A Conjoint Model of Quantity Discounts

Raghuram Iyengar
Assistant Professor of Marketing
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104
Email: riyengar@wharton.upenn.edu

Kamel Jedidi*
John A. Howard Professor of Marketing
Columbia Business School
Columbia University, New York
New York, NY – 10027
Email: kj7@columbia.edu

July, 2011

*Raghuram Iyengar is an Assistant Professor at the Wharton School, University of Pennsylvania. Kamel Jedidi is The John A. Howard Professor of Marketing at the Graduate School of Business, Columbia University. The authors thank Rajan Sambandam and Christi Clark from TRC Market Research for their support in data collection. The authors also thank Andre Bonfrer, Eric Bradlow, Stephanie Finnel, Arun Gopalakrishnan, Rajeev Kohli, Abba Krieger, Leonard Lodish, Eric Schwartz and Venkatesh Shankar for their insightful comments.

A Conjoint Model of Quantity Discounts

Abstract

Quantity discount pricing is a common practice used by business-to-business and business-to-consumer companies. A key characteristic of quantity discount pricing is that the marginal price declines with higher purchase quantities. In this paper, we propose a choice-based conjoint model for estimating consumer-level willingness-to-pay (WTP) for varying quantities of a product and for designing optimal quantity discount pricing schemes. Our model can handle large quantity values and produces WTP estimates that are positive and increasing in quantity at a diminishing rate. In particular, we propose a tractable utility function which depends on both product attributes and product quantity and which captures diminishing marginal utility. We show how such a function embeds standard utility functions in the quantity discount literature as special cases and how to use it to estimate the WTP function and consumer value potential. We also propose an experimental design approach for implementation.

We illustrate the model using data from a conjoint study concerning online movie rental services. The empirical results show that the proposed model has good fit and predictive validity. In addition, we find that marginal WTP in this category decays rapidly with quantity. We also find that the standard choice-based conjoint model results in anomalous WTP distributions with negative WTP values and non-diminishing marginal willingness-to-pay curves. Finally, we identify four segments of consumers that differ in terms of magnitude of WTP and volume potential and derive optimal quantity discount schemes for a monopolist and a new entrant in a competitive market.

Keywords: Quantity discounts; willingness to pay; choice models; mixed logit; conjoint analysis.

1. Introduction

Quantity discounts represent a popular pricing practice used by business-to-business and business-to-consumer companies. For example, Blockbuster charges $8.99, $13.99 and $16.99 for one, two and three DVDs out-at-a-time plans (see Blockbuster.com). Disney charges admission rates for Disneyworld that depend on the number of days. For a one day admission, Disney charges adults $79, and for ten consecutive days it charges $243 (see Disneyworld.com). Similarly, consumer goods companies often charge lower per-unit price for large packages of products, such as detergents, beers, and paper towels (Allenby et al. 2004). Based on a sample of 472 brands, Gerstner and Hess (1987) found that a large majority, 91.5%, were sold at a quantity discount in a supermarket in North Carolina. Other examples include print advertising rates that vary with respect to the number of ads placed per year and express mail service rates that depend on shipment volume. One key aspect of quantity discount pricing is that the per-unit or marginal price declines with a higher purchase quantity.[1]

From a demand perspective,[2] the rationale for quantity discounts is that often consumers' marginal willingness-to-pay (WTP) decreases with increasing quantity. A pricing scheme that mirrors consumers' WTP patterns is more profitable to the firm than a mere uniform price that charges the same price regardless of the number of purchased units (Dolan and Simon 1997). A second rationale for quantity discounts is consumers' heterogeneity in WTP: heavy users have higher marginal WTP for large quantities than light users (Dolan and Simon 1997 p. 174; Wilson 1993). Thus knowledge of *consumer-level* WTP for successive units of a product or service is critical for designing optimal quantity discount schemes.

[1] There are several forms of non-linear pricing such as multi-part tariff, multi-block tariff, and price points (see Dolan and Simon 1997, p. 164). In this paper, we focus on the price-points form of quantity discounts.
[2] There is also a supply side rationale for quantity discounts that stems from the supplier's cost savings (e.g., reduced production, inventory, and transportation costs) when selling larger quantities (Dolan and Simon 1997).

Conjoint analysis (Green and Srinivasan, 1990) has been gainfully utilized to assess the impact of price on demand and estimate consumer WTP for products and services. Kohli and Mahajan (1991) introduce an approach for measuring reservation price, which corresponds to the price that equates the utility of a new product to that of a status quo product. Jedidi and Zhang (2002) further develop this method to allow for the effect of new product introduction on category-level demand. Chung and Rao (2003) and Jedidi et al. (2003) describe methods for estimating consumer WTP for product bundles. More recently, Ding et al. (2005) and Park et al. (2008) propose incentive-compatible conjoint procedures for eliciting consumer WTPs for product attributes. Miller et al. (2011) compare the performance of four commonly used approaches to measure consumers' WTP to real purchase data. They find that conjoint analysis does well in inferring the true demand curves and determining the right pricing decisions.

Most pricing applications of conjoint analysis do not include quantity as an attribute in the design. They implicitly assume that a consumer buys one unit of a product at a single price and that consumer purchase rates do not depend on price (Iyengar et al. 2008; Kim et al. 2004). While it may seem trivial to add quantity as a factor, there are several design and analysis issues that traditional conjoint may encounter when estimating WTP for successive units of a product.

First, the traditional conjoint design requires as many price factors as quantity levels (i.e., a price factor for the first unit and a discount factor for each of the subsequent quantity levels). For example, for a product with six quantity levels, one needs to create six corresponding price factors. If each price/discount factor has three levels, then the full factorial is 6×3^6. Thus a traditional conjoint design may work in situations where the range of quantity offered is limited, but is not efficient when the range is large, making the respondent task tedious. Second, the conjoint part-worth function, while flexible, may not result in WTP measures that are positive, monotonic in quantity, and characterized by diminishing return. These properties are required for a proper WTP estimation (see Haab and McConnell 1998). Failure to enforce these constraints can lead to nonsensical measures of WTP and erroneous demand curves. For example, in a conjoint study on midsize sedans, Sonnier et al. (2007) obtain negative WTP estimates for between 13% and 23% of the participants. In our study, the standard choice-based conjoint model resulted in only two respondents (out of 250) with WTP estimates that satisfy the constraints of diminishing marginal WTP and positivity.

Recently, a few models have been proposed to account for volume in conjoint analysis. Kim et al. (2004) introduce a volumetric conjoint model in which product attributes are related to satiation parameters. Iyengar et al. (2008) propose a choice-based conjoint model which infers consumer usage levels as functions of the product features and the price components of a three-part tariff. Schlereth et al. (2010) use a WTP function approach to derive optimal two-part tariffs. However, none of these models is built to directly handle quantity discounts in conjoint analysis.

In this paper, we build on this emerging literature and propose a choice-based conjoint model for estimating consumer-level WTP values for varying quantities of a product and for designing optimal quantity discount pricing schemes. Our model can handle large quantity values and produces WTP estimates that are positive and increasing in quantity at a diminishing rate. In particular, we propose a tractable utility function which depends on both product attributes and product quantity and which captures diminishing marginal utility. We show how the proposed utility function embeds two standard utility functions used in the quantity discount literature as special cases. One attractive feature of the proposed utility function is the decomposition of the WTP function in terms of WTP for the first unit (which captures price premium) and a WTP multiple (which captures volume potential).

We also propose an experimental design approach for implementation that does not entail as many quantity and price factors as required by a standard conjoint design. Two critical features of the design are needed for determining the WTP values for different quantities of a product: (i) The experiment must include purchase quantity of the product as an attribute, and (ii) All choice sets in the conjoint experiment must include the no-purchase option. This latter feature is critical for obtaining unambiguous dollar-metric estimates of WTP.

We test our proposed model using data from a conjoint experiment involving consumer choice of online movie rental plans and compare our WTP distributions to those obtained from a standard choice-based conjoint (CBC) model. We find that the marginal WTP in this category decays rapidly with quantity. We also find that the standard CBC model results in anomalous WTP distributions with negative WTP values and non-diminishing marginal WTP estimates. For example, 16% of the respondents would not purchase a one DVD plan from Netflix when offered for free even though 63% of these respondents are current Netflix subscribers. We identify four segments of consumers that differ in terms of their WTP premium and purchase volume potential. An online movie rental company could use such information to target its customers based on their value potential to the firm. Finally, we use the parameter estimates to characterize consumer demand for online movie rental services and to design optimal quantity discount schemes that maximize gross contribution.

The rest of the paper is organized as follows. In Section 2, we describe the proposed model. In Section 3, we report an application of the model to the pricing of online movie rental services. Section 4 discusses the empirical results. In Section 5, we use the estimation results to characterize consumer demand for online movie rental services and in Section 6 we use them to derive optimal quantity discount schedules. Section 7 concludes the paper.

2. The Conjoint Model

In this section, we first present the indirect utility model and describe the form of the (pre-purchase) transaction utility that a consumer receives from purchasing multiple units. We then show how the proposed transaction utility function embeds two standard utility functions used in the literature as special cases. Next, we derive the WTP function that describes the maximum amount that a consumer is willing to pay for a given quantity of a product (Wilson 1993). Finally, we present the Bayesian multi-level procedure we use for model estimation.

2.1 The Indirect Utility Function

Consider a choice set consisting of J alternatives. Each choice alternative j (j=1, …, J) represents a product or a service that is described in terms of attribute levels, product size or quantity of service offered, and price (e.g., a two DVDs out-at-a-time movie rental plan from Blockbuster for $13.99 a month). Thus in contrast to standard conjoint analysis, consumer choice is based on both product attributes and quantity offered. Embedding such a quantity component in the conjoint design is a critical part of our measurement of consumers' WTP for successive units.

Let q_j be the quantity offered for product alternative j. We assume that consumer i (i=1, …, I) cannot choose more than one alternative. Let $p(q_j)$ be the price associated with q_j units of product j. The price schedule $p(q_j)$ represents a quantity discount scheme whereby the marginal price successively decreases with quantity. For example, Netflix charges a monthly fee of p(1)=$9.99 for a one movie at-a-time plan and p(2)=$14.99 for a two movie at-a-time plan. We specify the following indirect utility function for consumer i and product j:

(1) $$u_{ij}(q_j, y_i, p(q_j)) = v_{ij}(q_j) + \beta_i(y_i - p(q_j)) + \eta_{ij},$$

where $v_{ij}(q_j)$ is the transaction utility that consumer i associates with q_j units of product j, $\beta_i > 0$ is the income effect or price sensitivity, and η_{ij} is an error term that is observable to consumers but unobservable to the researcher. This error structure implies that the researcher cannot perfectly predict what a consumer will choose (Chandukala et al., 2007; p. 115).[3]

The form of the indirect utility specified in Equation (1) can be derived from a quasilinear utility function, which is free of wealth effects. This assumption is reasonable for products and services whose price is relatively small compared to the total budget but may not be adequate for products whose demand depends on income such as cars (Nevo 2000, p. 518). We check for the robustness of such an assumption in our empirical application.

Before discussing our specification of the transaction utility $v_{ij}(q_j)$, it is important to note that the parameter β_i confounds two price effects: the informational and allocative effects of price (Rao 1984). The informational effect is positive since it captures the role of price as a quality signal. The allocative effect is negative since it captures the role of price as an expense. These two effects cannot be disentangled empirically unless two separate conjoint studies are conducted (e.g., Rao and Sattler 2003). While the decomposition of these effects is managerially relevant (e.g., assessing the value of price as a signal), the decomposition is not necessary for the purposes of inferring consumer WTP (see Miller et al. 2011, p. 174).

