Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0000-0000 | EISSN 0000-0000 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000

Optimal Pricing Metrics for Digital Goods

Ke-Wei Huang
Leonard N. Stern School of Business, New York University, 44 West 4th Street, New York, New York 10012,
khuang0@stern.nyu.edu, http://pages.stern.nyu.edu/khuang0

The recent advances of information technologies make price discrimination more prevalent and more complicated. Flooded by a large number of variables found by data mining algorithms, pricing managers are perplexed by the task of selecting variables for price discrimination to maximize profits. However, relevant literature remains scarce. This paper attempts to investigate how a monopoly seller should determine the optimal combination of pricing variables. The criterion found is similar to the selection of independent variables for linear regression; it is revenue-maximizing to select the variable that best reduces the residual variance of buyer's willingness-to-pay. When incorporating costs associated with pricing metrics into this model, I propose a linear-running-time algorithm to solve this problem with any general cost structure. As to the implications for public policy, the welfare analysis shows that the monopoly seller will use an excessive number of metrics compared with the socially optimal level. By linking this theoretical model to a probit model, I demonstrate how to use this model to solve the price discrimination problem of an electronic greeting cards website.

Key words: pricing; price discrimination; information goods; digital goods; third-degree price discrimination; two-part tariff; probit model.
History: Preliminary: Please DO NOT CITE without consulting the author. Current Draft: July 2006

1. Introduction

Price discrimination has been ubiquitous in the business world for decades. More recently, advances in information technologies have enabled sellers to collect and store customer information much more cost-effectively. Equipped with analytical tools from burgeoning research in data mining, marketing practitioners can learn more about each customer's purchasing pattern, and have begun to personalize prices and tailor product offerings to each customer. Given a lot of customer information, pricing based on different variables may have significant impacts on the profits of the firms. However, it is puzzling that similar information technology (IT) products have different pricing schemes and even the same products may have different pricing schemes over time. For example, IBM mainframes and most mainframe applications have been priced based on the horsepower (million instructions per second, MIPS) of CPUs for years. Enterprise database software has been priced based on the number of CPUs of the server that installs the database. Oracle, the most dominant database vendor, once had a two-year long experiment with pricing based on the speed of CPU but failed. In 2001, Oracle abandoned this pricing scheme and changed back to per-CPU pricing. In 2003, Sun Microsystems reshaped software industry with a new Java enterprise system priced at $100 per employee with infinite right to use. As another instance, 23.2% of informational database services (Gale's Directory of Databases 2000) charge users based on the number of searches, while 62.3% of these services charge users based on the time spent on each search (Jain and Kannan (2002)).

The selection of pricing variables is becoming more important for enterprise software and Internet advertising pricing. The complexity of enterprise software pricing is growing because of the trend towards licensing software on a usage basis. Unlike utilities such as electricity or gas, the measurement of the usage of software or server products is not standardized. As a consequence, software vendors are experimenting with different proxy variables for pricing. For example, HP, IBM, and Sun are independently developing new usage metrics for on-demand computing (utility computing), and have proposed different metrics. Similarly, in the context of Internet advertisements, Google and Yahoo! have been successful in charging advertisers based on the number of clicks on a link. This apparent precision of price discrimination has helped boost Internet advertising to $12.5 billion in 2005, up from $9.6 billion in 2004, according to PricewaterhouseCoopers. However, fraudulent clicks have forced the Internet advertisers to come up with far more sophisticated methods for tracking customer connections. The CEO of DoubleClick Inc. says that every Internet campaign now allows for 50 different types of metrics (Baker and Hempel (2006)). The design of pricing schemes is also becoming more important because information technologies keep reshaping several industries by bringing in new products and/or services. For example, in 2005 the advance of CPU technology enabled a processor chip to come with more than one processing units (cores). Database vendors responded with different pricing schemes to adapt to this new chip technology. Some vendors (e.g., Microsoft SQL Server) priced one chip as one CPU whereas the others (e.g., IBM DB2 and Oracle before 2005) priced it as multiple CPUs (Sliwa (2004)). Similar pricing experiments were repeated when new IT products were introduced. A theoretical model of pricing-scheme design could provide helpful guidelines for successfully launching a new product.

Stimulated by rich business applications, economists and marketing researchers have studied price discrimination extensively for years. However, the literature on the metrics selection problem, which is the focus of this study, is scarce. In this paper, pricing metrics are defined as verifiable, non-adjustable variables associated with each buyer. In other words, a pricing metric is a noisy measurement of the willingness-to-pay (WTP)[1] of each buyer. For simplicity, this study does not

[1] In this study, WTP is defined as the largest amount of money that an individual or group could pay for a product or service without being worse off.

consider the self-selection behaviors of buyers and can be categorized as a third-degree price discrimination model[2] rather than a second-degree price discrimination model[3]. Standard examples of third-degree price discrimination include senior discounts (age as the pricing metric), student discounts (student status as the pricing metric), and Ladies' Nights (gender as the pricing metric). In the previous examples, the number of clicks is the metric in the Internet advertising case, the number of CPUs is the pricing metric in the database case, and MIPS is the metric in the IBM mainframe case.

These phenomena motivate the following research questions: (1) How should the design of a pricing schedule vary with the number and combination of metrics on which pricing is based? (2) How should a seller choose pricing metrics? (3) What are the welfare effects of price discrimination based on more metrics?

An empirically testable model is proposed to examine these questions. The formulation of this model closely resembles the probit model. The logarithm of willingness-to-pay is assumed to be a linear combination of a noise term and the pricing metrics, which can be any verifiable variable, ranging from the specification of hardware to the financial information of the corporate buyer. All variables are assumed to be mutually independent and normally distributed. The result of this model indicates that the logarithm of the optimal price is a linear function of pricing metrics (a two-part tariff). This model also shows that the expected revenue is invariant to the mean of the selected pricing metrics and it is decreasing exponentially in the residual variance of WTP. As a consequence, it is optimal for the sellers to adopt those pricing metrics with high correlation to WTP or high variance of the values of the metrics. In other words, the "variance reduction" criterion found is the same as that of selecting independent variables for a linear regression. It is surprising that unlike linear regression, this model does not assume a quadratic objective function resulting from minimizing mean-square error. Nevertheless, the optimal rule of selecting variable is the same.

Although using more metrics always improves the firm's revenue, using more metrics also incurs higher costs, including the administrating costs of more pricing metrics, mental accounting costs of consumers to choose among complex pricing plans, or the consumers' aversion to price discrimination. Given any cost function associated with each metric, a linear-running-time algorithm is proposed to solve the metrics selection problem. This algorithm suggests that if the cost of pricing

[2] Third-degree price discrimination means that different purchasers are charged different prices, but each purchaser pays a constant amount for each unit of the good bought (Varian (1992)).

[3] Second-degree price discrimination occurs when prices differ depending on the number of units of the good bought, but not across consumers (Varian (1992)).

metrics is linear in the residual variance, corner solutions appear naturally, in which either uniform pricing or pricing based on many metrics (personalized pricing) is optimal. The welfare analysis shows that the monopoly seller uses an excessive number of metrics compared with the socially optimal level. The reason is that in this model, rent extraction effect always dominates market expansion effect. Hence, whenever a monopoly seller increases the number of metrics, he can extract more rents from the buyers in addition to the rents from the market expansion effect.

A pilot empirical study is conducted using the data from Bluemountain.com. First, a restricted probit model is used to estimate the distribution of WTP. Next, all parameters of the theoretical model can be derived using the estimated distribution of WTP. From the closed-form solutions of optimal revenue and pricing scheme predicted by the theoretical model, the revenue can be estimated when Bluemountain.com prices based on any combination of pricing metrics. This model predicts that if Bluemountain.com prices based on the eldest age of the members of a family, its net income can be improved around 20%. The prescribed optimal pricing scheme ranges from \$5.74 to \$17.82, compared with the actual uniform pricing at \$11.95.

This paper is organized as follows: Section 2 is devoted to the literature review. Section 3 presents the setup and the results of the monopoly model. A comparison with the probit model and a pilot empirical study is provided in Section 4. Concluding remarks are given in Section 5.

2. Literature Review

This study is relevant to the literature of pricing information goods and discrete choice models. In short, compared with the theoretical pricing literature, this study poses a new research question and solves this problem with a new approach. Compared with existing econometric models, this model resembles the probit model and is relevant to discrete choice models, which have been widely applied in the empirical pricing literature. Hence, this model also provides a new, alternative tool for empirical researchers to examine pricing problems.

