuation – low-demand customers; (3)  $(z_3, p_3, q^h)$  for low-valuation – high-demand customers; (4)  $(z_4, p_4, q^l)$  for the low-valuation – low-demand customers. When the demand of a customer is verifiable the prices of these bundles are (1)  $p_1 = \theta_h z_1 q^h - \theta_h z_3 q^h + \theta_l z_3 q^h$ , (2)  $p_2 = \theta_h z_2 q^l - \theta_h z_4 q^l + \theta_l z_4 q^l$ , (3)  $p_3 = \theta_l z_3 q^h$ , and (4)  $p_4 = \theta_l z_4 q^l$ . Consider a high-valuation – low-demand customer. This customer purchases  $(z_2, p_2, q^l)$  only if her utility from this bundle is no less than her utility from other bundles. We already establish that she does not buy  $(z_4, p_4, q^l)$  which is designed for the low-valuation – low-demand customers. The utility of a high-valuation – low-demand customer from  $(z_1, p_1, q^h)$  is  $\theta_h z_1 q^h - p_1$ . A high-valuation – low-demand customer prefers  $(z_2, p_2, q^l)$  over  $(z_1, p_1, q^h)$  only if  $\theta_h z_2 q^l - p_2 \geq \theta_h z_1 q^h - p_1$  holds. The utility of a high-valuation – low-demand customer from  $(z_3, p_3, q^h)$  is  $\theta_h z_3 q^h - p_3$ . A high-valuation – low-demand customer prefers  $(z_2, p_2, q^l)$  over  $(z_3, p_3, q^h)$  only if

$\theta_h z_2 q^l - p_2 \geq \theta_h z_3 q^h - p_3$  holds. Due to the incentive compatibility constraint of the high-valuation – high-demand customers, we have  $\theta_h z_1 q^h - p_1 = \theta_h z_3 q^h - p_3$ . Hence, if  $\theta_h z_2 q^l - p_2 \geq \theta_h z_1 q^h - p_1$  holds, the high-valuation – high demand customers prefer bundle 2 over the rest. This implies that when  $\frac{q^l}{q^h} \geq \frac{n_l^h \theta_l - n_h^h (\theta_h - \theta_l)}{n_l^l \theta_l - n_h^l (\theta_h - \theta_l)}$  is satisfied, the utility of a high-valuation – low-demand customer from  $(z_2, p_2, q^l)$  is higher than her utility from  $(z_1, p_1, q^h)$ . Then, when  $\frac{q^l}{q^h} \geq \frac{n_l^h \theta_l - n_h^h (\theta_h - \theta_l)}{n_l^l \theta_l - n_h^l (\theta_h - \theta_l)}$  is not satisfied, the service provider must lower the price to  $p_2 = \theta_h z_2 q^l - (\theta_h - \theta_l) z_3 q^h$  to make the bundle  $(z_2, p_2, q^l)$  incentive compatible.

*Proof of Corollary 1.* As the profit of the service provider is  $\Pi = n_h^h p_h^h + n_h^l p_h^l + n_l^h p_l^h + n_l^l p_l^l - (C(z_h^h, q^h) + C(z_h^l, q^l) + C(z_l^h, q^h) + C(z_l^l, q^l))$ , it is trivial to identify the optimal profit under the regions of  $\frac{\theta_h - \theta_l}{\theta_l}$  identified in Proposition 1.

*Proof of Lemma 3.* When the service provider does not discriminate the high-demand customers with respect to quality and offers a service that all the high-demand customers prefer to buy, the maximum price that the service provider is able to charge is determined by the participation constraint of the low-valuation – high-demand customers. That is,  $p^h = \theta_l z^h q^h$ . Likewise, the price of the service that is designed for all the low-demand customers is determined by the participation constraint of the low-valuation – low-demand customers. That is,  $p^l = \theta_l z^l q^l$ . Then, the optimal quality for the service with high quantity is  $z^h = \text{argmax}\{(n_h^h + n_l^h)\theta_l z^h q^h - c(z^h)^2 q^h\}$  and the optimal quality for the service with low quantity is  $z^l = \text{argmax}\{(n_h^l + n_l^l)\theta_l z^l q^l - c(z^l)^2 q^l\}$ . It is trivial to show that  $z^h = \frac{(n_h^h + n_l^h)\theta_l}{2c}$  and  $z^l = \frac{(n_h^l + n_l^l)\theta_l}{2c}$ .