2.1.1 Transaction Utility Specification

In line with utility theory, we assume that the transaction utility consumer i associates with q_j units of product j, $v_{ij}(q_j)$, increases with quantity, but at a decreasing rate. Such an assumption of diminishing marginal utility or satiation has been a cornerstone of economics and psychology (e.g., Baucells and Sarin 2007). In addition, we set the utility of zero units to zero (Dyer and

[3] Structural models of consumer choice (e.g., Nevo 2000) specify an additional error term (ξ_j), which corresponds to unobserved product characteristics for product j. This source of error, being observed to the firm and the consumer, induces an endogeneity problem in studies using observational data. In contrast to observational studies, endogeneity is not an issue in our model since the independent variables (price and other attributes) are manipulated experimentally in a conjoint study (Nevo 2000, p.528). In addition, our model captures any unobserved brand characteristics by specifying brand-specific fixed effects.

Sarin 1979). This means that the consumer derives no utility if s/he does not possess the product or subscribe to the service (i.e., $v_{ij}(0)=0$). Finally, we propose a discounted utility specification where the marginal utility from the qth unit (q > 1) of product j is a fraction of the marginal utility from the (q-1)th unit. Suppose for now that the utility decay constant is invariant across products i.e., $\lambda_{iq}^{j}=\lambda_{iq}\,(0<\lambda_{iq}\le 1)\,\forall\, j=1,\cdots,J.$ (Later we discuss the more general case where the decay constant can vary as a function of product features.) Then:

(2) $$v_{ij}(q_j)=v_{ij1}\sum_{k=1}^{q_j}\prod_{m=1}^{k}\lambda_{im}, \quad q_j\ge 1,$$

where $v_{ij1}>0$ is the transaction utility consumer i derives from the first unit of product j, $0<\lambda_{im}\le 1$ is the utility decay constant or the rate of satiation for the m[th] unit (m > 1) of product j. By definition, $\lambda_{i1}=1$. Thus, the marginal utilities from the first, second, …, and q[th] unit are $v_{ij1},\ \lambda_{i2}v_{ij1},\ \lambda_{i2}\lambda_{i3}v_{ij1},\ldots,\ \lambda_{i2}\lambda_{i3}\ldots\lambda_{iq}v_{ij1}$, respectively. The utility specification in Equation (2) is similar to utility functions used for the study of inter-temporal preferences (Koopmans 1960). Our discounting, however, is over quantity and not time.

As a special case, suppose the rate of satiation $\lambda_{im}=\lambda_i$ (for all m > 1) is constant. Then the marginal utilities from the first, second, …, and q[th] unit are $v_{ij1},\ \lambda_i v_{ij1}, \lambda_i^2 v_{ij1},\ldots,\ \lambda_i^{q-1}v_{ij1}$, respectively. Note that as $0<\lambda_i\le 1$, the marginal utilities are positive and decreasing. In addition, the closer λ_i is to 1 (0), the smaller (larger) is the diminishing of the marginal utility. Note that the utility function in Equation (2) reduces to $v_{ij}(q_j)=v_{ij1}$ when consumers buy only one unit of a product.

The total utility from q units in Equation (2) is the discounted sum of incremental utilities from each unit of product j and is additively separable across quantities. Such a utility specification is suitable for product categories characterized by diminishing marginal utility of consumption. This assumption holds for most products and is commonly made in the economics literature (e.g., Gerstner and Hess 1987; Wilson 1993). The utility specification, however, may not be suitable for addictive product categories (Gordon and Sun 2010) or products that command quantity premia. Furthermore, some product categories require a minimum number of units of the product for the purchase to have meaningful value to consumers. For example, shoes and earrings have no value unless bought in pairs. In these situations, one needs to redefine quantity q_j in terms of minimum purchase sizes (e.g., pairs).

Substituting Equation (2) for $v_{ij}(q_j)$ in Equation (1), we obtain the following, fully specified, indirect utility function for consumer i and product j:

(3) $$u_{ij}(q_j, y_i, p(q_j)) = v_{ij1}\sum_{k=1}^{q_j}\prod_{m=1}^{k}\lambda_{im} + \beta_i(y_i - p(q_j)) + \eta_{ij}.$$

Note that the above indirect utility function is nonlinear in quantity. In addition, the budget set implied by the quantity discount p(q) is convex. Both of these conditions imply that utility maximization can lead to an interior solution. That is, it is optimal for a consumer to spend a fraction of her budget on purchasing a certain quantity q of the product and the remaining budget on the composite good. See Online Appendix A for details.

2.1.2 Special Cases for Transaction Utility

Our proposed transaction utility is general and subsumes standard utility functions commonly used in the quantity discount literature as special cases (see Online Appendix B for a proof). Two such functions are the power utility function (e.g., Shugan 1985) and the quadratic utility function (e.g., Lambrecht et al. 2007). The power utility function is defined as:

(4) $$v_{ij}(q_j) = v_{ij1}q_j^{\beta_i}, \quad q_j \geq 1,$$

where $v_{ij1} \geq 0$ is the utility from the first unit and $0 < \beta_i \leq 1$ is a parameter that enforces the diminishing marginal utility. Translated in terms of Equation (2), the power function implies the following pattern of decay constants (see Online Appendix B):

(5) $$\lambda_{iq} = \frac{(q)^{\beta_i} - (q-1)^{\beta_i}}{(q-1)^{\beta_i} - (q-2)^{\beta_i}}, \quad q > 1.$$

Note that $\lambda_{iq} \geq 0$ increase with q. For example, if $\beta_i = 0.6$ then $\lambda_{i2} = 0.52$, $\lambda_{i3} = 0.87, \lambda_{i4} = 0.90, \ldots$. This means that the marginal utility decays at a slower rate with increasing quantity.

The quadratic utility function is defined as:

(6) $$v_{ij}(q_j) = \begin{cases} \beta_{i0} + \beta_{i1} q_j - 0.5\beta_{i2} q_j^2, & \text{if } q_j \leq \dfrac{\beta_{i1}}{\beta_{i2}}, \\ \beta_{i0} + \dfrac{(\beta_{i1})^2}{2\beta_{i2}} & \text{if } q_j > \dfrac{\beta_{i1}}{\beta_{i2}} \end{cases}$$

where β_{i0} is an intercept term and $\beta_{i1} \geq 0$ and $\beta_{i2} \geq 0$ are parameters whose ratio represents a utility threshold beyond which marginal utility is zero. The quadratic utility function implies the following pattern of decay constants (see Online Appendix B):

(7) $$\lambda_{iq} = \begin{cases} \dfrac{\beta_{i1} - 0.5\beta_{i2}((q)^2 - (q-1)^2)}{\beta_{i0} I_{q=2} + \beta_{i1} - 0.5\beta_{i2}((q-1)^2 - (q-2)^2)}, & \text{if } 2 \leq q \leq \dfrac{\beta_{i1}}{\beta_{i2}}, \\ 0 & \text{if } q > \dfrac{\beta_{i1}}{\beta_{i2}} \end{cases}$$

where $I_{q=2}$ is an indicator variable that takes a value of 1 if q=2 and zero otherwise. Note that when $\beta_{i0} = 0$, the decay constants λ_{iq} are a decreasing function of q. For example, if $\beta_{i0} = 0$, $\beta_{i1} = 3.0$, and $\beta_{i2} = 0.5$ then $\lambda_{i2} = 0.82$, $\lambda_{i3} = 0.77, \lambda_{i4} = 0.71, \ldots$. This means that the marginal utility decays at a faster rate with increasing quantity. When $\beta_{i0} > 0$, the decay constant λ_{iq} has a lower value at q=2 but then decreases afterwards.

A priori, we do not know whether the rate of satiation is constant, declining, or increasing over successive quantities. Past research on satiation has indicated that the rate at which consumers satiate differs across products, people, and contexts (e.g., Redden 2008). For instance, research on eating behavior has found that people satiate at a lower rate with food when they have wine or beer as aperitifs as compared to when they have water or fruit juice (Westerterp-Plantenga and Verwegen 1999). Thus the quadratic utility function maybe a better model for consumers in the water or fruit juice condition whereas the power utility function maybe a better model for those in the wine or beer condition. Restricting the analysis to either the power or the quadratic utility function may therefore result in an erroneous conclusion about the nature of consumer satiation.[4] The transaction utility function in Equation (2) is sufficiently flexible to capture any pattern of diminishing marginal utility.

To summarize, the power utility function implies an increasing λ pattern over successive quantities whereas the quadratic function implies a decreasing pattern. Therefore using either of these functions implicitly imposes a certain pattern of decay on the data. Thus, our utility specification can be useful in empirical applications where the pattern of decay is unknown *a priori* and/or where the researcher is interested in testing certain hypotheses about decay pattern.

2.2 The Willingness to Pay Function

Our interest in this paper lies in inferring consumers' willingness to pay (WTP) for a given quantity of a product or service. WTP or reservation price is the price that equates the expected utility of q units of product j to the expected utility of no-choice (see Jedidi and Zhang 2002).

Let j=0 denote the no-choice option. Then using Equation (1), the expected utility of allocating the whole budget to the composite good (i.e., no-choice) for consumer i is $u_{i0}(0, y_i, 0) = \beta_i y_i$ since $v_{i0}(0) = 0$. Thus, equating the expected utility of q units to the expected

[4] Simulation results show that the quadratic utility model does poorly in fitting data generated from a power utility model and vice versa.

utility of no-choice and solving for p(q) gives the following WTP function:

$$\mathrm{WTP}_{ij}(q_j) = w_{ij1}\sum_{k=1}^{q_j}\prod_{m=1}^{k}\lambda_{im}. \tag{8}$$

where $w_{ij1} = v_{ij1}/\beta_i$ is consumer i's WTP for the first unit. This function describes the maximum price a consumer is willing to pay for a given quantity of product j. It has desirable properties. It admits non-negative values and is an increasing function of quantity, but at a decreasing rate (Wilson 1993). In addition, the WTP for q units is a multiple of the WTP for the first unit, w_{ij1}. For example, in the special case $\lambda_{im} = \lambda_i$ for all m > 1, the WTP for q units is given by $w_{ij1} \times (1+\lambda_i+\lambda_i^2+\ldots\lambda_i^{q-1})$. One benefit of this property is that one can compute a WTP multiple for an infinite quantity q. For the case $\lambda_{im} = \lambda_i$, this multiple is $\frac{\lambda_i}{1-\lambda_i} = (1+\lambda_i+\lambda_i^2+\lambda_i^3+\cdots)$. To illustrate, for a consumer with $\lambda_i = 0.5$, this WTP multiple is equal to two. That is, this consumer is willing to pay a maximum of twice his or her WTP for the first unit for an offer with an infinite number of units. Thus one could segment consumers based on their WTP for the first unit, w_{ij1}, as well as their WTP multiple. In addition, one could score consumers based on their value potential which is the product of the WTP multiple and the WTP for the first unit. Note that we do not apply a time discounting when computing the WTP multiple because of the transaction (versus consumption) nature of our utility measurement.