2.1. Price Discrimination and Pricing Information Goods

Economists have studied price discrimination extensively for years. Stole (2003) provides a recent survey of price discrimination. Price discrimination implies selling the same or similar products to different consumers at different prices. This study can be categorized as a study of noisy first-degree price discrimination or third-degree price discrimination. A brief review is provided in the following paragraphs.

First-degree price discrimination involves charging a different price for every customer based on their WTP. Due to the popularity of target couponing/pricing or personalized pricing in recent years, first degree price discrimination has aroused the attention of researchers in marketing and information systems (see for example, Villas-Boas (1999); Shaffer and Zhang (2002); Villas-Boas (2004); Acquisti and Varian (2005); Choudhary et al. (2005)). Third-degree price discrimination can be achieved when the market is segmented and different segments are charged with different prices. This stream of literature in economics started from the seminal article of Schmalensee (1981) and Varian (1985). This line of research have focused only on the welfare implications of price discrimination but not the metrics selection problem.

Second-degree price discrimination involves charging different prices for products with different quantities (nonlinear pricing) or qualities (versioning). Several studies investigate nonlinear pricing of information goods (see Sundararajan (2004a) and Sundararajan (2004b)). A number of studies investigate the versioning of information goods. (e.g., Nault (1997); Bhargava and Choudhary (2001); Bhargava and Choudhary (2004); Parker and Alstyne (2005)). Recently, Chen and Hitt (2005) apply the second-degree price discrimination model to evaluate customized bundling, a pricing strategy that gives consumers the right to choose up to a quantity M of goods drawn from a larger pool of N different goods for a fixed price. The present study focuses on the pricing metrics that are not adjustable and the self-selection behavior of consumers is assumed away. Extending along this direction may be fruitful but it involves a complicated multi-dimensional screening model and is beyond the scope of this study.

2.2. Empirical Pricing Studies and Discrete Choice Models

Since the seminal work of McFadden (1974), researchers have developed many strands of applications using discrete choice models (random utility models). Two papers provide a complete review of the applications (Allenby and Rossi (1999) and Baltas and Doyle (2001)). Recent advances in this field stem from the seminal work of Berry et al. (1995). The authors develop a framework to obtain the estimates of demand and cost parameters for differentiated products. Rossi et al. (1996) assess the information content of the purchase history as well as demographic characteristics. Their results indicate that even rather short purchase histories can produce a net gain in revenue from target couponing. Along this line of research, several studies have also demonstrated the influences of price discrimination (e.g., Besanko et al. (2003) and Leslie (2004)).

In addition to the discrete choice approach, several marketing studies attempt to derive optimal prices based on the estimated consumer preferences based on other methods. For example, Jedidi and Zhang (2002) propose a conjoint-based approach to estimate consumer-level reservation prices. Jedidi et al. (2003) develop a model for estimating the joint normal distribution of reservation prices for products and bundles. Based on a market experiment, Danaher (2002) derives a revenue-maximizing strategy for subscription services. The results of the present study can be compared with those of this line of research.

In information systems, discrete choice models and the probit regression have been widely applied in examining various topics. For example, by the probit regression, Mukhopadhyay and Kekre (2002) investigate the benefits of the electronic integration for industrial procurement, Gopal et al. (2003) study the choice of contracts and offshore software development projects, and Wu et al. (2005) assess the promotional performance of pure online firms. Using discrete choice models, Chen and Hitt (2002) propose an approach to measure the magnitudes of switching costs and brand loyalty for online service providers, Tam and Hui (2001) and Hui (2004) estimate the demand of PC industry, and several articles investigate the adoption of channels or technologies (e.g., Hitt and Frei (2002); Kim et al. (2002); Forman (2005)).

3. Monopoly Model

This model is a third-degree price discrimination problem. The monopoly seller knows the joint distribution of pricing metrics and WTP but not the exact value of WTP of each buyer. Hence, the monopolist has to estimate the WTP from observable pricing metrics. All pricing metrics are assumed to be verifiable by the seller and not adjustable by the buyers. There are three stages in this game. In the first stage, the seller selects a combination of pricing metrics and posts a pricing schedule in terms of the chosen pricing metrics. Buyers value only one unit of this product. In the second stage, the buyers decide between buying one unit and not buying based on their WTP minus the posted price. In other words, we assume quasi-linear utility functions. In the last stage, buyers pay the product price according to the pricing schedule. They receive zero surplus if they decide not to purchase the product.

Pricing metrics are denoted by an N-dimensional random vector with coordinates denoted by X_i and realizations denoted by x_i. The WTP of any on consumer is denoted by W and is assumed to have the following functional form,

$$ln(W) = \sum_{i=1}^{N} X_i + U, \text{ with } U \sim N(\mu_u, \sigma_u^2) \text{ and } X_i \sim N(\mu_i, \sigma_i^2), \tag{1}$$

where, in the population of buyers, U, $X_1, ..., X_N$, are mutually independent, normally distributed random variables with means and variances denoted as in the preceding equation. If the pricing metrics are not mutually independent, we can apply principle-component analysis to orthogonalize them. Also, in many cases the metrics can be transformed so that the normality assumption is satisfied. The independence among U and the vector of metrics expresses the standard assumption that it represents unobserved "noise". Lastly, U models the unobservable noisy component of WTP.

The prior distribution of $ln(W)$ is denoted as $N(M,V)$, with $M = \sum_{i=1}^{N} \mu_i + \mu_u$ and $V = \sum_{i=1}^{N} \sigma_i^2 + \sigma_u^2$. Without loss of generality, the coefficient of X_i is assumed to be one since the coefficients can be combined into in (μ_i, σ_i^2). For example, if the specification of the model is $ln(W) = \beta X + U$, then it is equivalent to the model with coefficients being one and the associated distributions of each metric being $N(\beta_i \mu_i, \beta_i^2 \sigma_{xi}^2)$. However, note that the interpretation of the variance of X_i is different. When the coefficients are assumed to be one, the variance of X_i includes two components, β_i^2 and σ_{xi}^2. A higher variance may result from higher β_i^2 or σ_{xi}^2.

Also, our formulation is similar to the widely used log-linear hedonic regression. The difference is that in hedonic regression, the independent variables are product characteristics and the dependent variable is the product price. In our model, the independent variables are buyer characteristics and the dependent variable is the WTP. Lastly, bearing in mind in the case of digital goods, the variable cost of production is assumed to be zero.

There are several reasons for assuming a log-normally distributed WTP. First, since W is lognormally distributed, W is nonnegative. Theoretically, it may be acceptable to have negative WTP. However, negative WTP leads to less reliable empirical predictions when we apply the model in actual pricing applications[4]. Second, normal distribution and log-normal distribution are among the most common distributional assumptions in empirical pricing literature (For example, Jedidi and Zhang (2002) and Jedidi et al. (2003)). Furthermore, Section 4 demonstrates that this functional form can be linked to the probit model, which has been applied extensively in the literature. Third, the additivity of normal distributions leads to one convenient property for our problem. Whichever combination of pricing metrics used by the seller, the conditional WTP is always lognormally distributed. Hence, with lognormality assumption, we can derive a closed-form solution of optimal revenue function from price discrimination[5]. Also, this setup adds a new alternative to both theorists or empiricists because many pricing models assume a linear demand curve, which is generated from uniformly distributed WTP. This model assumes log-normally distributed WTP and equivalently, a convex demand curve, which is a more realistic assumption than a linear demand curve.

[4] As demonstrated in section 4, most of the consumers in the population do not pay for the product or service. If negative WTP is allowed, the estimated distribution of WTP will have a very negative mean. In contrast, log-normally distributed WTP will lead to an estimated distribution of WTP with high density around zero. The later one seems to be a more convincing representation of the consumer preferences.

[5] Problems with other distributional assumptions may not have closed-form solutions but can be solved numerically.