*Proof of Proposition 2.* When the service provider prefers a non-discriminatory strategy for at least one group of customers, there are three possibilities: (1)  $(p^h, q^h, z^h)$  and  $(p^l, q^l, z^l)$ ; (2)  $(p_h^h, q^h, z_h^h)$ ,  $(p_l^h, q^h, z_l^h)$  and  $(p^l, q^l, z^l)$ ; (3)  $(p^h, q^h, z^h)$ ,  $(p_h^l, q^l, z_h^l)$  and  $(p_l^l, q^l, z_l^l)$ . In case (1), the service provider does discriminate at all. All customers that have the same demand buy the same bundle. Then, as Lemma 3 establishes the quality of the service that the high-demand customers purchase becomes  $z^h = \frac{(n_h^h + n_l^h)\theta_l}{2c}$  and the quality of the service that the low-demand customers purchase becomes  $z^l = \frac{(n_h^l + n_l^l)\theta_l}{2c}$ . It follows that the prices due to the participation constraints become  $p^h = \theta_l \frac{(n_h^h + n_l^h)\theta_l}{2c} q^h$  and  $p^l = \frac{(n_h^l + n_l^l)(\theta_l)^2}{2c} q^l$ . Then, the profit under (1) is  $\Pi_{(1)} = \frac{(n_h^h + n_l^h)\theta_l^2}{4c} q^h + \frac{(n_h^l + n_l^l)\theta_l^2}{4c} q^l$ . In case (2), the service provider only discriminates the high-demand customers and offers two bundles for the high-demand customers and one bundle for all the low-demand customers. The quality levels offered to the high-demand customers are  $z_h^h = \frac{n_h^h \theta_h}{2c}$  and  $z_l^h = \frac{n_l^h \theta_l - n_h^h (\theta_h - \theta_l)}{2c}$  if  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h}$ . Otherwise,  $z_l^h = 0$ . The prices are determined by the incentive compatibility and participation constraints of the high-demand customers. Then,  $p_l^h = \theta_l z_l^h q^h$  and  $p_h^h = \theta_h z_h^h q^h - (\theta_h - \theta_l) z_l^h q^h$ . The bundle that is designed for the low-demand customers is the same as the one in case (1). Then, the profit of

the service provider becomes  $\Pi_{(2)} = \frac{(n_h^h \theta_h)^2}{4c} q^h + \frac{(n_l^h \theta_l - n_h^h (\theta_h - \theta_l))^2}{4c} q^h + \frac{((n_h^l + n_l^l) \theta_l)^2}{4c} q^l$  when  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h}$ . Otherwise,  $\Pi_{(2)} = \frac{(n_h^h \theta_h)^2}{4c} q^h + \frac{((n_h^l + n_l^l) \theta_l)^2}{4c} q^l$ . In case (3), the service provider only discriminates the low-demand customers and offers two bundles, one for the low-demand customers and one for all the high-demand customers. The quality levels offered to the low-demand customers are  $z_h^l = \frac{n_h^h \theta_h}{2c}$  and  $z_l^l = \frac{n_l^l \theta_l - n_h^h (\theta_h - \theta_l)}{2c}$  if  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h}$ . Otherwise,  $z_l^l = 0$ . The prices are determined by the incentive compatibility and participation constraints of the low-demand customers. Then,  $p_l^l = \theta_l z_l^l q^l$  and  $p_h^l = \theta_h z_h^l q^l - (\theta_h - \theta_l) z_l^l q^l$ . The bundle that is designed by the high-demand customers is the same as the one in case (1). Then, the profit of the service provider becomes  $\Pi_{(3)} = \frac{((n_h^h + n_l^h) \theta_l)^2}{4c} q^h + \frac{(n_h^l \theta_h)^2}{4c} q^l + \frac{(n_l^l \theta_l - n_h^h (\theta_h - \theta_l))^2}{4c} q^l$  if  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h}$ . Otherwise,  $\Pi_{(3)} = \frac{((n_h^h + n_l^h) \theta_l)^2}{4c} q^h + \frac{(n_h^l \theta_h)^2}{4c} q^l$ . Then, one can prove that  $\Pi_{(1)} > \max\{\Pi_{(2)}, \Pi_{(3)}\}$  for all  $\frac{\theta_h - \theta_l}{\theta_l} < \frac{n_l^h}{n_h^h}$  and that  $\Pi_{(2)} > \max\{\Pi_{(1)}, \Pi_{(3)}\}$  for all  $\frac{\theta_h - \theta_l}{\theta_l} < \frac{n_l^h}{n_h^h}$ .