The marginal WTP function gives the maximum price a consumer is willing to pay for the qth (i.e., incremental) unit of product j. It is given by

$$\mathrm{MWTP}_{ij}(q_j) = \begin{cases} w_{ij1} & \text{if } q_j = 1, \\ w_{ij1}\prod_{m=1}^{q_j}\lambda_{im} & \text{if } q_j > 1. \end{cases} \tag{9}$$

Note that the MWTP for each incremental unit is a fraction of the MWTP of the preceding unit.

Jedidi and Zhang (2002, p. 1353) show that utility maximization is equivalent to surplus maximization when the utility function is quasilinear.[5] Thus maximizing $u_{ij}(q_j, y_i, p(q_j))$ in Equation (3) is equivalent to maximizing the consumer surplus function given by:

$$S_{ij}(q_j, p(q_j)) = WTP(q_j) - p(q_j) + \varepsilon_{ij} = w_{ij1}\sum_{k=1}^{q_j}\prod_{m=1}^{k}\lambda_{im} - p(q_j) + \varepsilon_{ij} = s_{ij}(q_j, p(q_j)) + \varepsilon_{ij}, \tag{10}$$

where $S_{ij}(q_j, p(q_j))$ is the consumer surplus (WTP – price) that consumer i derives from purchasing q_j units of alternative j and $\varepsilon_{ij} = \eta_{ij} / \beta_i$ is a scaled error term. Note that as $v_{i0}(0) = 0$, the surplus from no-choice is $S_{i0} = \varepsilon_{i0}$.

2.3. Capturing the Impact of Product Attributes

To capture the impact of product attributes (e.g., brand name, product features) on WTP and ensure positivity of the WTP of the first unit, we reparametrize w_{ij1} as follows:

$$w_{ij1} = \exp(\sum_{l=1}^{L}\alpha_{il} x_{jl}), \text{ for } i = 1, \ldots, I;\ j = 1, \cdots, J; \text{and } l = 1, \cdots, L, \tag{11}$$

where x_{jl} is the value of product j on attribute l, and α_{il} measures the impact of x_{jl} on w_{ij1}.

In Equation (2), we specified different decay constants for different quantities. This non-parametric specification works well for products sold in small quantities but is infeasible for products sold in large quantities. Moreover, we assumed a common decay constant for all products variants (j=1, …, J). While this assumption may be acceptable for some products (e.g., online DVD plans), it may not be reasonable for others (e.g., light versus dark beers). To

[5] For quasilinear utility functions, the income effect $\beta_i y_i$ is irrelevant to the choice decision since it is a consumer-specific constant across alternatives (see Haaijer, Kamakura and Wedel 2000).

accommodate these issues and ensure that the decay constants fall in the (0, 1] interval, we reparametrize them as a logistic function of both quantity and product attributes. That is:

$$\lambda_{iq}^{j} = 1/(1+\exp(\theta_{i0}+\theta_{i1}q+\theta_{i2}q^2+\sum_{l=1}^{L}\gamma_{il}x_{jl})), \text{ for all i, q} > 1. \tag{12}$$

where θ_{i0} is an intercept, θ_{i1}, θ_{i2} and γ_{il} (l=1, …, L) capture the impact of quantity and product features, respectively, on the alternative-specific decay constant. The specification allows different product variants (e.g., brands) to have different decay constants. In addition, it allows for different decay patterns of WTP over quantities. For example if both θ_{i1} and θ_{i2} are zero, then the decay rate is constant across quantities (i.e., $\lambda_{i}^{j} = 1/(1+\exp(\theta_{i0}+\sum_{l=1}^{L}\gamma_{il}x_{jl}))$). However, if θ_{i1} is positive (negative) and θ_{i2} is zero then the decay coefficient becomes smaller (larger) with increasing quantity (i.e., $\lambda_{iq}^{j} = 1/(1+\exp(\theta_{i0}+\theta_{i1}q+\sum_{l=1}^{L}\gamma_{il}x_{jl}))$). Such a specification captures the quadratic (power) utility model. In the empirical application, we will test these different nested versions as well as a more general non-parametric specification.

2.4 Model Estimation

Consider a sample of I consumers, each choosing at most one product alternative from a set of J alternatives. Let t indicate a choice occasion. If consumer i contributes T_i such observations, then the total number of observations in the data is given by $T = \sum_{i=1}^{I} T_i$. Let z_{ijt}=1 if the choice of alternative j is recorded for choice occasion t, otherwise, z_{ijt}=0. Let j=0 denote the index for the no-choice alternative. Thus, z_{i0t}=1 if the consumer chooses none of the alternatives.

We assume that consumers are surplus maximizers. On choice occasion t, let S_{ijt} = $S_{ijt}(q_j, p(q_j)) = s_{ijt}(q_j, p(q_j)) + \varepsilon_{ijt}$ and $S_{i0t} = \varepsilon_{i0t}$ denote the surplus from alternative j and the no-choice option, respectively. Thus, a consumer would choose alternative j on choice occasion t if it has the maximum surplus $\{S_{ijt} > S_{ikt},\ k = 0,\ \cdots,\ J,\ k \neq j\}$ and would choose none of the alternatives if the no-choice option (j=0) has the maximum surplus $\{S_{i0t} > S_{ijt}, j = 1, \cdots, J\}$.

We assume that ε_{ijt} follows an iid extreme value distribution.[6] Therefore, consumer i's choice probability for product j on choice occasion t, Pr_{ijt}, and no-choice probability, Pr_{i0t}, are given by:

$$\text{(13)} \quad Pr_{ijt} = \frac{\exp(\mu_i s_{ijt}(q_j, p(q_j)))}{1 + \sum_{k=1}^{J} \exp(\mu_i s_{ikt}(q_k, p(q_k)))} \quad \text{and} \quad Pr_{i0t} = \frac{1}{1 + \sum_{k=1}^{J} \exp(\mu_i s_{ikt}(q_k, p(q_k)))},$$

where, $\mu_i > 0$ is a scale parameter (see Ben-Akiva and Lerman 1985, pp. 104-105). The scale parameter μ_i is necessary because the price coefficient is normalized to one in the surplus Equation (10).

As we model consumer surplus, the parameter estimates directly provide the *indifference* reservation price that makes a consumer indifferent between buying and not buying a certain quantity (i.e., 50% chance of buying). We can also use the parameter estimates to calculate reservation prices that correspond to other levels of probability of purchase. For instance, we can compute a *floor* reservation price at or below which a consumer would buy q units of product j with almost certainty (e.g., 95% chance of buying). We compute this quantity by setting the no-choice probability (Pr_{i0t}) to 5% and solving for p(q). Similarly, we can compute a *ceiling* reservation price that would make a consumer almost certainly not buy the product (e.g., 5% chance of buying). Thus, we can compute a WTP range for each consumer and quantity level.[7]

For an individual i, let $\alpha_i = (\alpha_{i1}, \cdots, \alpha_{iL})', \gamma_i = (\gamma_{i1}, \cdots, \gamma_{iL})', \theta_i = (\theta_{i0}, \theta_{i1}, \theta_{i2})'$ and

[6] We assume that the errors are independent because of the cyclical design approach that we use for constructing the choice sets (see Section 3.1). This is consistent with past work in choice-based conjoint (e.g., Iyengar et al. 2008).
[7] Note that our model-based WTP range is distinct from the ICERANGE proposed by Wang, Venkatesh and Chatterjee (2007) as the latter arises from consumer-level uncertainty in WTP.

$\psi_i = (\alpha_i, \gamma_i, \theta_i, \mu_i)$ be the joint vector of parameters. We use the choice data to estimate the vector of parameters, ψ_i, for each individual. As it is not possible to obtain sufficient choice data to estimate separate models for each individual, we use a Bayesian multi-level structure (Gelman and Hill 2007) that specifies how the individual-level parameters vary in the population and thereby statistically pool information across individuals. We assume that:

(14) $$\psi_i \sim N(\overline{\psi}, \Sigma),$$

where $\overline{\psi}$ and Σ are population level parameters to be estimated.

The model parameters are estimated using a standard Bayesian estimation procedure using Markov Chain Monte Carlo (MCMC) methods (see Online Appendix C). This mixed logit procedure allows one to compute WTP measures as part of the MCMC iteration process and provides confidence intervals for WTP and MWTP for different quantities and at different levels of aggregation. Hence managers can use such information to design optimal quantity discount schemes or customized pricing strategies for each consumer or consumer segments.

3. An Empirical Application

We illustrate the model using data from a choice-based conjoint experiment on DVD movie rentals by mail. Subscribers to this service rent movies online, receive them in DVD format by mail, and return them by mail free of charge after watching. The sample consists of 250 consumers. The online DVD rental category was chosen for several reasons. DVD rental is a product category that most consumers are familiar with. In addition, consumers are familiar with the various DVD rental plans offered by the two major competitors (Netflix and Blockbuster).

3.1 Design of Conjoint Experiment

We used four attributes to create online movie rental plans (conjoint profiles): (1) Service provider, (2) Number of movies out-at-a-time offered under the plan, (3) Monthly price of the plan, and (4) Blu-ray movies availability (Yes, No). These are the same attributes that online movie rental companies (e.g., Netflix) use to describe their plans at the time of the study.

The service provider attribute has three levels: A hypothetical new service with the generic name MovieMail and the two leading brand names in the category (Netflix and Blockbuster online). These two leading brands jointly account for 77.42% market share of the online DVD rental market in 2008.[8] We included a hypothetical new service to examine the impact of brand name on the WTP curve. This new service was described to respondents as follows:

MovieMail.com is a new online movie rental service about to enter the market. Like Netflix and Blockbuster Online, MovieMail operates by mail and promises to have same movie selection, search capabilities, and mail delivery time.

Note that the attribute-level details of MovieMail (e.g., price) were not included in the description; however, they were included as treatment variables in the conjoint experiment.

The number of movies out-at-a-time q has three levels: Low (1 or 2 DVDs out-at-a-time), Medium (3 or 4 DVDs out-at-a-time), and High (5 or 6 DVDs out-at-a-time). Note that though each level has two values, respondents will see only one of these values in a particular DVD plan. For example, if the number of movies at a time is "Low" in a particular conjoint profile, then we assign the respondent a value of either one or two DVDs out-at-a-time randomly.