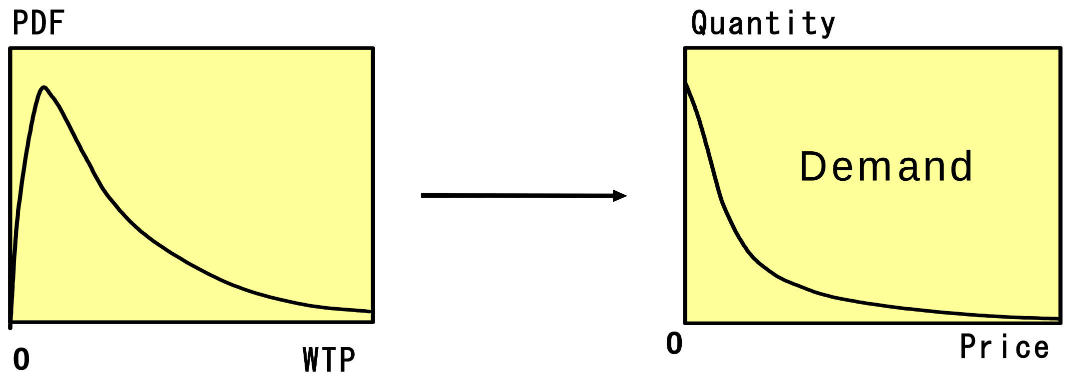

Figure 1: Log-Normal WTP and Demand Curve

3.1. Optimal Price Schedule

Since the monopoly seller always faces a demand curve with log-normally distributed WTP, without loss of generality, we assume the monopoly seller chooses X_1 through X_m as the pricing metrics; then the residual distribution of $\sum_{i=m+1}^{N} X_i + U$ will be normally distributed with mean and variance denoted by μ and σ^2, respectively. In general, the monopoly seller only needs to solve a pricing problem with arbitrary values of μ and σ^2. Formally, the monopolist's pricing problem is to choose a pricing function, $p(x_1 \dots x_m)$, to maximize its expected profit function, which is given by

$$\max_{p(x_1 \dots x_m)} \int_{-\infty}^{\infty} (\pi | x_1 \dots x_m) \mathrm{d}F(x_1 \dots x_m).$$

Note that for each type of buyer, the demand is determined by

$$\Pr[W \geq p(x_1 \dots x_m)] = \Pr[\ln W \geq \ln p(x_1 \dots x_m)].$$

Hence, the objective function can be rewritten as

$$\max_{p(x_1 \dots x_m)} \int_{-\infty}^{\infty} p(x_1 \dots x_m) \cdot \Pr\left[\sum_{i=1}^{m} x_i + \sum_{i=m+1}^{N} X_i + U \geq \ln p(x_1 \dots x_m)\right] \mathrm{d}F(x_1 \dots x_m). \tag{2}$$

Note that in this section, all costs are assumed to be zero and thus the revenue is equal to the profit of the monopolist. Since the seller can perfectly verify the values of customer demographic information, $(x_1 \dots x_m)$, this problem can be further simplified and solved by maximizing the integrand pointwise. Hence, the maximization problem of the integrand can be rewritten as

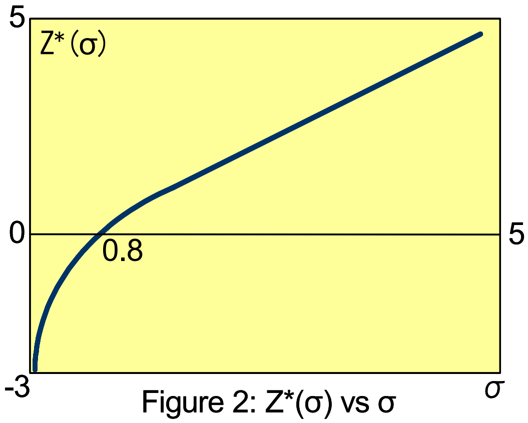

Figure 2: Z*(σ) vs σ

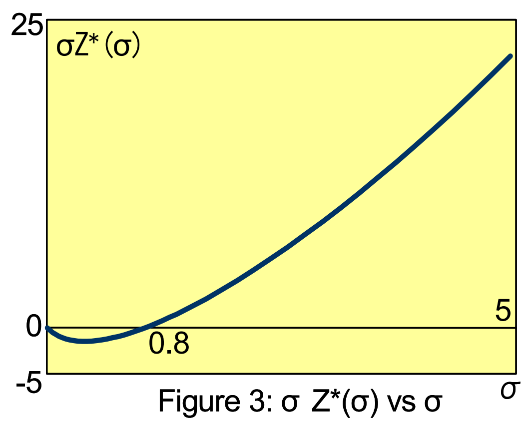

Figure 3: σ Z*(σ) vs σ

$$\max_p(\pi|x_1 \dots x_m) = \max_p \; p \cdot \left[1 - F\left(\ln p - \sum_{i=1}^{m} x_i; \mu, \sigma^2\right)\right], \tag{3}$$

where the cumulative distribution function of $N(m, s^2)$ is denoted by $F(\cdot; m, s^2)$ and the corresponding density function is denoted by $f(\cdot; m, s^2)$. For abbreviation, the probability density function (p.d.f.) and cumulative density function (c.d.f.) of standard normal distribution is denoted by $F_0(\cdot)$ and $f_0(\cdot)$, respectively. All of the conditional notations of $x_1 \dots x_m$ will be dropped for ease of exposition. The following theorem states the optimal pricing schedule of this maximization problem.

THEOREM 1. *The log of the optimal pricing schedule,* $\ln p^*$*, is linear in the realized values of* X_1 $\dots$ X_m*, and* μ*.*

$$\ln p^* = \mu + \sum_{i=1}^{m} x_i + \sigma \times z^*(\sigma).$$

Proof: All of the proofs are relegated to the Appendix. **Q.E.D.**

In this study, $z^*(\sigma)$ is defined as the solution of the following equation,

$$\frac{f_0(z^*)}{1 - F_0(z^*)} = \sigma. \tag{4}$$

We can characterize the optimal pricing and revenue in terms of $z^*(\sigma)$. The left hand side of (4) is the well-known hazard rate function[6] (Bagnoli and Bergstrom (2005)). As a result, $z^*(\sigma)$ is simply a inverse hazard rate function of standard normal distribution in terms of σ. Although the analytical form of $z^*(\sigma)$ is not tractable, Figure 2 numerically shows that it is an increasing and concave function[7].

Theorem 1 indicates that the log of the optimal price is linear in μ and in the metric x_i. In other words, the log of the optimal price has a two-part tariff form, in which the fixed component depends

[6] It is also known as the survival rate function or the reliability function in different literature

[7] $z^*(\sigma) \approx -1.97 + 3.22 \cdot \sigma - 1.18 \cdot \sigma^2 + 0.28 \cdot \sigma^3 - 0.02 \cdot \sigma^4$ by OLS.

on the mean and variance of the residual noise, $(\mu + \sigma z^*(\sigma))$. We can also interpret this expression as the conditional expected WTP $(\mu + \sum_{i=1}^m x_i)$ plus $\sigma z^*(\sigma)$, in which $\sigma z^*(\sigma)$ is an uncertainty discount or premium as depicted in Figure 3.

The relationship between $\ln p^*$ and σ is also interesting. The optimal price is first decreasing and then increasing in σ. This is because there are two countervailing effects: a "right tail effect" and an "information effect". When σ is larger, the seller cares more about the right tail of the distribution; the higher the variance, there exist more consumers with very high willingness-to-pay, and the higher the optimal price. When σ is very small, the information effect dominates. Higher σ means the information needed to predict the WTP is less accurate, and thus the optimal price is lower.

The fact that $\sigma z^*(\sigma)$ is not monotonic in σ has an important managerial implication. When there exists heteroscedasticity of σ, the optimal pricing will be non-monotonic in pricing metrics. Particularly, even though consumers with higher income have higher expected WTP, it is not optimal to charge them higher prices. This is because the variance of WTP may be larger and optimal price should be lower due to the uncertainty discount. The following example does not satisfy the heteroscedasticity assumption of our baseline model but it provides a case in which the optimal pricing is not monotonic in the expected WTP of consumers.

EXAMPLE 1. *A website is designing her pricing scheme for banner advertisements. She believes the WTP of advertisers has the following functional form*

$$\ln W = \frac{X}{2} + U,$$

where X is the number of clicks (in thousands) on that customer's banner advertisement, $U \sim N(0, x)$, and $X \sim N(\frac{1}{2}, \frac{1}{64})$. Theorem 1 prescribes the following optimal pricing scheme

$$\ln p^* = \frac{1}{2}x + x \times z^*(x),$$

Figure 4 depicts the optimal pricing in terms of the pricing metric. It is interesting that the seller should offer deepest discounts to consumers with metric values around 0.25 given the fact that consumers with higher x have higher expected WTP.

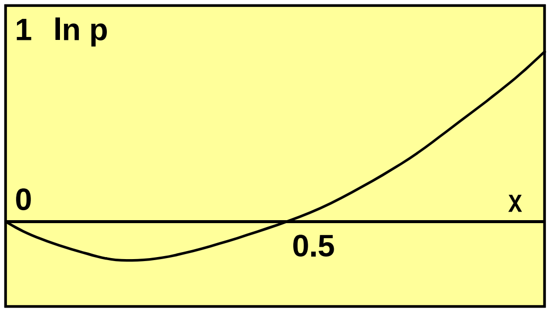

Figure 4: Optimal Price versus x: An Example.