When  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h}$ , the service provider offers two different bundles to the market:  $s_7$  with  $(z_7, p_7, q^h)$  that targets all the high-demand customers and  $s_8$  with  $(z_8, p_8, q^l)$  that targets all the low-demand customers. Consider a high-valuation – low-demand customer. This customer prefers  $s_8$  only if  $\theta_h z_8 q^l - p_8 \geq \theta_h z_7 q^h - p_7$ . This condition is satisfied when  $z_8 q^l \geq z_7 q^h$  as  $p_7 = \theta_l z_7 q^h$  and  $p_8 = \theta_l z_8 q^l$ . Hence, if  $\frac{q^l}{q^h} \geq \frac{n_h^l + n_l^l}{n_h^h + n_l^h}$  holds, the high-valuation – low-demand customers always prefer  $s_8$  even when the demand is not verifiable. If  $\frac{q^l}{q^h} \geq \frac{n_h^l + n_l^l}{n_h^h + n_l^h}$  is not satisfied, then the service provider must lower the price of  $(z_8, p_8, q^l)$ . Its price becomes  $p_8 = \theta_h z_8 q^l - (\theta_h - \theta_l) z_7 q^h$ . When  $\frac{n_l^h}{n_h^h} < \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^l}$ , the service provider offers two different bundles to the market:  $s_5$  with  $(z_5, p_5, q^h)$  that only targets the high-valuation – high-demand customers and  $s_8$  with  $(z_8, p_8, q^l)$  that targets all low-demand customers. A high-valuation – low-demand customer prefers  $s_8$  only if  $\theta_h z_8 q^l - p_8 \geq \theta_h z_5 q^h - p_5$  where  $p_8 = \theta_l z_8 q^l$  and  $p_5 = \theta_h z_5 q^h$ . As  $\theta_h z_8 q^l - p_8 > 0$  and  $\theta_h z_5 q^h - p_5 = 0$ , the high-valuation – low-demand customers always prefer  $s_8$ . When  $\frac{\theta_h - \theta_l}{\theta_l} > \frac{n_l^l}{n_h^l}$ , the service provider offers two different bundles to the market:  $s_5$  with  $(z_5, p_5, q^h)$  that only targets the high-valuation – high-demand customers and  $s_6$  with  $(z_6, p_6, q^l)$  that only targets the low-valuation – low-demand customers. A high-valuation – low-demand customer prefers  $s_6$  only if  $\theta_h z_6 q^l - p_6 \geq \theta_h z_5 q^h - p_5$  where  $p_5 = \theta_h z_5 q^h$  and  $p_6 = \theta_h z_6 q^l$ . As the high-valuation – low-demand customers are indifferent between  $s_6$  and  $s_5$  for all parameter values, we can adjust the prices to make them strictly prefer their own bundle.

*Proof of Corollary 2.* As Proposition 2 establishes, the service provider prefers to offer two bundles one for each demand level when  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h}$ . Hence, the profit of the service provider, as shown in the proof of Proposition 2, becomes  $\Pi = \frac{((n_h^h + n_l^h) \theta_l)^2}{4c} q^h + \frac{((n_h^l + n_l^l) \theta_l)^2}{4c} q^l$ . When  $\frac{\theta_h - \theta_l}{\theta_l} > \frac{n_l^h}{n_h^h}$ , the service provider prefers to offer one bundle to all the low-demand customers.