The monthly price attribute p(q) has two components: the base price level of the plan and the depth of quantity discount. The base price for a plan p is based on the price of the one DVD out-at-a-time plan and has three levels: Low ($5.99 or $6.99), Medium ($7.99 or $8.99), and High ($9.99 or $10.99). The monthly price for a plan with no quantity discount (i.e., uniform pricing) is $p(q)=p*q$. One way to capture quantity discounts is through $p(q) = p*q^b$, where $b<1$ (decreasing

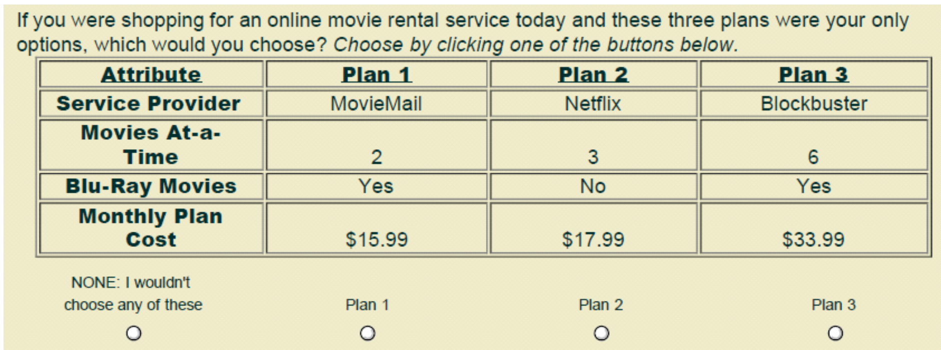
[8] The U.S. DVD video sales and rental market is valued at $7.6 billion in 2008, of which brick-and-mortar stores claimed 69% of the revenue share. Mail-order companies such as Blockbuster and Netflix together commanded 24% of the market, while kiosks had a mere 6% share and online streaming or download options an even smaller 1%. Thus Netflix and Blockbuster Online command 77.42% (=24/31) share of the online DVD rental market. See http://www.sramanamitra.com/2009/10/28/netflix-leads-video-market-forward.

block) measures the percent increase in total monthly price when quantity increases by 1%. We specify three levels for the depth of quantity discount: Low (b=.88 or .84), Medium (b=0.80 or .76), and High (b=.71 or .67). Similar to the number of movies out-at-a-time attribute, only one base price value and one quantity discount rate appear in a particular conjoint profile. When q=6, the b values correspond to the following quantity discount rates: Low (20% or 25% discount), Medium (30% or 35%), and High (40% or 45%). Suppose q=3, p=$9.99, and b=.88 for a particular conjoint profile. Then the monthly rate for such a plan is $p(q)=\$9.99*3^{.88}=\$26.14$.

Our experimental design has two novelties. First, because price and quantity can take large sets of values, we adopt a randomized-block-design type approach where we initially establish low, medium, and high intervals (the blocks) for each of the factors and then randomly assign specific values from the intervals for each respondent. This approach ensures that each quantity, price, and discount value is tested in the experiment. Second, unlike traditional conjoint, we do not specify different prices for different quantities. Instead, we decompose the price variable into two components: the price of the first unit and the depth of discount. This will result in a more parsimonious experimental design. In the context of our study where we have three brand levels, two levels for Blu-ray availability, six quantity levels, a full-blown traditional conjoint design would necessitate six price/discount factors (one for each quantity level). Assuming three levels for each price factor, this results in a $3x2x6x3^6$ design (26,244 profiles in the full factorial) whereas our design is only 3x2x3x3x3 (135 possible profiles). Even if one reduces the quantity levels to three, the traditional design still results in $486=3x2x3x3^3$ full factorial profiles. However, this design does not allow the testing of every quantity and price value.

We used a cyclic design approach for constructing choice sets (see Huber and Zwerina 1996). We first used Proc Optex in SAS to generate six orthogonal designs of 18 profiles from the $3^4 \times 2$ full factorial. For each orthogonal plan, we then used the cyclic design procedure to generate 18 choice sets with three online movie rental plans each.

Figure 1: An Example of a Choice Set


If you were shopping for an online movie rental service today and these three plans were your only options, which would you choose? *Choose by clicking one of the buttons below.*

Attribute	Plan 1	Plan 2	Plan 3
Service Provider	MovieMail	Netflix	Blockbuster
Movies At-a-Time	2	3	6
Blu-Ray Movies	Yes	No	Yes
Monthly Plan Cost	$15.99	$17.99	$33.99

NONE: I wouldn't choose any of these | Plan 1 | Plan 2 | Plan 3

Each participant in the study was randomly assigned to one of the six choice designs. After the conjoint task was explained, each participant was presented a sequence of 18 choice sets of movie rental plans in show-card format. The participant's task was to choose at most one of the three alternatives (i.e., no-choice is possible) from each choice set shown. See Figure 1 for an example of a choice set that we used in the conjoint experiment. We controlled for order effects by randomizing the order of profiles across subjects. We randomly selected 15 out of the 18 choice sets for model estimation and the remaining three for holdout prediction.

3.2 Descriptive Results

As part of the conjoint survey, we also collected information about respondents' demographics (e.g., income), their current movie rental provider, the type of plan they subscribe to, and how many DVDs they actually receive by mail in a month. Of the 250 respondents, 73.2% (25.2%) have Netflix (Blockbuster) as their current provider. The remaining 1.6% subscribe to other online movie rental companies. Overall, 29% of these respondents have a plan with one DVD out-at-a-time, 25% have a two DVDs out-at-a-time plan, 38% have a three DVDs out-at-a-time plan, 4% have a four DVDs out-at-a-time and the remaining 4% have 5 or higher DVDs out-at-a-time. This percentage breakdown compares very well with that reported in Feedfliks[9] and suggests that our sample is representative of online DVD rental users. Finally, we find that respondents on a plan with one DVD out-at-a-time receive an average of 4.9 DVDs per month from their service provider. Those with two, three, and four DVDs out-at-a-time plan receive an average of 7.7, 10.4, and 13.8 DVDs per month, respectively. Thus consumers with lower DVD plans are costlier to serve (per DVD) than those with higher plans.

3.3 Model Specifications

We used the data from the conjoint experiment to estimate four nested models. The models were selected to investigate various patterns in the decay of the marginal WTP for successive quantities. In all models, we initially specify decay constants that vary over quantities but not product features (e.g., brand). That is $\lambda_{iq}^{j} = \lambda_{iq}\,(0 < \lambda_{iq} \leq 1)\,\forall\, j = 1, \cdots, J$. Later, we generalize the models to allow the decay parameters to vary by product features as well. Let MM_j, NF_j, BB_j be 0/1 dummy variables indicating whether or not MovieMail, Netflix, Blockbuster, respectively, is the service provider of plan j. Let BR_j indicate whether Blu-ray movies are offered in plan j. Then the general model is specified as:

$$s_{ij}(q_j, p(q_j)) = w_{ij1} \sum_{k=1}^{q_j} \prod_{m=1}^{k} \lambda_{im} - p(q_j), \tag{15}$$

$$w_{ij1} = \exp(\alpha_{i0} + \alpha_{i1} NF_j + \alpha_{i2} BB_j + \alpha_{i3} BR_j), \tag{16}$$

where q_j is the number of DVDs out-at-a time offered under plan j and $p(q_j)$ is the monthly price for the plan. Note that MovieMail is used as the base service provider in Equation (16). Thus the brand coefficients should be interpreted relative to MovieMail.

[9] Feedfliks.com collects self-stated information from their registered users on various plan features such as number of DVDs at-a-time, the average rental period, typical queue sizes.

The models vary in terms of how we specify the decay coefficients. The most general model is non-parametric with decay coefficients represented by five separate coefficients. That is:

(17) $$\lambda_{\mathrm{im}} = 1/(1+\exp(\gamma_{\mathrm{im}})) \text{ for } \mathrm{m} = 2,\ldots,6,$$

where γ_{im} (m= 2, …, 6) are individual-specific parameters. Note that the logistic function ensures that the decay coefficients fall in the (0, 1] interval. We refer to this model (defined by Equations 15, 16, and 17) as the "Non-Parametric Decay Model." Because of its non-parametric form, this decay function is flexible. One drawback, however, is that the specification is not parsimonious especially in cases where the quantity variable takes a large set of values. In such cases, it is difficult to estimate a model with a decay coefficient for each quantity unit.

The next model is a nested parametric form where the decay constants are reparametrized as a quadratic function of quantity. That is:

(18) $$\lambda_{\mathrm{im}} = 1/(1+\exp(\theta_{\mathrm{i0}} + \theta_{\mathrm{i1}}\mathrm{m} + \theta_{\mathrm{i2}}\mathrm{m}^2)),\ m = 2,\cdots,6,$$

where the parameters θ_{i0}, θ_{i1} and θ_{i2} capture how the decay coefficients vary with quantity. For comparison purpose, we refer to the model in Equations (15), (16), and (18) as the "Quadratic Decay" model. A comparison of fit of this model relative to the non-parametric model provides evidence for the suitability of the parametric form of the decay function.

To test other patterns in the decay of the marginal WTP of successive quantities, we estimate two other nested versions. The first model sets both θ_{i1} and θ_{i2} to zero. In this model, for a consumer i, the decay coefficient is constant across quantities, i.e., $\lambda_{\mathrm{im}} = \lambda_{\mathrm{i}} = 1/(1+\exp(\theta_{\mathrm{i0}}))$ for m > 1. Thus, the marginal WTP for the first, second, third, … and qth unit are, respectively, $w_{\mathrm{ij1}}, \lambda_{\mathrm{i}} w_{\mathrm{ij1}}, \lambda_{\mathrm{i}}^2 w_{\mathrm{ij1}}, \ldots, \lambda_{\mathrm{i}}^{q-1} w_{\mathrm{ij1}}$. We call this model the "Constant Decay Model." The second model sets θ_{i2} to zero. In this model, the decay coefficient varies with increasing quantity. If

θ_{i1} is positive (negative), the decay coefficient becomes smaller (larger) with increasing quantity and hence the marginal WTP decays at a faster (slower) rate. We refer to this specification as the "Linear Decay Model." This model captures the WTP function from a quadratic (power) utility model if θ_{i1} is positive (negative). Note that the words "Linear" and "Quadratic" refer to the linear and quadratic terms, respectively, in the exponential function in Equation (18). They do not connote that marginal WTP decays linearly or quadratically.

4. Results

We used MCMC methods for estimating the models (see Online Appendix C). For each model, we ran sampling chains for 50,000 iterations. We assessed convergence by monitoring the time-series of the draws. We report the results based on 30,000 draws retained after discarding the initial 20,000 draws as burn-in iterations.

Table 1: Model Performance Comparison

Model Specification	**LML**[1]	LogBF	Holdout Hit Rate	Holdout LL[2]	Actual Plan Hit Rate
Non-Parametric Decay	-2144.45	-	70.2	-514.14	56.1
Quadratic Decay	-2127.95	16.50	71.2	-510.30	57.4
Linear Decay[3]	**-2099.33**	45.12	72.8	-485.58	60.1
Constant Decay	-2210.92	-66.47	69.7	-529.46	48.3

[1] LML denotes Log-Marginal Likelihood.
[2] Holdout LL denotes Holdout Log-Likelihood.
[3] Denotes "Selected Model."

4.1 Model Comparisons

We use log Bayes Factor (log BF) to compare the models. This measure accounts for model fit and automatically penalizes model complexity (Kass and Raftery 1995). In our context, log BF is the difference between the log-marginal likelihood of the non-parametric model (LML_{M1}) and that of a nested model (LML_{M2}). We use the MCMC draws to obtain an estimate of the log-marginal likelihood for each of the models. Table 1 reports the results for all four models.

Kass and Raftery (1995) suggest that a value of log BF= LML_{M2}-LML_{M1} greater than 5 provides strong evidence for the superiority of a model. Hence the LML results in Table 1 provide strong evidence for the superiority of the linear decay model relative to all other models.

The “Constant Decay” model where $\lambda_{im} = \lambda_i$ for all i and m performed relatively poorly (lowest LML). This result suggests that the rate of decay in marginal WTP, λ_i, is not constant over successive quantities. The superiority of the “Quadratic Decay” model over the “Non-Parametric Decay” model suggests that there is no need to estimate a decay coefficient for each quantity unit. Similarly, the superiority of the “Linear Decay” model over the “Quadratic Decay” model suggests that a linear specification is sufficient for capturing the pattern of decay in marginal WTP. Thus this parametric specification is not only parsimonious but does very well in capturing the shape of the WTP function.