3.2. Optimal Revenue and Optimal Metrics Selection

Given the closed-form solution of prices, the optimal revenue (conditional on metric values) can be derived by substituting the optimal price function into (3). After integrating the conditional revenue function, the expected revenue can be derived. The result is summarized in the following theorem.

THEOREM 2. *The maximum expected profit from price discrimination using metrics $x_1 \ldots x_m$ is*

$$\pi = \exp(M + \frac{V}{2}) \times \frac{1}{\sigma} f_0(z^*(\sigma) - \sigma). \tag{5}$$

COROLLARY 1. *When there is no cost associated with the number of pricing metrics, utilizing more pricing metrics improves profit.*

There are two main observations from Theorem 2. First, the (maximum) expected profit does not depend on the expected value of each metric, but only on M. In other words, the expected value of each pricing metric will not affect the optimal selection of pricing metrics. Second, $\frac{1}{\sigma} f(z^*(\sigma) - \sigma; 0, 1)$ is exponentially decreasing in σ and σ^2 as shown in Figure 5 and Figure 6, respectively. As a result, we can conclude that the higher the residual variance, the lower the profit. When choosing metrics, the monopoly seller only needs to consider the residual σ^2, which should be as small as possible. Similar to selecting independent variables for linear regressions, we should select pricing metrics with higher σ_i to reduce residual variance as much as possible. This high variation, σ_i, may result from two sources as described in the beginning of this section: high coefficient between WTP or high variance of the value of this metric itself. *These two factors together are the main criteria for monopoly sellers to select pricing metrics.*

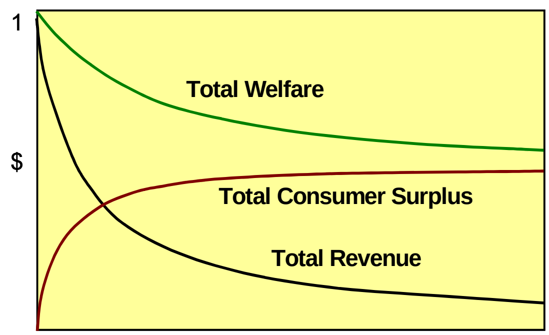

Figure 5: Revenue versus σ: Residual Standard Deviation

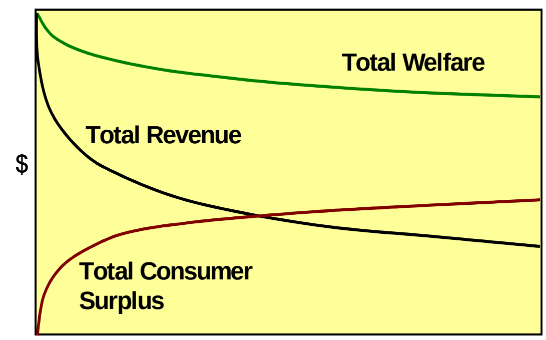

Figure 6: Revenue versus σ²: Residual Variance

3.3. Surplus and Welfare Comparison

The primary research question in the third-degree price discrimination literature is "Does price discrimination leads to higher social welfare?". The following theorem provides a closed-form solution and an affirmative answer to this question.

THEOREM 3. *The total welfare from price discrimination using $x_1 \ldots x_m$ is*

$$\exp(M + \frac{V}{2}) \times [1 - F_0(z^*(\sigma) - \sigma)] .$$

Given this theorem and the fact that total welfare equals total profit plus total consumer surplus, the total welfare, profit, and consumer surplus are numerically plotted in Figure 5 and 6 (when $\exp(M + \frac{V}{2}) = 1$). An immediate observation from Figure 5 and 6 is that total welfare and profit is decreasing, convex in σ whereas consumer surplus is increasing, concave in σ. In other words, price discrimination by more metrics leads to higher total profits and total social welfare but lower consumer surplus.

In this model, when the monopoly seller price discriminates by more metrics, he has more flexibility in pricing, which leads to higher revenue (and profit). The intuition is that when pricing based on N metrics, the monopoly seller can still set the pricing scheme as if pricing based on $N - 1$ metrics, but not vice versa. As a result, pricing based on more metrics always lead to higher revenue. One caveat of this conclusion is that I do not consider the cost of using more metrics for price discrimination, which will be discussed in details in the next section.

The total welfare is also higher when the monopoly seller pricing based on more metrics because he can segment consumers better and sell to more consumers without cannibalizing his existing consumers who have higher WTP. In our model, since the variable cost of production is zero and all consumers have positive WTP, the more consumers are served, the higher the social welfare is. Note that pricing based on more metrics does not guarantee more consumers will be served.

However, Theorem 3 shows that under log-normality assumptions, the monopoly seller will serve more consumers when pricing based on more metrics.

In contrast, the consumer surplus is lower when the monopoly seller prices based on more metrics. The intuition is that as σ becomes smaller, there are two effects on the consumer surplus: (1) less total rent because of better rent extraction by the monopolist and (2) more total rent from market expansion (more consumers with low WTP served by the monopolist). In our model, the first effect dominates the second one and thus the consumer surplus is increasing in σ.

When considering the cost of utilizing more metrics, the seller will price based on an excessive number of metrics compared with the socially optimal level. In the cases of interior solutions, the seller will price based on too many metrics because the absolute value of the slope of the total profit is greater than that of the total welfare. At the maximum[8], it is necessary and sufficient that $\frac{\partial \pi}{\partial \sigma} = \frac{\partial c}{\partial \sigma}$. Since $\frac{\partial \pi}{\partial \sigma} < \frac{\partial W}{\partial \sigma} < 0$, it follows that $\frac{\partial W}{\partial \sigma} - \frac{\partial c}{\partial \sigma} > 0$, which implies the seller uses an excessive number of pricing metrics compared with the socially optimal level (σ is too small). In the cases of corner solutions, the seller either adopts personalized pricing (pricing based on all metrics) or uniform pricing (pricing based on zero metric). Since the gap between total welfare and total profit is increasing in σ, there exist cases in which the monopolist prices based on all metrics when it is socially optimal to charge a uniform price to all consumers.

3.4. A Metrics Selection Algorithm with General Cost Functions

In a real business pricing problem, there exist various kinds of costs associated with implementing different types of price discrimination. When a product is sold at different prices, the seller usually incurs administrative costs such as menu costs, accounting costs, and pricing variables tracking costs. Price discrimination may also incur heavy (psychological) costs on the buyers including the concerns from unfairness, aversion of complicated pricing plans due to bounded rationality, hassle costs of price discrimination by some controversial attributes[9].

Formally, this optimization problem can be formulated as follows,

$$\max_{I_1 \ldots I_N} \pi \left[\sum_{i=1}^{N} I_i \times \sigma_i^2 \right] - \sum_{i=1}^{N} I_i \times c_i,$$
$$\text{subject to } I_i = 0, 1.$$

Here, I_i is a binary decision variable (indicator variable) representing the adoption of this metric. It is one when the seller chooses this metric. For example, if the seller chooses metric 1 and 3, then

[8] We abuse the notation only here and assume for this moment that we have many metrics so that the cost function, c, is a differentaiable function of σ.

[9] An ideal model should model some of these costs as lower WTP, which is beyond the scope of this paper.

only I_1 and I_3 are one and other $I_j s$ are zero. Derived in Theorem 2, $\pi(\cdot)$ is the optimal revenue function in terms of only one argument, the residual variances. Equivalently, it is a function of total variance from the pricing metrics since the total variance of all metrics is a constant V. c_i denotes the fixed cost associated with each pricing metric. Hence, the second term, $\sum_{i=1}^{N} I_i \times c_i$, is simply the total fixed cost associated with a combination of metrics. In general, this is a nonlinear integer programming problem, which is known to be NP-hard. However, by the following lemma, this problem can be solved in linear-running-time.

LEMMA 1. *Denote the optimal set of metrics utilized as S^1 and the set of metrics not utilized as S^0. A necessary condition for the optimal solution is*

$$\frac{c_i}{\sigma_i^2} \leq \frac{c_j}{\sigma_j^2},\ \forall i \in S^1,\ \forall j \in S^0.$$

This lemma states that the cost per unit variance of the metrics utilized must be greater than that of the metrics that are not utilized. This is because the marginal revenue (MR) is increasing in σ_i^2. By comparing the marginal revenue with the average cost of each pricing metric, the firm can improve profits by including pricing metrics whose average cost is lower than the marginal revenue curve. As a result, the optimal selection of metrics will include all metrics with average cost lower than the MR. This criterion is similar to the criteria in other decision problems such as capital budgeting, corporate financing alternatives, or Knapsack problem in computer science. Given this lemma, I propose the following Greedy algorithm.