However, the server provider prefers serving only the high-valuation segment of the high-demand customers as the optimal quality for the low-valuation segment of the high-demand customers is  $z_l^h = 0$ . Then, as established in Proposition 2, the profit becomes  $\Pi = \frac{(n_h^h \theta_h)^2}{4c} q^h + \frac{((n_h^h + n_l^l) \theta_l)^2}{4c} q^l$ .

*Proof of Theorem 1.* In order to identify the optimal pricing strategy of the service provider, we need to compare the profits under different pricing schemes. When  $\frac{\theta_h - \theta_l}{\theta_l} > \frac{n_l^h}{n_h^h}$ , there are two options for the service provider. When the service provider discriminates, the service provider offers four different bundles to the market. Bundle 1 targets the high-valuation – high-demand customers. Bundle 2 targets the high-valuation – low-demand customers. Bundle 3 targets the low-valuation – high-demand customers. Bundle 4 targets the low-valuation – low-demand customers. Then, the profit of the service provider becomes  $\Pi_1$  established in Corollary 1. When the service provider does not discriminate, the service provider offers two different bundles to the market. Bundle 1 targets all the high-demand customers and bundle 2 targets all the low-demand customers. Then, the profit of the service provider becomes  $\Pi_2$  established in Corollary 2. When we compare these two profits, for all parameter values, the service provider prefers not to discriminate and offer two bundles one for each demand valuation. That is, when  $\frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^h}{n_h^h}$ , the service provider offers two different bundles to the market. Bundle 1 targets the high-demand customers with the quality  $z = \frac{n^h \theta_l}{2c}$ , the price  $p = \theta_l \frac{n^h \theta_l}{2c} q^h$  and the quantity  $q^h$ . Bundle 2 targets the low-demand customers with the quality  $z = \frac{n^l \theta_l}{2c}$ , the price that depends on the relative size of low-demand customers and the quantity  $q^l$ . The price of this bundle becomes  $p = \frac{n^l (\theta_l)^2}{2c} q^l$  if  $\frac{q^l}{q^h} \geq \frac{n^l}{n^h}$  holds; otherwise  $p = \frac{n^l \theta_h \theta_l q^l - (\theta_h - \theta_l) \theta_l n^h q^h}{2c}$ .

When  $\frac{n_l^h}{n_h^h} < \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^l}$ , there are two options for the service provider. When the service provider discriminates, the service provider offers three different bundles to the market. Bundle 1 targets the high-valuation – high-demand customers. Bundle 2 targets the high-valuation – low-demand customers. Bundle 3 targets the low-valuation – low-demand customers. Then, the profit of the service provider becomes  $\Pi = \frac{(\theta_h)^2 ((n_h^h)^2 q^h + (n_h^l)^2 q^l)}{4c} + \frac{(n_h^l (\theta_h - \theta_l) - n_l^l \theta_l)^2 q_l}{4c}$ . When the service provider does not discriminate, the service provider offers two different bundles to the market. Bundle 1 targets the high-valuation – high-quantity customers and bundle 2 targets all the low-demand customers. Then, the profit of the service provider becomes  $\Pi = \frac{(n_h^h \theta_h)^2}{4c} q^h + \frac{(n_h^l + n_l^l) \theta_l)^2}{4c} q^l$ . When we compare these two profits, for all parameter values, the service provider prefers not to discriminate and offer two bundles. That is, when  $\frac{n_l^h}{n_h^h} < \frac{\theta_h - \theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^l}$ , the service provider offers two different bundles to the market. Bundle 1 targets the high-valuation – high-quantity customers with the quality  $z = \frac{n_h^h \theta_h}{2c}$ , the price  $p =$

$\frac{n_h^h(\theta_h)^2}{2c}q^h$  and the quantity  $q^h$ . Bundle 2 targets the low-demand customers with the quality  $z = \frac{(n_h^l+n_l^l)\theta_l}{2c}$ , the price  $p = \frac{(n_h^l+n_l^l)(\theta_l)^2}{2c}q^l$  and the quantity  $q^l$ .