4.2 Predictive Validity

To assess predictive validity, we calculate the holdout hit rate and validation log-likelihood (VLL) for each model. This latter statistics has been used in the Bayesian literature for assessing predictive validity (e.g., Montoya et al. 2010). The estimated parameters for each model were used to test that model’s predictive validity for holdout samples. Recall that the calibration data for each respondent included 15 choice sets and the holdout sample included three choice sets. The results in Table 1 indicate that the selected linear decay model has the highest holdout hit rate and VLL. The smaller differences in predictive validity among the linear, quadratic, and non-parametric decay models is expected since the former model is a special case of the latter models. However, the improvement in predictive validity for the linear decay model over the constant decay model is more noticeable when measured by VLL versus the holdout hit rate. This is expected since VLL is a more sensitive measure, which explains its use in practice.

As a further validation, we use the individual-level parameters and market prices for the available online DVD plans to predict the respondents' actual plans. The two major players in the market are Blockbuster and Netflix. In our sample, both companies account for 98.4% of the market. Blockbuster offers three plans (1 to 3 DVDs out-at-a-time) whereas Netflix offers four (1 to 4 DVDs out-at-a-time). Thus there are seven choices available to consumers. Using the MCMC draws, we predicted the choice probability of each subject for each of these plans given monthly fee and brand name. Consistent with the holdout task, we find that the "Linear Decay" model predicts real behavior well: a 60% hit rate compared to a 14% chance criterion and 35% maximum chance criterion. This performance fares well with the more general models and is superior to the "Constant Decay" model, which results in a 48% hit rate. See Table 1.[10]

4.3 Robustness Checks

We conducted two robustness checks. The first checks for the robustness of the quasi-linear utility specification of our model (see Equation 1). The second tests whether the decay parameters vary over product variants or brands.

Robustness of the Quasi-Linear Utility Assumption

We checked for the robustness of the quasi-linearity assumption by estimating two models with different non quasi-linear utility specifications. In each model, we use the linear decay reparametrization for the transaction utility, $v_{ij}(q_j)$, as it is empirically superior.

The first model specifies the utility of the outside good in a logarithmic form as follows:

$$u_{ij}(q_j, y_i, p(q_j)) = v_{ij}(q_j) + \beta_i \log(\theta_i y_i - p(q_j)) + \eta_{ij}, \tag{19}$$

where θ_i is a heterogeneous parameter that captures the proportion of income spent by the

[10] This validation exercise is rather a test of consistency than a test of predictive validity. A more stringent test entails a delayed holdout task involving real behavior and using a data collection format that is different from the one used in the calibration task. We thank an anonymous reviewer for raising these points.

consumer on the entertainment category. This specification has been used in the past by Sudhir (2001) to model consumer choice of automobiles.

The estimation results show that the logarithmic model has worse fit and predictive validity than our proposed quasi-linear utility model. Specifically, the log marginal likelihood (LML) of this model is -2135.38 and the Validation log likelihood (VLL) is -500.14. Both quantities are significantly worse than those of our proposed linear decay model (see Table 1).

The second model specifies the utility of the outside good in a power form as:

$$u_{ij}(q_j, y_i, p(q_j)) = v_{ij}(q_j) + \beta_i[\theta y_i - p(q_j)]^{\delta_i} + \eta_{ij}, \tag{20}$$

where $0 < \delta_i < 1$ is a parameter that captures diminishing marginal utility. This model reduces to our proposed indirect utility model in Equation (1) when $\delta_i=1$. We enforced the $0<\delta_i<1$ constraint by reparametrizing $\delta_i=1/(1+\exp(\tau_i))$. Note that under this reparamerization δ_i is restricted to be strictly less than 1. (This is because $\delta_i=1$ only when $\tau_i=\infty$.) Because of identification, we set the proportion of income spent on the entertainment category, θ, equal to 4%. This value is equal to the proportion of monthly income allocated to entertainment services, as reported in the New York Times (2008, February 10).[11]

The estimation results show that the parameter δ has a posterior mean estimate almost equal to one (mean=0.98 and posterior confidence interval is [0.97,0.99]). More importantly, the LML and VLL of this model (-2104.57 and -486.75, respectively) are slightly worse than those of the selected linear decay model (see Table 1).

In summary, the results of both analyses suggest that our assumption of quasi-linearity appear to be robust.

Assessing the impact of plan features on the decay parameters

[11] "You are what you spend." Available at : http://www.nytimes.com/2008/02/10/opinion/10cox.html

To test whether the decay constants vary across product variants, we re-estimated the models with decay parameters varying in terms of *both* quantity and plan features (i.e., MovieMail, Netflix, Blockbuster and Blu-ray). See Equation (12). In all four models, we find that none of the plan features significantly impact the decay parameters. In addition, all the LMLs are worse than the corresponding values reported in Table 1. For instance, the linear decay model with plan features in the decay coefficients has LML and VLL equal to -2109.3 and -493.87, respectively. Both quantities are significantly worse than those of the Linear Decay Model (see Table 1). Thus, in this application, the decay coefficients don't appear to vary across brands or affected by whether the plan has Blu-ray or not.

Table 2: Parameter Estimates: Posterior Means And 95 % Posterior Intervals

	Parameter	Parameter Label	Constant Decay	Linear Decay	Quadratic Decay	Parameter Label	Non-Parametric Decay
WTP First Unit Parameters	Intercept	α_0	**2.51***	**2.36**	**2.41**	α_0	**2.37**
			(2.39, 2.62)**	(2.27, 2.45)	(2.31, 2.52)		(2.28, 2.43)
	Netflix	α_1	**0.08**	**0.09**	**0.09**	α_1	**0.09**
			(0.05, 0.12)	(0.06, 0.13)	(0.06, 0.12)		(0.06, 0.12)
	Blockbuster	α_2	-0.02	-0.01	0.00	α_2	0.00
			(-0.06, 0.02)	(-0.04, 0.03)	(-0.03, 0.03)		(-0.04, 0.03)
	Blu-Ray	α_3	**0.05**	**0.06**	**0.06**	α_3	**0.06**
			(0.02, 0.08)	(0.03, 0.09)	(0.03, 0.09)		(0.03, 0.09)
Decay Parameters	Intercept	θ_0	-0.01	**-4.24**	**-4.99**	γ_2	**-1.19**
			(-0.24, 0.25)	(-5.68, -2.61)	(-8.31, -1.68)		(-1.59, -0.73)
	q	θ_1		**1.68**	**2.45**	γ_3	0.37
				(1.33, 1.90)	(0.19, 4.78)		(-0.19, 0.92)
	q^2	θ_2			-0.26	γ_4	-0.25
					(-0.65, 0.11)		(-2.19, 0.66)
						γ_5	-0.79
							(-4.59, 0.84)
						γ_6	-4.16
							(-4.59, -1.07)
Scale Parameter		μ	0.42	0.48	0.47	μ	0.45
			(0.33, 0.51)	(0.39, 0.58)	(0.39, 0.57)		(0.39, 0.52)

* Posterior mean for parameter. Coefficients for which zero lies outside the 95% interval are highlighted in boldface.

** 95% posterior confidence interval for parameter.

4.4 Parameter Values

We now discuss the parameter estimates from the models. Table 2 summarizes the posterior distributions of the parameters by reporting their posterior means and 95% posterior intervals. The middle panel of the table reports the estimates for the three parametric decay models whereas the right-most panel reports those for the non-parametric decay model.

Scale Parameter. All models resulted in scale parameter estimates that are statistically indistinguishable (i.e., their 95% posterior intervals overlap). As the scale parameter is inversely related to the price coefficient, this result indicates that all the model specifications are consistent in their estimate of price sensitivity.

WTP for First Unit. All the models produced parameter estimates that are similar in magnitude. Netflix (the market leader) has the highest mean part-worth value. The mean part-worth value for Blockbuster is not significantly different from that of the unbranded online movie rental service MovieMail, which we used as the base brand. The mean part-worth value for Blu-ray is positive and is significant (zero value is outside the 95% posterior interval). Translated in WTP values, for the selected Linear Decay model, consumers are willing to pay an average of \$12.47, \$11.35, \$11.37 (\$11.78, \$10.73, \$10.74) for a one DVD out-at-a-time plan with (without) Blu-ray that is offered, respectively by Netflix, Blockbuster, and MovieMail.[12] Thus, on average, consumers are willing to pay an additional \$1.10 for Netflix compared to MovieMail or Blockbuster and about \$0.65 to have movies in Blu-ray format. Currently, Netflix charges an additional \$2 for the Blu-ray option in the one DVD out-at-a-time plan while Blockbuster charges no additional fees. The free Blu-ray option may indicate a strategic move by Blockbuster to compensate for its weaker brand equity. See http://news.cnet.com/8301-

---

[12] The corresponding 95% posterior intervals for Netflix, Blockbuster, MovieMail with [without] Blu-ray are, respectively, (12.44, 12.49), (11.32, 11.37), (11.34, 11.39) [(11.75, 11.81), (10.69, 10.75), (10.71, 10.76)].

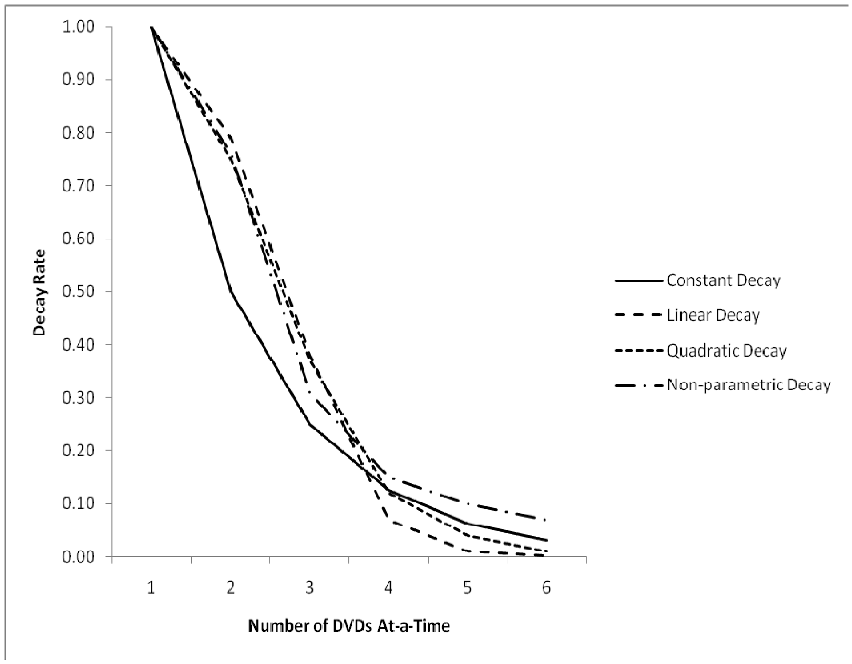
13506_3-10208093-17.html.

Decay Parameters. In the constant decay model, the decay parameter is not significantly different from zero. As $\lambda_{im} = \lambda_i = 1/(1+\exp(\theta_{i0}))$ for m >1, this means that the average decay rate in the sample is about λ=0.5. Thus the marginal WTP for the second unit is half of the WTP of the first unit, the marginal WTP for the third unit is one fourth of the first unit, and so on. One could use this decay pattern to estimate a WTP multiple by calculating the sum of the geometric series $1+\lambda+\lambda^2+\lambda^3+\cdots=1/(1-\lambda)$. Thus the constant decay model implies a WTP multiple of 2 $(=1/(1-0.5))$. That is, on average consumers are willing to pay a maximum of twice their WTP for the first unit for a plan that offers an "infinite" number of movies out-at-a-time.