Algorithm 1

1. (Step 1): Sort all metrics by the average cost per unit variance. Denote the one with smallest unit cost by $i = 1$.
2. (Step 2): Compute N profit numbers, $\left[\pi(\sum_{i=1}^{M} \sigma_i^2) - \sum_{i=1}^{M} c_i\right]$, $\forall M = 1, 2, 3 \ldots N$.
3. (Step 3): Find the maximal one among $M = 1, 2, 3 \ldots N$.

From the computational point of view, step 1 is a typical sorting problem. Step 2 and step 3 are linear in the number of metrics. Sorting is well-studied in computer science and known to achieve average linear-running-time by Quicksort algorithm. As a result, Algorithm 1 has an average linear running time. An interesting special case of our problem is when all metrics have the same fixed cost and variance, which is equivalent to total cost is linear in total variance of metrics. Since the revenue function is convex as shown in Theorem 2, if the cost of metrics is linear in the total variance, then a corner solution, either uniform pricing or using all metrics, is optimal.

4. A Pilot Empirical Application

This section presents a pilot application to illustrate how to solve a real pricing problem using the proposed model. The primary purpose of this application is to show how the proposed approach allows a practitioner to estimate WTP at the individual consumer level with ease. By analyzing the estimated WTP, a practitioner can then easily make complicated price discrimination decisions (or promotion decisions). This pilot study uses the dataset of Bluemountain.com, a famous electronic greeting card website, from the publicly available comScore dataset. Although this dataset does not perfectly satisfy all of our assumptions, it provides enough variables for the demonstration purpose. In the following sections, the link between the baseline theoretical model and probit model is first discussed. Next, the results of the estimated parameters will be reported. Based on the estimated distribution of WTP, we can derive the optimal price scheme by Theorem 1 when pricing based on any combination of pricing metrics.

4.1. Model: A Restricted Probit Model

The proposed model in Section 3 closely resembles the probit model (see Greene (2003)). In probit model, the dependent variable Y is binary and is generally zero or one, and the independent variables x_i^P are estimated in

$$\Pr(Y=1) = F_0(\beta_0 + \sum_{i=1}^{N} \beta_i \cdot x_i^P). \tag{6}$$

Note that $F_0(\cdot)$ is the cumulative distribution function of standard normal distribution. In the following empirical analysis, Y equals one when the consumers pay for Bluemoutain.com. x_i^P are the demographic variables associated with each customer. β_0 and β_i are the coefficients for estimation. The baseline theoretical model can also give the probability that a consumer will pay for the service. Remember that the WTP has the following functional form,

$$\ln W = \sum_{i=1}^{N} x_i + U.$$

A consumer will pay for the service if and only if

$$\Pr(Y=1) = \Pr(p \le W) = \Pr(\ln p \le \ln W).$$

By definition, the distribution function of $\ln W$ is normally distributed with mean $\sum_{i=1}^{N} x_i + \mu$ and variance σ^2. Expressed in terms of distribution function, the previous equation is equivalent to

$$\Pr(Y=1) = 1 - F\left[\ln p; \sum_{i=1}^{N} x_i + \mu, \sigma^2\right].$$

After standardizing this distribution and applying the symmetry of standard normal distribution, this equation can be rewritten as

$$\Pr(Y=1) = 1 - F_0\left(\frac{\ln p - \sum_{i=1}^{N} x_i - \mu}{\sigma}\right), \tag{7}$$

$$= F_0\left[\frac{1}{\sigma}\left(\sum_{i=1}^{N} x_i + \mu - \ln p\right)\right]. \tag{8}$$

Comparing the coefficients of (6) and (8), we have

$$\beta_0 = \frac{1}{\sigma}(\mu - \ln p); \tag{9}$$

$$\beta_i \cdot x_i^P = \frac{x_i}{\sigma}. \tag{10}$$

β_0 and β_i can estimated by a standard probit model and obtain $\ln p$ from the data. However, note that this model has a identification problem to estimate μ and σ at the same time[10]. An additional equation is needed to resolve this problem. In practice, if other data sources are available (such as a market survey of consumers) to estimate the prior mean of WTP, μ, then σ can be derived by (9). Given the values of μ and σ, the optimal pricing and revenue can be derived by Theorem 1 and 2 when pricing based on any combination of metrics.

If other data sources are not available, a restricted probit model is used to estimate μ and σ simultaneously. The idea is to estimate a probit model with a revenue-maximizing constraint. In other words, it is assumed that the observed actual price is also the revenue-maximizing price. The details of this estimation problem is relegated in Appendix B. The results of this pilot study will be estimated using this approach.

4.2. Data

The data is obtained from comScore Internet database at Wharton Research Data Services(WRDS). The comScore panelist-level database captures detailed browsing and buying behavior by 100,000 Internet users across the United States from July 1st, 2002 to December 31st, 2002. This dataset keeps tracks of purchasing behavior of "Online Subscriptions", which is obviously one type of information goods. Data fields include the product name, product price, and the ID of the buyer (machine ID) who bought the product. Since only the machine rather than the user can be identified, the database has only the demographics of the family own that machine. Demographics information

[10] If we estimate only β_0 and β_i by the probit model, then we cannot derive both μ and σ by one equation, eq(9). If we try to estimate β_0, β_i, and σ at the same time by probit model, then the first order condition of maximum likelihood function will be a system of linearly dependent equations.

Descriptive Statistics	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Highest Education	11864	0	99	31.0539	43.9521	0.8978	-1.1910
Census Region	11864	1	4	2.5739	1.0112	-0.1666	-1.0631
Household Size	11864	1	6	2.9782	1.3354	0.6170	-0.4089
Eldest Age	11864	1	11	7.0619	2.6749	-0.3536	-0.6183
Household Income	11864	1	7	4.1763	1.6631	-0.0873	-0.7592
Child Present	11864	0	1	0.4432	0.4968	0.2288	-1.9480
Racial Background	11864	1	5	1.3232	0.9649	3.1307	8.6553
Connection Speed	11864	0	1	0.4335	0.4956	0.2684	-1.9283
Country of Origin	11864	0	1	0.0708	0.2565	3.3471	9.2044

Table 1: Descriptive Statistics of Demographic Data (All Samples)

Descriptive Statistics	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Highest Education	184	0	99	32.5815	44.6336	0.8272	-1.3269
Census Region	184	1	4	2.7228	1.0319	-0.3572	-1.0015
Household Size	184	1	6	2.7772	1.2455	0.8945	0.1483
Eldest Age	184	1	11	7.9728	2.4594	-0.5789	-0.3128
Household Income	184	1	7	4.0109	1.6759	0.1094	-0.6750
Child Present	184	0	1	0.3967	0.4906	0.4256	-1.8390
Racial Background	184	1	5	1.2663	0.9112	3.5641	11.5894
Connection Speed	184	0	1	0.4511	0.4990	0.1982	-1.9824
Country of Origin	184	0	1	0.0924	0.2904	2.8384	6.1229

Table 2: Descriptive Statistics of Demographic Data (Only Buy)

include (1) Highest education (2) Family Size (3) Region (4) Eldest Age (5) Income (6) Child present (7) Racial Background (8) Connection Speed (9) Country of origin.

The present model is tested using the data from Bluemountain.com. Bluemountain is one of the largest electronic greeting cards retailers. The annual membership fee during the sample period is \$11.95. Currently, the annual membership fee is \$13.99. This site is chosen due to the following reasons: (1) It has large enough sample size; (2) It has only one product with only one price in the sample period; (3) It has significant results (only) on non-categorical variables[11].

There are 11864 users who surfed this site with 184 users paid the annual membership fee during the sample period. The descriptive statistics are reported in Table 1 and Table 2. One variable, “highest education”, is dropped because it contains too many missing values. As a consequence, eight independent variables will be included in the preliminary probit model.

[11] There are only three sites satisfying condition (1) and (2). Carfax.com is not significant at any demographic variables. Classmate.com is significant at categorical demographic variables.