When  $\frac{\theta_h-\theta_l}{\theta_l} > \frac{n_l^l}{n_h^l}$ , there are two options for the service provider. The service provider offers two different bundles to the market. Bundle 1 targets the high-valuation – high-demand customers. Bundle 2 targets the high-valuation – low-demand customers. Then, the profit of the service provider becomes  $\Pi = \frac{(\theta_h)^2((n_h^h)^2q^h+(n_h^l)^2q^l)}{4c}$ . When the service provider does not discriminate, the service provider offers two different bundles to the market. Bundle 1 targets the high-valuation – high-quantity customers and bundle 2 targets all the low-demand customers. Then, the profit of the service provider becomes  $\Pi = \frac{(n_h^h\theta_h)^2}{4c}q^h + \frac{((n_h^l+n_l^l)\theta_l)^2}{4c}q^l$ . When we compare these two profits, for all parameter values, the service provider prefers discriminating. That is, when  $\frac{\theta_h-\theta_l}{\theta_l} > \frac{n_l^l}{n_h^l}$ , the service provider offers two different bundles to the market. Bundle 1 targets the high-valuation – high-demand customers with the quality  $z = \frac{n_h^h\theta_h}{2c}$ , the price  $p = \frac{n_h^h(\theta_h)^2}{2c}q^h$  and the quantity  $q^h$ , and bundle 2 targets the high-valuation – low-demand customers with the quality  $z = \frac{n_h^l\theta_h}{2c}$ , the price  $p = \frac{n_h^l(\theta_h)^2}{2c}q^l$  and the quantity  $q^l$ .

*Proof of Corollary 3* When  $\frac{\theta_h-\theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^h}$ , the service provider prefers to serve all the high-demand customers at the same quality level of  $z = \frac{(n_h^h+n_l^h)\theta_l}{2c}$  and the same price level of  $p = \theta_l \frac{(n_h^h+n_l^h)\theta_l}{2c}q_h$ , and serve all the low-demand customers at the same quality level of  $z = \frac{(n_h^l+n_l^l)\theta_l}{2c}$  and the same price level of  $p = \frac{(n_h^l+n_l^l)(\theta_l)^2}{2c}q^l$ . The profit the service provider, as established in Corollary 2, becomes  $\frac{((n_h^h+n_l^h)\theta_l)^2}{4c}q^h + \frac{((n_h^l+n_l^l)\theta_l)^2}{4c}q^l$ . When  $\frac{n_l^h}{n_h^h} < \frac{\theta_h-\theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^h}$ , the service provider prefers to serve only the high-valuation segment of the high-demand customers and serve all the low-demand customers at the same quality level. The quality of the service that is designed for the high-valuation – high-demand customers is  $z = \frac{n_h^h\theta_h}{2c}$ , and its corresponding price is  $p = \frac{n_h^h(\theta_h)^2}{2c}q^h$ , whereas the quality of the service that is designed for all the low-demand customers is  $z = \frac{(n_h^l+n_l^l)\theta_l}{2c}$ , and its corresponding price is  $p = \frac{(n_h^l+n_l^l)(\theta_l)^2}{2c}q^l$ . The profit the service provider, as established in Corollary 2, becomes  $\frac{(n_h^h\theta_h)^2}{4c}q^h + \frac{((n_h^l+n_l^l)\theta_l)^2}{4c}q^l$ . Finally, when  $\frac{n_l^h}{n_h^h} < \frac{\theta_h-\theta_l}{\theta_l} \leq \frac{n_l^l}{n_h^h}$ , the service provider prefers to serve only the high-valuation segment of the high- and low-demand customers. The quality

of the service that is designed for the high-valuation – high-demand customers is  $z = \frac{n_h^h\theta_h}{2c}$ , and its corresponding price is  $p = \frac{n_h^h(\theta_h)^2}{2c}q^h$ , whereas the quality of the service that is designed for the high-valuation – low-demand customers is  $z = \frac{n_h^l\theta_h}{2c}$ , and its corresponding price is  $p =$

$\frac{n_h^l(\theta_h)^2}{2c} q^l$ . The profit of the service provider, as established in Corollary 1, becomes

$$\frac{*(\theta_h)^2((n_h^h)^2 q^h + (n_h^l)^2 q^l)}{4c}.$$