The linear decay model shows that the decay rate increases rapidly with larger quantity. On average, we find that the marginal WTP for the second unit is 79% of the WTP of the first unit (i.e., the average of $\lambda_{i2} = 1/(1+\exp(\theta_{i0}+2\theta_{i1}))$ in the sample) and the marginal WTP for the third unit is $\lambda_2\,\lambda_3$ =38% of the WTP of the first unit. For the fourth, fifth, and sixth the marginal WTP is, respectively, $\lambda_2\,\lambda_3\,\lambda_4$=7%, $\lambda_2\,\lambda_3\,\lambda_4\,\lambda_5$=1%, and $\lambda_2\,\lambda_3\,\lambda_4\,\lambda_5\,\lambda_6$=0.1% of the WTP of the first unit. Summed over an infinite quantity, the Linear Decay model results in a WTP multiple of 2.25 $(=1+\lambda_2+\lambda_2\lambda_3+\lambda_2\lambda_3\lambda_4+\cdots)$. That is, on average consumers are willing to pay a maximum of 2.25 times the WTP for the first unit for a plan offering an "infinite" number of DVDs out-at-a-time.

The quadratic (non-parametric) decay model resulted in decay rates similar to those from the linear decay model. For the quadratic (non-parametric) decay model, the marginal WTP of the second, third, fourth, fifth, and sixth unit are, respectively, 0.75, 0.37, 0.12, 0.04, and 0.01 (0.76, 0.31, 0.15, 0.10, and 0.07) of the WTP of the first unit. For the quadratic (non-parametric) model the average WTP multiple is estimated to be 2.30 (2.40). Figure 2 depicts the decay rates for all the four estimated models. As the figure shows, all the models except the constant decay model have decay functions that are similar. The latter model appears to understate the decay rate for the first few units and overstate it for the larger units.

Figure 2: Decay Rates as a Function of Number of DVDs Out-at-a-Time



WTP Range. Recall that we can use the parameter estimates to calculate the floor (ceiling) reservation price below (above) which a consumer would almost certainly buy (not buy) a plan with q DVDs out-at-a-time. To illustrate, Table 3 reports the average floor and ceiling reservation prices for Netflix plans without Blu-ray that we obtained using the selected Linear Decay model parameter estimates. For completeness, the table also reports the average indifference reservation prices or WTP.

In summary, the empirical results show that the Linear Decay model has the best statistical fit and predictive validity. These results suggest that the marginal WTP in the online movie rental service category decays rapidly with quantity.[13]

[13] To test if there are any systematic differences between Netflix and Blockbuster customers, we estimated the models only on respondents who are current Netflix subscribers. The estimation results show that the parameters of the full sample and those of the Netflix sample are statistically indistinguishable.

Table 3: Floor, Indifference and Ceiling Reservation Prices for Netflix Plans without Blu-ray

Number of DVDs at-a-time	Floor Price ($)	Indifference Price ($)	Ceiling Price ($)
1	5.59	11.78	17.96
2	14.94	21.12	27.31
3	19.51	25.69	31.88
4	20.41	26.59	32.77
5	20.48	26.67	32.84
6	20.51	26.73	32.86

5. Demand Analysis

We now use the individual level parameter estimates to examine the extent of consumer heterogeneity in WTP and characterize consumer demand for online movie rental services.

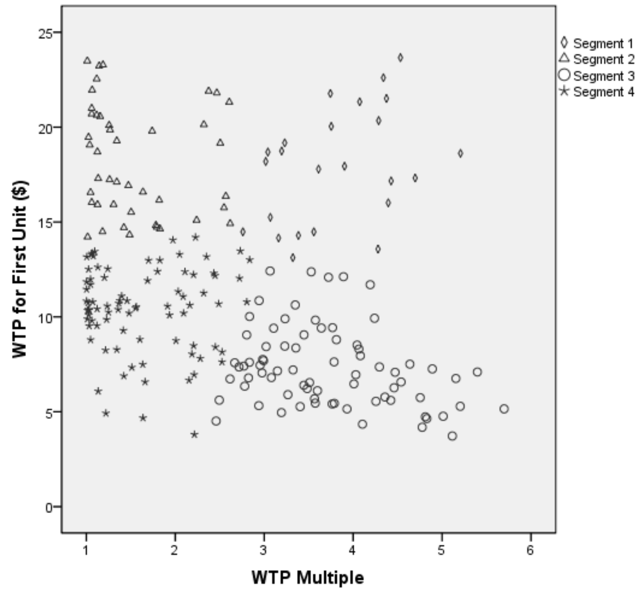
5.1 Consumer Heterogeneity

To explore the extent of heterogeneity in the WTP for the first unit and the WTP multiple in the sample, we used the MCMC draws of the selected linear decay model parameters to compute the posterior mean values of these statistics for each consumer in the sample. To illustrate, Figure 3 depicts the consumer-level estimates for Netflix. Across consumers, the average WTP for the first DVD without Blu-ray is $11.78 and the 95% heterogeneity interval is ($3.72, $23.67); the WTP multiple has a posterior mean of 2.25 and 95% heterogeneity interval of (1.0, 5.77).

We used K-Means clustering to segment consumers in our sample based on their mean WTP for first unit and WTP multiple.[14] We identified four segments of consumers shown in Figure 3 based on a scree plot of the percentage of variance explained by the clusters. To profile these segments, we use self-stated behavioral data we collected in our survey. Table 4 reports the descriptive statistics of the four segments.

[14] Equivalently, we could segment consumers based on their WTP for a one, two, …, and six DVDs out-at-a-time plans. We could also segment them based on their WTP range for successive units.

Figure 3: WTP for the First DVD from Netflix without Blu-ray and WTP Multiple



Segment 1 consists of 10.7% of the consumers in the sample who have a high WTP for the first unit (mean=$17.93) and a high WTP multiple (mean=3.81). Multiplying each consumer's WTP multiple by his or her WTP for the first unit is a measure of the value potential of the consumer. Thus segment 1 consumers are the most attractive with an average value potential of $68.99 per consumer. We call this segment the "High Value" segment.

Segment 2 consists of light users who have high WTP for the first unit. The mean WTP for the first unit in this segment is $18.22 and the mean WTP multiple is 1.54. 19.1% of the consumers belong to this segment. The average value potential per customer in this segment is $27.81. We label this segment the "High Premium" segment.

Segment 3, which represents 33.3% of the consumers, consists of heavy users with low willingness to pay. The mean WTP multiple for this segment is 3.72 and the mean WTP for the first unit is $7.31. Thus the average value potential per customer in this segment is similar to Segment 2 and is equal to $26.71. We name this segment the "High Volume" segment.

Segment 4, the least attractive segment, embodies 36.9% of the consumers. The mean WTP for the first unit is $10.40 and the mean WTP multiple is 1.63. Hence the mean value potential for this segment is $16.93. Hence, we call this segment the "Low Value" segment.

The segmentation results appear to be concordant with respondents' self-stated behavior. The high volume segments 1 and 3 currently subscribe to plans with higher number of DVDs out-at-a-time and appear to watch more movies than respondents in the low volume segments 2 and 4. These results shed some face validity for the proposed segmentation scheme.

Table 4: Segments' Description

Segment	1	2	3	4
Name	High Value	High Premium	High Volume	Low Value
Segmentation Bases:				
WTP of First Unit ($)	17.93*	18.22	7.31	10.40
WTP Multiple	3.81	1.54	3.72	1.63
Behavioral Descriptors:				
Number of DVDs in plan	3.50	2.09	2.94	1.76
Number of movies per week	3.83	2.80	3.65	2.61
Segment Value Potential ($)	68.99	27.81	26.71	16.93
Segment Size (%)	10.70	19.10	33.30	36.90

*Average across consumers in a segment

5.2 Willingness-to-Pay Distribution for Successive Units

To further explore heterogeneity, Figure 4a displays the cumulative WTP distribution for each successive DVD of an online movie rental plan without Blu-ray offered by Netflix (i.e., the percent of consumers whose WTP for the qth DVD is greater than a given price) that we obtained from the selected linear decay model. From the figure, we can determine that 68% of consumers have WTP greater or equal than $9.00 for the first DVD from Netflix. Similarly, 65.6% of consumers have WTP greater or equal to $5.00 for the second DVD. Note that, for the first DVD, one could determine the potential demand at any given price. The demand for the second DVD, however, depends jointly on the prices of the first and second DVD. Similarly, the demand for the qth DVD depends on the prices of 1, …, q DVDs. Consequently, the information in the figure should not be construed as demand curves for successive units.

Figure 4: Cumulative WTP Distributions for Successive Units

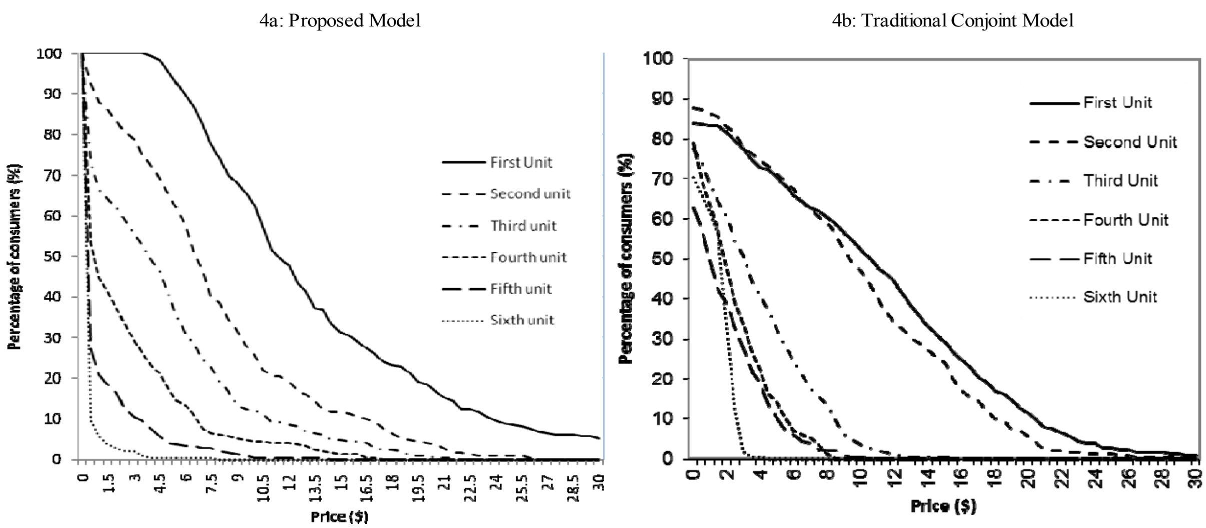


Figure 4b shows the corresponding WTP distributions that we obtained using the standard choice-based conjoint model. The figure illustrates the limitations of using traditional conjoint for measuring WTP over successive units that we discussed in the introduction.

First, at zero price, not all consumers purchase the online DVD service from Netflix. Thus there are 16% of consumers who would not purchase (i.e., have negative WTP for) a one DVD plan when it is offered for free even though 63% of these respondents are current Netflix subscribers. Similarly, 12.4%, 22.4%, 21.2%, 37.2%, and 29.6% would not purchase a second, third, …, and, sixth DVD, respectively, if offered for free. This anomalous result occurs because traditional conjoint does not constrain WTP to be positive. This finding is consistent with past research (e.g., Sonnier et al. 2007). In contrast, our proposed model ensures that WTP is always positive as illustrated in Figure 4a.