	Coef.	Std. Err.	z	P>|z|		[95% Conf.Int.]	
Region1	-0.1287	0.0932	-1.3800	0.1670		-0.3113	0.0540
Region2	-0.1764	0.0887	-1.9900	0.0470	**	-0.3504	-0.0025
Region3	-0.0915	0.0780	-1.1700	0.2410		-0.2443	0.0614
Size	-0.0312	0.0319	-0.9800	0.3290		-0.0937	0.0314
Eldest age	0.0528	0.0120	4.4100	0.0000	***	0.0293	0.0763
Income	-0.0242	0.0178	-1.3600	0.1730		-0.0591	0.0106
Child present	0.0120	0.0814	0.1500	0.8830		-0.1475	0.1715
Racial2	-0.1136	0.1716	-0.6600	0.5080		-0.4500	0.2228
Racial3	-0.1543	0.1960	-0.7900	0.4310		-0.5385	0.2299
Racial5	-0.1016	0.1504	-0.6800	0.5000		-0.3964	0.1932
Connection Speed	0.0601	0.0597	1.0100	0.3140		-0.0568	0.1771
Country of origin	0.1960	0.1178	1.6600	0.0960	*	-0.0349	0.4269
Constant term	-2.2984	0.1567	-14.6700	0.0000	***	-2.6054	-1.9913

*: <10%, **: <5%, ***: <1%.

Table 3: Results of the Probit Regression.

There are two caveats of using this dataset. First, all independent variables are discretized. In this study, fortunately, only "Eldest Age" is statistically significant. Although this variable is discretized, this variable can be assumed to be normally distributed in the following analysis. Second, the effective duration of the membership is one year but the sample period lasts only 6 months. As a result, there are some users who paid before the sample period and visited this website during the sample period. Treating these customers as the ones who didn't pay will underestimate the significance of probit model. As a result, the following method is adopted to alleviate this problem.

This dataset indicates that the usage (total duration) differs dramatically (averaging 16.94 minutes versus 130.45 minutes) for customers who paid and did not pay during the sample period. As a result, total usage can be a proxy variable of Y. Three cases are investigated in the following analysis. Case 1: $Y = 1$ only when the consumer paid for the service during the sample period. Case 2: $Y = 1$ if the consumer paid or the total usage is greater than 224.8267 minutes, which is the value of four standard deviations greater than the average usage of customers who didn't pay. Case 3: $Y = 1$ if the consumer paid or the total usage is greater than 172.855 minutes, which is the value of three standard deviations greater than the average usage of customers who didn't pay. The number of samples with $Y = 1$ is 184, 284, and 336, respectively.

4.3. Results

After the preliminary standard probit regression, only the variable "Eldest Age" is significant at 5% level (see Table (3)). Next, by the restricted maximum likelihood approach in Appendix B, the following specification of consumer preference is estimated.

$$\ln W_i = b_0 + b_1 \cdot X_i + U, \tag{11}$$

where X_i is the value of the "eldest age" of a sample i and U is a normally distributed with mean zero and standard deviation σ_U^2. The results are reported in Table 4. Note that by Theorem 2, three variables are needed to derive the optimal revenue when pricing based on any combination of metrics[12]. Since only one metric is included, it is sufficient to compare the revenue from uniform pricing with the revenue from pricing based on "eldest age". Given the values of prior mean and variance of lnW, by Theorem 2, the optimal revenue of uniform pricing can be calculated. Together with the estimated σ_U^2, the optimal revenue from pricing based on "eldest age" can be derived explicitly. The results are reported in Table 4. It indicates that the revenue improvement ranges from 0.689% to 1.782%. Although the absolute value of improvement is not large at the first sight, however, note that this improvement is generated almost at no cost. Also, the net profit margin[13] of American Greetings Corp.(Bluemountain is a subsidiary of American Greetings Corp.) is around 5% in 2005. An 1% improvement in revenue is equivalent to 20% improvement in net income.

The optimal pricing scheme is reported in Table 5. In practice, retailers can implement this pricing scheme by setting the price at the highest level (e.g., \$17.816) and offering coupons (via emails) to other groups of consumers. A more conservative approach is to email coupons only to those groups of consumers with optimal prices lower than the original price while keeping the posted price at the original level (\$11.95). Lowering prices to those customers should improve the seller's revenue without the risk of offending customers by raising prices.

5. Conclusion

This study presents the results of a general monopoly pricing metrics selection model. With these results, pricing managers can design different (promotional) pricing plans offered to different customers based on their demographic information. Moreover, pricing managers can apply the proposed model to design pricing plans for pay per usage services (e.g., on-demand computing) easily. The results of this study indicate that when the costs of pricing variables are the same, marketing managers should use "the most informative" metrics, including metrics with high correlation with willingness-to-pay (WTP) or metrics with very disperse values. To determine the "informativeness"

[12] These variables are residual variance, prior mean and variance of $\ln W$. Residual variance is estimated from the probit model. The prior mean is derived by $b_0 + b_1 \cdot \mathrm{E}(X_i)$ and prior variance is derived by $b_1^2 \cdot \mathrm{var}(X_i) + \sigma_U^2$.

[13] (Net Income Applicable To Common Shares)/(Total Revenue) for American Greetings Corp, the parent company of Bluemountain.com, is 6.607%, 5.21%, and 5.01% in 2003, 2004, and 2005, respectively.

	Original Sample	4 Sigma Sample	3 Sigma Sample
b0	-3.972	-3.637	-3.467
b1	0.022	0.031	0.033
Std.Dev. Of Noise	2.487	2.317	2.246
Prior expected WTP	-2.902	-2.174	-1.883
Prior variance of WTP	6.280	5.547	5.253
Prior standard deviation	2.506	2.355	2.292
Prior Z*	2.149	1.980	1.908
Posterior Z*	2.128	1.937	1.856
Normalized Revenue from Uniform Pricing	0.189	0.287	0.340
Normalized Revenue from Price Discrimination	0.191	0.291	0.346
Revenue Improvement	0.689%	1.439%	1.782%

Table 4: Results Derived Based on the Theoretical Model.

Eldest Age	Original Sample	4 Sigma Sample	3 Sigma Sample
19	$ 5.737	$ 4.195	$ 3.792
22.5	$ 6.206	$ 4.671	$ 4.260
27	$ 6.866	$ 5.362	$ 4.946
32	$ 7.681	$ 6.250	$ 5.840
37	$ 8.592	$ 7.286	$ 6.895
42	$ 9.613	$ 8.494	$ 8.140
47	$ 10.754	$ 9.901	$ 9.611
52	$ 12.030	$ 11.542	$ 11.347
57	$ 13.459	$ 13.455	$ 13.397
62	$ 15.057	$ 15.685	$ 15.817
69.5	$ 17.816	$ 19.741	$ 20.291

Table 5: Estimated Optimal Pricing Scheme.

of pricing metrics, this study also provides a proper transformation to estimate WTP from the probit regression. In other words, similar to selecting variables for a linear regression, this study shows that the most profitable variable for price discrimination is also the variable that can best reduce the residual variance in a probit regression. When pricing based on different metrics incurs different fixed costs, marketing managers should use the metrics-selecting algorithm developed in Section 3 to find the optimal combination of pricing metrics.

This study also demonstrates the proposed procedures in a pilot application using the data from

Bluemountain.com. Specifically, the probit regression indicates that "Eldest Age" is the most profitable demographic variable for price discrimination. Furthermore, it estimates that the improvement of net income is around 20%. This result shows that the proposed theoretical model indeed can provide a distinct perspective on the consumer demand and can help managers to make better pricing decisions.

There are three main limitations of this study. First, this study does not allow buyers to adjust their characteristics (second-degree price discrimination). After incorporating this feature, it becomes a multi-dimensional second-degree price discrimination problem, which is fairly difficult and is beyond the scope of this paper. Second, this model does not address the issues of discrete metrics and heteroscedasticity of WTP of different groups of consumers. However, this model is still applicable to most real world pricing problems since a model with discrete metrics and/or heteroscedasticity of WTP is numerically (but not analytically) tractable. Lastly, this study does not consider several common pricing constraints, such as legal restrictions or bounded rational consumers. For example, laws or social norms may constrain the differences in prices charged to different consumers. Incorporating these constraints may make the baseline model intractable analytically and lose transparency. However, a real world pricing application can be solved numerically following a procedure similar to the one proposed in this study.

A logical direction for future research would be to extend the model of this paper to a competitive setting. It is not clear whether in a duopoly, sellers will benefit from pricing based on the same set of pricing metrics or not. Extending this model from a single product monopoly seller to a multi-product monopoly seller may be another direction with important business implications. I hope this research paves the way for future work in this domain.

Acknowledgments

This paper is a chapter of my dissertation at New York University. I would like to thank my advisor, Roy Radner, for encouragement and many helpful suggestions. I would also like to thank the following persons for their comments on the earlier drafts of this paper: Anindya Ghose, William Greene, Joel Steckel, Arun Sundararajan, Foster Provost, Ying-Ju Chen, seminar participants at Workshop on Information Systems and Economics 2005, Stern School of Business, and Wharton School of Business. This project has been supported by Taiwan Merit Scholarship(TMS-094-1-A-043). The usual disclaimer applies.