Second, note that the WTP curves in Figure 4b intersect each other. For instance, at a price of \$2.50, 79% of the consumers would purchase the first unit and 82% would purchase the second unit. This is anomalous because one would expect the demand for the second unit to be lower than the demand of the first unit. This happens because traditional conjoint does not impose diminishing marginal WTP. Researchers in Economics (e.g., Baucells and Sarin 2007; Wilson 1993) have emphasized the need to impose such a restriction. Without it, it is possible to find situations such as the one described above where consumers may be willing to pay a higher amount for a successive unit than a previous one. In contrast, our proposed model explicitly accounts for such a restriction.

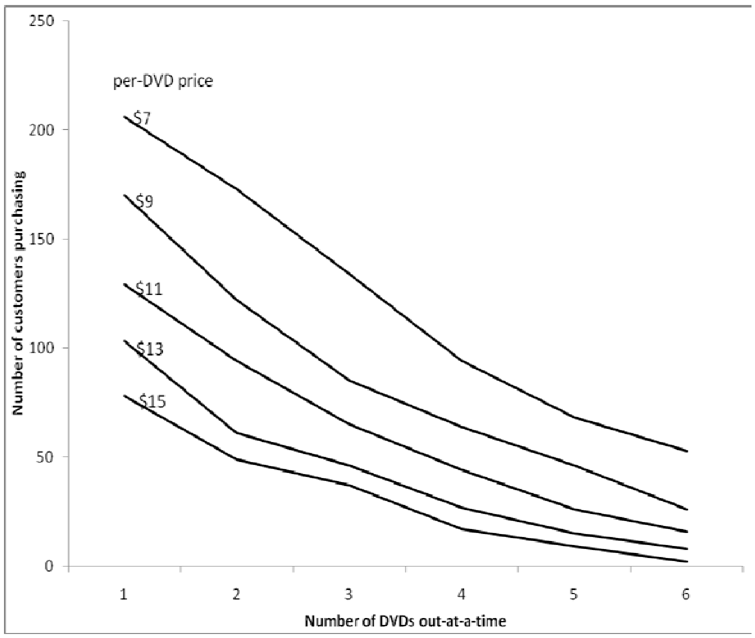
5.3 Demand Profile

One approach to depict the demand curve for successive units is to use Wilson's (1993, p. 50) demand profile method, which specifies for each (per-unit) price p, the number of consumers purchasing at least q units. Figure 5 presents the demand profile for Netflix online movie rental plans without Blu-ray for p=$7, $9, $11, $13, and $15.

For each price p, the demand profile represents the distribution of purchase sizes q at that price. For example, when the per-unit price is $9, 122 out of 250 (or 48%) consumers would be willing to buy Netflix plans with two or more DVDs out-at-a-time. Similarly, for each unit q, the demand profile reveals the distribution of marginal WTP for that unit. For example, 65 out of 250 (or 26%) consumers have marginal WTP greater or equal to $11 for the third (q=3) unit.

Following Wilson's (1993, p. 50), we used the demand profile information to compute the price elasticity for each successive unit of demand. For Netflix, we find an average price elasticity of -1.15 for the first DVD demand. That is if Netflix increases its price by 1%, its demand for the one DVD out-at-a-time would decrease by 1.15%. For the second and third DVD, we find an average elasticity of -1.42. For the fourth, fifth, and sixth DVD, the price elasticities are -1.96, -2.25, and -3.24, respectively. For comparison, we also computed the average price elasticities for Blockbuster. These elasticities are -1.17, -1.65, -1.79, -2.29, -2.57 and -3.57 for the one to six DVD out-at-a-time movie rental plans, respectively. As expected, consumers have higher price sensitivity for Blockbuster than they do for Netflix.

Figure 5: Demand Profile for Netflix



In summary, the demand analysis results illustrate the kind of managerial insights that can be derived from our proposed model. We now discuss how to use the estimation results to design optimal quantity discount schemes.

6. Quantity Discount Schedule Design

In this section, we use the demand profile method (Wilson 1993) to design a quantity discount schedule for a monopolist. Online Appendix D discusses the design of a discount schedule for a new entrant in a competitive setting. In both analyses, we assume that the DVD rental service is available to all consumers in the market and enjoys full awareness immediately after launch.

Suppose that Movie Mail is a monopolist and is considering offering four online movie rental plans. What quantity discount schedule should it offer? To examine this question, we need to estimate the variable cost that MovieMail would incur while serving customers in different plans. Currently Netflix incurs a marginal cost of $1.22 per rented DVD. This cost includes mailing, packaging, and royalty costs.[15] Recall that our survey results indicate that consumers rent on average 4.9, 7.7, 10.4, and 13.8 DVDs per month under a one, …, and four DVDs out-at-a-time plan, respectively (see Section 3.2). Thus assuming that MovieMail has a cost structure similar to Netflix, the plan-specific marginal costs would be c_1=$5.98 (=$1.22*4.9), c_2=$9.39, c_3=$12.69, and c_4=$16.84 for the one, two, three, and four DVDs out-at-a-time plans, respectively.

Table 5: Demand Profile of MovieMail and Optimal Discount Schedule for a Monopolist

	Demand for			
Price per DVD	1st DVD	2nd DVD	3rd DVD	4th DVD
$6	215	193	145	106
$7	190*	145	116	82
$8	173	127	**93**	**63**
$9	155	115	76	47
$10	139	**102**	58	40
$11	120	77	51	30
$12	106	63	42	26
$13	**91****	51	37	17
$14	78	42	27	12
Marginal Unit Cost	$5.98	$3.41	$3.30	4.15***
Optimal Gross Contribution	$638.82	$672.18	$437.10	$242.68
Optimal Marginal Price	$13	$10	$8	$8
Optimal Plan Price	$13	$23	$31	$39

* Reads as follows: 190 consumers would subscribe to a plan of one DVD or more if the per DVD price is $7.
** Entries in boldface have maximum gross contribution and correspond to optimal choices for the prices.
*** This is the difference between the cost of a four DVDs out-at-a-time plan and three DVDs out-at-a-time plan.

MovieMail will choose a price discount scheme that maximizes its gross contribution. To solve this problem, we use the price point method suggested by Wilson (1993). Table 5 reports the demand profile for MovieMail for q=1 to 4 DVDs and unit prices varying from $6 to $14 per

[15] Netflix reports that it mails about two million DVDs per day (Netflix Annual Report 2008). There are 313 mailing days (i.e., excluding Sundays) per year. Therefore Netflix ships a total of 626 million DVDs per year. The total subscription cost which includes mailing, packaging, and royalty fees is reported to $761,133,000. Therefore the cost per DVD is $1.22.

DVD. At unit price p=$6 for example, 215 consumers would subscribe to a one DVD out-at-a-time or more from MovieMail and 193 of these consumers would subscribe to at least a two DVDs out-at-a-time plan. Thus the number who would subscribe to exactly one DVD out-at-a-time plan is 22 (=215-193). The table also reports the marginal cost of each plan.

We use the demand profile and marginal cost information for each unit to determine the optimal price for each successive DVD that maximizes gross contribution. For the first DVD, the profit maximizing price is $13 with a gross contribution of $638.82 (= (13-5.98)*91). Thus under a monopolist scenario, MovieMail would achieve a market penetration of 36.4% (=91/250). That is 36.4% of the consumers would subscribe to *at least* one DVD plan if it is priced at $13. Similarly, for the second DVD, the profit maximizing price is $10 with a gross contribution of 672.18 (=(10-3.41)*102). Thus, for a two DVDs plan, the optimal price is $23 (=$13 + $10). Using the information from the table, we can estimate how many (among 91) consumers will subscribe to at least two DVDs plan. The table indicates that there are 77 consumers willing to pay $22 (=2*11) and 63 consumers willing to pay $24 (=2*12) for a two DVDs plan. Interpolating between these two demand predictions, we note that there are about 70 (among 91) consumers willing to pay at least $23 for a two DVD plan. Similarly, the optimal price for a three DVDs plan is determined to be $31. This price appeals to about 58 (among 70) consumers. Finally the optimal price for a four DVDs plan is $39 and would attract 40 (among the 58) consumers. Under this discount scheme, 21 consumers (or 8.4%) would subscribe to the one DVD plan; 12 consumers (or 5%) would subscribe to a two DVDs plan; 18 consumers (or 7%) to a three DVDs plan; and 40 consumers (or 16%) to a four DVDs plan. We obtain a similar discount schedule when we use finer price intervals with $0.50 increments.

For comparison, suppose MovieMail is entering a market where Netflix and Blockbuster are incumbents. Presently Netflix (Blockbuster) offers four (three) online movie rental plans. Both firms charge \$8.99, \$13.99, \$16.99 for the one, two, three DVDs plans, respectively. For the four DVDs plan, Netflix charges \$23.99. Suppose MovieMail decided to offer four online movie rental plans. Then the optimal prices under this competitive scenario are \$8.22, \$12.69, \$16.40, and \$21.82 for the one to four DVDs plans (see Online Appendix D). As expected, competitive prices are much lower than the ones under monopolist setting. For example, for 2 DVDs plan, a monopolist charges about \$11.50 per DVD (i.e., \$23 for the plan) whereas the new entrant charges about \$6 per-DVD (i.e., \$12 for the plan). Thus, compared to the monopolist case, consumers receive a price reduction of about \$5 per DVD because of competition.

7. Conclusions

Quantity discount pricing is commonly used by firms. This pricing scheme charges consumers a per-unit price that declines with purchase quantity. The critical information for designing such a quantity discount scheme is knowledge of consumers' WTP for successive units of a product.

In this paper, we propose a choice-based conjoint model for estimating consumer-level WTP values for varying quantities of a product. We use a novel utility function that embeds standard utility functions in the literature as special cases. The derived WTP function can handle large quantity values and allows WTP to be positive and to increase with quantity at a decreasing rate. A key benefit of this formulation is that it enables the segmentation of consumers in terms of WTP potential for the first unit (which measure price premium) and WTP multiple (a measure of volume potential) and the scoring of consumers based on their value potential to the firm. We also propose an experimental design approach for implementation that does not entail as many price/discount and quantity factors as required by a standard conjoint design.

We estimate four variants of the proposed model using data we collected in a conjoint experiment involving consumer choice of online movie rental plans. We find that the linear decay model has the best fit and predictive validity. Although parsimonious, the constant decay model has poorer performance, suggesting that WTP decays rapidly with larger quantity.

We use the parameter estimates to quantify the distribution of WTP and WTP multiple in the sample. For Netflix, the average WTP for a one DVD plan is about $11 and the average WTP multiple is 2.25, suggesting an average of about $300 value potential per customer per year. Additionally, we find that consumers are willing to pay an extra $1.10 for a one DVD plan from Netflix relative to MovieMail (a fictitious brand name) and an extra $0.65 for the Blu-ray option. Blockbuster, however, has no differential brand value relative to MovieMail.

We compare the WTP distributions from our model with those from a standard choice-based conjoint model. The latter model gives anomalous WTP distributions with negative WTP values and non-diminishing marginal WTP values. These results illustrate the need to enforce the positivity and the diminishing marginal utility in a conjoint model when estimating WTP.