Appendix A: Proofs of Theorems

A.1. Proof of Theorem 1

The first order condition for a maximum is given by

$$\frac{f(\ln p^* - \sum_{i=1}^m x_i; \mu, \sigma^2)}{1 - F(\ln p^* - \sum_{i=1}^m x_i; \mu, \sigma^2)} = 1, \tag{12}$$

which is equivalent to the condition that the hazard rate equals one[14]. After normalizing $\ln p^* - \sum_{i=1}^m x_i$ by μ and σ^2, the first order condition can be rewritten as

$$\frac{f(\ln p^* - \sum_{i=1}^m x_i; \mu, \sigma^2)}{1 - F(\ln p^* - \sum_{i=1}^m x_i; \mu, \sigma^2)} = 1 \Leftrightarrow \frac{f_0(z^*)}{1 - F_0(z^*)} = \sigma, \tag{13}$$

where

$$z^*(\sigma) = \frac{\ln p^* - \sum_{i=1}^m x_i - \mu}{\sigma}. \tag{14}$$

This establishes this theorem.

A.2. Proof of Theorem 2

By the F.O.C. and the optimal price schedule, we can further simplify the expected revenue function in several steps. First, by (12), $1 - F(\cdot)$ in the objective function can be substituted by $f(\cdot)$ and it follows that

$$\mathbf{E}[\pi] = \mathbf{E}\left[p \cdot f\left(\ln p - \sum_{i=1}^m X_i; \mu, \sigma^2\right)\right]. \tag{15}$$

It is equal to

$$\mathbf{E}\left[\exp\left[\mu + \sum_{i=1}^m X_i + \sigma \cdot z^*(\sigma)\right] \cdot f\left(\ln p - \sum_{i=1}^m X_i; \mu, \sigma^2\right)\right], \tag{16}$$

which leads to

$$\mathbf{E}\left[\exp\left[\mu + \sum_{i=1}^m X_i + \sigma \cdot z^*(\sigma)\right]\right] \cdot \frac{f\left(z^*(\sigma); 0, 1\right)}{\sigma}. \tag{17}$$

Note that the only random term is $\mathbf{E}[\exp \sum_{i=1}^m X_i]$. By the independence assumption among X_is,

$$\mathbf{E}[\exp \sum_{i=1}^m X_i] = \Pi_i \; \mathbf{E}[\exp X_i]. \tag{18}$$

Thus, it is sufficient to derive $\mathbf{E}[\exp X_i]$. Taking the expectation on $\exp X_i$, we have

$$\mathbf{E}[\exp X_i] = \int_{-\infty}^{\infty} \frac{1}{\sigma_i\sqrt{2\pi}} \exp(x_i - \frac{1}{2\sigma_i^2}\left(x_i^2 - 2x_i\mu_i + \mu_i^2\right))\mathrm{d}x_i, \tag{19}$$

which simplifies to

$$\exp(\mu_i + \frac{1}{2}\sigma_i^2)\int_{-\infty}^{\infty} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_i^2}\left(x_i - \mu_i - \sigma_i^2\right)\right]\mathrm{d}x_i = \exp(\mu_i + \frac{1}{2}\sigma_i^2). \tag{20}$$

As a consequence, the optimal expected revenue is given by

$$\begin{aligned}
\mathbf{E}[\pi] &= \exp\left[\mu + \sigma \cdot z^*(\sigma)\right] \times \Pi_i \; \exp(\mu_i + \frac{1}{2}\sigma_i^2) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{z^*(\sigma)^2}{2}\right], && (21)\\
&= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\left(\mu + \sum_{i=1}^m \mu_i\right) + \frac{1}{2}\sum_{i=1}^m \sigma_i^2 - \frac{z^*(\sigma)^2}{2} + \sigma \times z^*(\sigma)\right], && (22)\\
&= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[M + \frac{1}{2}V - \frac{1}{2}\sigma^2 - \frac{z^*(\sigma)^2}{2} + \sigma \times z^*(\sigma)\right], && (23)
\end{aligned}$$

which is equal to the expression in the main text after rearranging terms.

[14] See Bagnoli and Bergstrom (2005). It is a standard result that normal distribution has a strictly increasing hazard rate function, which guarantees the uniqueness of the solution. The sufficiency of this type of problem is also documented in the same study.

A.3. Proof of Corollary 1

By Theorem 2, the revenue can be decomposed as

$$\frac{\exp\left[M+\frac{1}{2}V\right]}{\sqrt{2\pi}} \times \frac{\exp\left[-\frac{1}{2}(\sigma - z^*(\sigma))^2\right]}{\sigma}. \tag{24}$$

The first term is a constant and the second term is decreasing, convex in σ as shown in Figure 4. As a consequence, introducing more variables strictly decreases σ and leads to higher profits. **Q.E.D.**

A.4. Proof of Lemma 1

For a solution to be optimal, it is necessary that using one more metric, x_j, will not increase the total profit.

$$\pi(\sum_{i\in S^1} I_i \times \sigma_i^2) - \sum_{i\in S^1} I_i \times c_i \geq \pi(\sigma_j^2 + \sum_{i\in S^1} I_i \times \sigma_i^2) - c_j - \sum_{i\in S^1} I_i \times c_i,\ \forall j \in S^0, \tag{25}$$

which is equivalent to

$$c_j \geq \pi(\sigma_j^2 + \sum_{i\in S^1} I_i \times \sigma_i^2) - \pi(\sum_{i\in S^1} I_i \times \sigma_i^2) > \sigma_j^2 \cdot \frac{d\pi(\sum_{i\in S^1} I_i \times \sigma_i^2)}{d\sigma_i^2},\ \forall j \in S^0. \tag{26}$$

The last inequality comes from the convexity of the revenue function (in the residual variance). Similarly, it is necessary that using one less metric from S^1 cannot increase the total profit.

$$\pi(\sum_{i\in S^1} I_i \times \sigma_i^2) - \sum_{i\in S^1} I_i \times c_i \geq \pi(-\sigma_k^2 + \sum_{i\in S^1} I_i \times \sigma_i^2) + c_k - \sum_{i\in S^1} I_i \times c_i, \forall k \in S^1, \tag{27}$$

which is equivalent to

$$c_k \leq \pi(\sum_{i\in S^1} I_i \times \sigma_i^2) - \pi(-\sigma_k^2 + \sum_{i\in S^1} I_i \times \sigma_i^2) < \sigma_k^2 \cdot \frac{d\pi(\sum_{i\in S^1} I_i \times \sigma_i^2)}{d\sigma_i^2}, \forall k \in S^1. \tag{28}$$

As a result, by (26) and (28), it follows that

$$\frac{c_k}{\sigma_k^2} < \frac{d\pi(\sum_{i\in S^1} I_i \times \sigma_i^2)}{d\sigma_i^2} < \frac{c_j}{\sigma_j^2}, \forall k \in S^1, \forall j \in S^0, \tag{29}$$

and the result is established.

A.5. Proof of Theorem 3

Since the variable cost of production is zero, the total welfare is the sum (integration) of all WTP of all consumers who have WTP higher than the prices offered by the monopolist. Formally, when the seller prices based on metrics $X_1 \dots X_m$, the conditional total welfare for a consumer with metric values $(x_1 \dots x_m)$ is

$$\begin{aligned} &\mathbf{E}_{X_{m+1}\dots X_N}\left[W(x_1 \dots x_m, X_{m+1} \dots X_N) - 0 | W(x_1 \dots x_m) > p(x_1 \dots x_m)\right] \\ &= \int_{\ln p - \sum_{i=1}^m x_i}^{\infty} \exp(y + \sum_{i=1}^m x_i) \cdot \mathrm{d}F(y;\mu,\sigma), \end{aligned} \tag{30}$$

which leads to

$$\int_{\ln p - \sum_{i=1}^m x_i}^{\infty} \left[\exp(y + \sum_{i=1}^m x_i) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]\right] \mathrm{d}y. \tag{31}$$

This term can be rewritten as

$$\exp(\sum_{i=1}^m x_i) \int_{\ln p - \sum_{i=1}^m x_i}^{\infty} \left\{\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y-\mu-\sigma^2)^2}{2\sigma^2} + \frac{2\mu+\sigma^2}{2}\right]\right\} \mathrm{d}y. \tag{32}$$

The integration of the first term in the bracket is equivalent to the cumulative density function of a normal distribution. Hence, the conditional welfare is equal to

$$\exp(\frac{2\mu+\sigma^2}{2}+\sum_{i=1}^{m}x_i)\times\left[1-F(\ln p-\sum_{i=1}^{m}x_i;\mu+\sigma^2;\sigma^2)\right]. \tag{33}$$

After standardizing the normal distribution, we have

$$\exp(\mu+\frac{\sigma^2}{2}+\sum_{i=1}^{m}x_i)\times[1-F_0(z^*(\sigma)-\sigma)]. \tag{34}$$

Given this expression of the conditional welfare, we can derive the prior, expected welfare similar to the procedure in the proof of Theorem 2. Note that only the first term depends on x_i, the total welfare can be written as

$$\begin{aligned} W &\equiv \mathbf{E}_{X_1\dots X_m}\left[\exp(\mu+\frac{\sigma^2}{2}+\sum_{i=1}^{m}X_i)\times[1-F_0(z^*(\sigma)-\sigma)]\right] \\ &= \exp(\mu+\frac{\sigma^2}{2})\times[1-F_0(z^*(\sigma)-\sigma)]\times\mathbf{E}_{X_1\dots X_m}\left[\exp(\sum_{i=1}^{m}X_i)\right], \end{aligned} \tag{35}$$

which leads to

$$W=\exp(\mu+\frac{\sigma^2}{2})\times[1-F_0(z^*(\sigma)-\sigma)]\times\Pi_i\ \exp(\mu_i+\frac{1}{2}\sigma_i^2). \tag{36}$$

By $\mathbf{E}[\exp X_i]=\exp(\mu_i+\frac{1}{2}\sigma_i^2)$ in the proof of Theorem 2, it follows that

$$W=\exp(M+\frac{V}{2})\times[1-F_0(z^*(\sigma)-\sigma)], \tag{37}$$

and the result is established.

Appendix B: Restricted Probit Model

The model is

$$(\textbf{Restricted Probit Model}):\ \Pr(Y_i=1)=F(\ln p;\beta_0+\beta_1\cdot x^P,\sigma^2). \tag{38}$$

The likelihood function is

$$L=\prod_{j=1}^{m}\left[1-F(\ln p;\beta_0+\beta_1\cdot x^P,\sigma^2)\right]^{N_j^1}\cdot\left[F(\ln p;\beta_0+\beta_1\cdot x^P,\sigma^2)\right]^{N_j^0}. \tag{39}$$

In comScore dataset, all of the independent variables are discretized and $m=11$ represents the number of groups of "eldest age". N_j^1 and N_j^0 are the number of positive and negative samples of each group. Because of the identification problem, I cannot estimate β_0, β_0, and σ^2 at once. Hence, the following revenue-maximizing constraint is proposed.

$$\sum_{i=1}^{N}\left[1-F(\ln p;\beta_0+\beta_1\cdot x_i^P,\sigma^2)-f(\ln p;\beta_0+\beta_1\cdot x_i^P,\sigma^2)\right]=0, \tag{40}$$

where $N=11864$ is the number of samples. As a consequence, β_0, β_0, and σ^2 are estimated by maximizing $\ln(L)$ subject to this constraint. The results are reported in Table 4.

References

Acquisti, Alessandro, Hal R. Varian. 2005. Conditioning prices on purchase history. *Marketing Science* **24** 367–381.

Allenby, Greg M., Peter E. Rossi. 1999. Marketing models of consumer heterogeneity. *Journal of Econometrics* **89** 57–78.

Bagnoli, Mark, Ted Bergstrom. 2005. Log-concave probability and its applications. *Economic Theory* **26** 445–469.

Baker, Stephen, Jessi Hempel. 2006. Wiser about the web. *Business Week* .

Baltas, George, Peter Doyle. 2001. Random utility models in marketing research: a survey. *Journal of Business Research* **51** 115–125.

Berry, Steven, James Levinsohn, Ariel Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* **63** 841–890.

Besanko, David, Jean-Pierre Dube, Sachin Gupta. 2003. Competitive price discrimination strategies in a vertical channel using aggregate retail data. *Management Science* **49** 1121–1138.

Bhargava, Hemant K, Vidyanand Choudhary. 2001. Information goods and vertical differentiation. *Journal of Management Information Systems* **18**(2) 89–107.

Bhargava, Hemant K, Vidyanand Choudhary. 2004. Economics of an information intermediary with aggregation benefits. *Information Systems Research* **15**(1) 22–37.

Chen, Pei-Yu, Lorin Hitt. 2002. Measuring switching costs and the determinants of customer retention in internet-enabled businesses: A study of the online brokerage industry. *Information Systems Research* **13**(3) 255–274.

Chen, Pei-Yu, Lorin Hitt. 2005. Bundling with customer self-selection: A simple approach to bundling low-marginal-cost goods. *Management Science* **51**(10) 1481–1493.

Choudhary, Vidyanand, Anindya Ghose, Tridas Mukhopadhyay, Uday Rajan. 2005. Personalized pricing and quality differentiation. *Management Science* **51** 1120–1130.

Danaher, Peter. 2002. Optimal pricing of new subscription services: Analysis of a market experiment. *Marketing Science* **21** 119–138.

Forman, Chris. 2005. The corporate digital divide: Determinants of internet adoption. *Management Science* **51**(4) 641–654.

Gopal, Anandasivam, Konduru Sivaramakrishnan, MS Krishnan, Tridas Mukhopadhyay. 2003. Contracts in offshore software development: An empirical analysis. *Management Science* **49**(12) 1671–1683.

Greene, William. 2003. *Econometric Analysis*. 5th ed. Prentice Hall.

Hitt, Lorin M, Frances X Frei. 2002. Do better customers utilize electronic distribution channels? the case of pc banking. *Management Science* **48**(6) 732–749.

Hui, Kai-Lung. 2004. Product variety under brand influence: An empirical investigation of personal computer demand. *Management Science* **50**(5) 686–700.

Jain, Sanjay, P. K. Kannan. 2002. Pricing of information products on online servers: Issues, models, and analysis. *Management Science* **48**(9) 1123–1142.

Jedidi, Kamel, Sharan Jagpal, Puneet Manchanda. 2003. Measuring heterogeneous reservation prices for product bundles. *Marketing Science* **22** 107–130.

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

Kim, Namwoon, Jin K Han, Rajendra K Srivastava. 2002. A dynamic it adoption model for the soho market: Pc generational decisions with technological expectations. *Management Science* **48**(2) 222–241.

Leslie, Phillip. 2004. Price discrimination in broadway theater. *The Rand Journal of Economics* **35** 520–541.

McFadden, Daniel. 1974. The measurement of urban travel demand. *Journal of Public Economics* **3** 303–328.

Mukhopadhyay, Tridas, Sunder Kekre. 2002. Strategic and operational benefits of electronic integration in b2b procurement processes. *Management Science* **48**(10) 1301–1314.

Nault, Barrie. 1997. Quality differentiation and adoption costs: The case for interorganizational information system pricing. *Annals of Operations Research* **71**(0) 115–142.

Parker, Geoffrey G, Marshall W Van Alstyne. 2005. Two-sided network effects: A theory of information product design. *Management Science* **51**(10) 1494–1504.

Rossi, Peter E., Robert E. McCulloch, Greg M. Allenby. 1996. The value of purchase history data in target marketing. *Marketing Science* **15** 321–340.

Schmalensee, Richard. 1981. Output and welfare implications of monopolistic third-degree price discrimination. *American Economic Review* **71** 242–247.

Shaffer, Greg, Z. John Zhang. 2002. Competitive one-to-one promotions. *Management Science* **48** 1143–1160.

Sliwa, Carol. 2004. Microsoft takes different tack on multicore pricing. *Computerworld* **38**(43).

Stole, Lars. 2003. Price discrimination and imperfect competition. *Handbook of Industrial Organization: Vol.III* .

Sundararajan, Arun. 2004a. Managing digital piracy: Pricing and protection. *Information Systems Research* **15**(3) 287–308.

Sundararajan, Arun. 2004b. Nonlinear pricing of information goods. *Management Science* **50**(12) 1660–1673.

Tam, Kar Yan, Kai-Lung Hui. 2001. A choice model for the selection of computer vendors and its empirical estimation. *Journal of Management Information Systems* **17**(4) 97–124.

Varian, Hal. 1992. *Microeconomic Analysis*. 3rd ed. WW Norton, New York NY.

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

Villas-Boas, Miguel. 1999. Dynamic competition with customer recognition. *The Rand Journal of Economics* **30** 604–631.

Villas-Boas, Miguel. 2004. Price cycles in markets with customer recognition. *The Rand Journal of Economics* **35** 486–501.

Wu, Jianan, Victor J Cook Jr, Edward C Strong. 2005. A two-stage model of the promotional performance of pure online firms. *Information Systems Research* **16**(4) 334–20.