We identify four consumer segments that vary in terms of WTP, WTP multiple, and value potential. An online movie rental company could use such information to target its customers based on their value potential to the firm. We also illustrate how to build a demand profile (Wilson 1993) using the consumer level WTP estimates. Finally, we use the parameter estimates to design an optimal quantity discount scheme that maximizes gross contribution. We consider two scenarios: a monopolist and a new entrant in a competitive market. As expected the optimal monopolist schedule resulted in higher per-DVD prices than the competitive schedule.

There are several avenues for future research. From a measurement perspective, our choice task didn't provide any incentive for respondents to be truthful. While CBC is found to do well in inferring WTP, future applications of the method should consider incentive alignment when collecting calibration and holdout choice data. In this paper, we assume a quasilinear utility function. While this assumption is standard in the WTP literature and is reasonable in the context of our application, it may not be reasonable for products whose demand is affected by income. Future research should generalize our model by fitting non quasi-linear utility functions. Finally, an area of managerial interest would be to apply the model in other product categories and especially in business-to-business settings where the quantity variable takes large values.

References

Allenby, Greg M., Thomas S. Shively, Sha Yang and Mark J. Garratt (2004), "A Choice Model for Packaged Goods: Dealing with Discrete Quantities and Quantity Discounts," *Marketing Science*, 23, 1, 95-108.

Baucells, Manel and Rakesh K. Sarin (2007), "Satiation in Discounted Utility," *Operations Research*, 55(1), 170–181.

Ben-Akiva, Moshe and Steven Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press, Cambridge, Massachusetts, USA.

Chandukala, Sandeep E., Jaehwan Kim, Thomas Otter, Peter E. Rossi and Greg M. Allenby (2007), "Choice Models in Marketing: Economic Assumptions, Challenges and Trends," *Foundations and Trends in Marketing*, 2, 2, 97-184.

Chung, Jaihak and Vithala R. Rao (2003), "A General Choice Model for Bundles and Multiple-Category Products: Application to Market Segmentation and Optimal Pricing for Bundles," *Journal of Marketing Research*, 40, May, 115-130.

Ding, Min, Rajdeep Grewal, and John Liechty (2005), "Incentive Aligned Conjoint Analysis," *Journal of Marketing Research,* 42 (1), 67-82.

Dolan, Robert J. and Hermann Simon (1997), *Power Pricing*, The Free Press, New York.

Dyer, James S. and Rakesh K. Sarin (1979), "Measurable Multiattribute Value Functions," *Operations Research*, 27, 810-822.

Ferjani, Madiha, Kamel Jedidi and Sharan Jagpal (2009), "A Conjoint Approach for Consumer- and Firm-level Brand Valuation," *Journal of Marketing Research*, 46, 6, 846-862.

Gelman, Andrew and Jennifer Hill (2007), *Data Analysis using Regression and Multilevel / Hierarchical Models*, Cambridge University Press, New York.

Gerstner, Eitan and James D. Hess (1987), "Why do Hot Dogs Come in Packs of 10 and Buns in

8s and 12s? A Demand-Side Investigation," *Journal of Business*, 60, 4, 491-517.

Gordon, Brett and Baohong Sun (2010), "A Dynamic Structural Model of Addiction, Promotions, and Permanent Price Cuts," Working paper, Columbia University.

Green, Paul E. and Srinivasan, V. (1990), "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice," *Journal of Marketing,* 54, October, 3-19.

Haab, Timothy C. and Kenneth E. McConnell (1998), "Referendum Models and Economic Values: Theoretical, Intuitive, and Practical Bounds on Willingness to Pay," *Land Economics*, 74, 2, 216-229.

Haaijer, Marinus E., Wagner A. Kamakura, and Michel Wedel (2000), "The 'No-Choice' Alternative in Conjoint Choice Experiments," *International Journal of Marketing Research*, 43 (1), 93–106.

Huber, Joel and Klaus Zwerina (1996), "The Importance of Utility Balance in Efficient Choice Designs," *Journal of Marketing Research*, 33, 307-317.

Iyengar, Raghuram, Kamel Jedidi and Rajeev Kohli (2008), "A Conjoint Approach to Multi-Part Pricing," *Journal of Marketing Research*, 45, 2, 195-210.

Jedidi, Kamel and Z. John Zhang (2002), "Augmenting Conjoint Analysis to Estimate Consumer Reservation Price," *Management Science*, 48, 1350-1368.

Jedidi, Kamel, Sharan Jagpal, and Puneet Manchanda (2003), "Measuring Heterogeneous Reservation Prices for Bundles," *Marketing Science*, 22, 107-130.

Kass, Robert E. and Adrian Raftery (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773-795.

Kim, Jaehwan, Greg M. Allenby and Peter E. Rossi (2004), "Volumetric Conjoint Analysis," Working paper, Ohio State University.

Kohli, Rajeev and Vijay Mahajan (1991), "A Reservation-Price Model for Optimal Pricing of Multiattribute Products in Conjoint Analysis," *Journal of Marketing Research*, 28, 3, 347-354.

Koopmans, Tjalling C. (1960), "Stationary Ordinal Utility and Impatience," *Econometrica*, 28, 2, 287-309.

Lambrecht, Anja, Katja Seim and Bernd Skiera (2007), "Does Uncertainty Matter? Consumer Behavior Under Three-Part Tariffs," *Marketing Science*, 26, 5, 698-710.

Miller Klaus M., Reto Hofstetter, Harley Krohmer and Z. John Zhang (2011), "How Should We Measure Consumers' Willingness to Pay? An Empirical Comparison of State-of-the-Art Approaches," *Journal of Marketing Research*, 48, 1, 172-184.

Montoya, Ricardo, Oded Netzer and Kamel Jedidi (2010), "Dynamic Allocation of Pharmaceutical Detailing and Sampling for Long-Term Profitability," *Marketing Science*, 29, 5, 909-924.

Nevo, Aviv (2000), "A Practitioners Guide to Estimation of Random Coefficients Logit Models of Demand," *Journal of Economics and Management Strategy*, 9, 513-548.

Park, Young-Hoon, Min Ding, and Vithala R. Rao (2008), "Eliciting Preference for Complex Products: A Web-Based Upgrading Method," *Journal of Marketing Research*, 45, October, 562-574.

Rao, Vithala R. (1984), "Pricing Research in Marketing: The State of the Art," *Journal of Business*, 57, 1, S39-S60.

Rao, Vithala R. and Henrik Sattler (2003), "Measurement of Price Effects with Conjoint Analysis: Separating Informational and Allocative Effects of Price," in *Conjoint Measurement. Methods and Applications*, Anders Gustaffsson, Andreas Herrmann, and Frank Huber, eds. Berlin: Springer, 47–66.

Redden, Joseph P. (2008), "Reducing Satiation: The Role of Categorization Level," *Journal of Consumer Research*, 34, 5, 624-634.

Schlereth, Christian, Tanja Stepanchuk and Bernd Skiera (2010), "Optimization and Analysis of the Profitability of Tariff Structures with Two Part Tariffs," *European Journal of Operational Research*, 206, 691-701.

Shugan, Steven M. (1985), "Implicit Understandings in Channels of Distribution," *Management Science*, 31, 4, 435-460.

Sonnier, Garret, Andrew Ainslie, and Thomas Otter (2007), "The Effects of Parameterization on Heterogeneous Choice Models," Working paper, University of California, Los Angeles.

Sudhir, K. (2001), "Competitive Pricing Behavior in the Auto Market: A Structural Analysis," *Marketing Science*, 20, 1, 42-60.

Wang, Tuo, R. Venkatesh, and Rabikar Chatterjee (2007), "Reservation Price as a Range: An Incentive-Compatible Measurement Approach," *Journal of Marketing Research*, 44, 200-213.

Westerterp-Plantenga Margriet S. and Christianne RT Verwegen (1999), "The Appetizing Effect of an Apéritif in Overweight and Normal-weight Humans," *The American Journal of Clinical Nutrition*, 69, 205-212.

Wilson, Robert (1993), *Nonlinear Pricing*, Oxford University Press, New York.

Online Appendix A: Interior Solutions

In this Appendix, we illustrate through an example how the consumer utility maximization problem subject to the budget constraint can lead to an interior solution even in the absence of error in the model. That is, it is optimal for a consumer to spend a fraction of her budget on a certain q^* units of the product and the remaining budget on the composite good.

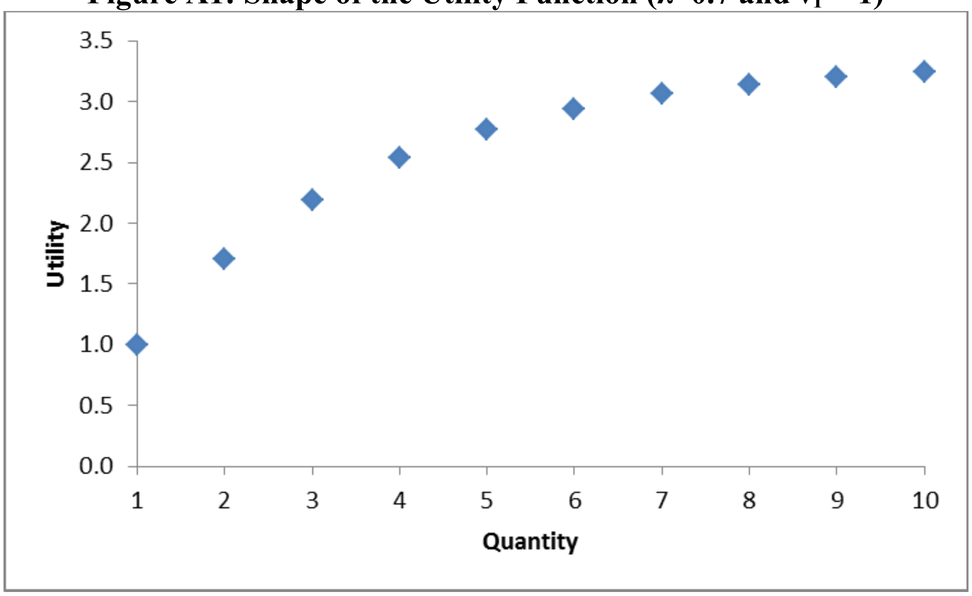
Utility Function

For simplicity, consider the constant decay transaction utility function, which we define as:

(A1)
$$\begin{aligned} v(q) &= v_1 + \lambda v_1 + \lambda^2 v_1 + \cdots + \lambda^{q-1} v_1, \\ &= v_1(1 + \lambda + \lambda^2 + \cdots + \lambda^{q-1}), \quad q \geq 1, \end{aligned}$$

where v_1 is the utility from the first unit and λ is the utility decay constant and where we suppressed the subscripts i, j for simplicity. Figure A1 plots this utility function when λ=0.7 and $v_1 = 1$.

Figure A1: Shape of the Utility Function (λ=0.7 and v_1 = 1)



Let f(q, w) be the direct utility a consumer obtains from q units of the product and w units of the composite good. The quasi-linear utility function is defined as:

(A2)
$$f(q, w) = v(q) + \beta(w),$$

where $\beta > 0$ is the income effect or price sensitivity. The budget constraint is defined as: $p(q) + w = y$, where p(q) is the price discount scheme, and y is the budget. Let y=\$5, $\beta = 1$, and the discount schedule $p(q)=q^{(1/2)}$. Then Figure A2 plots the indifference "curves" implied by this quasilinear utility function when the utility is set at f=4 and f=5 utils.

Figure A2: Indifference Curves

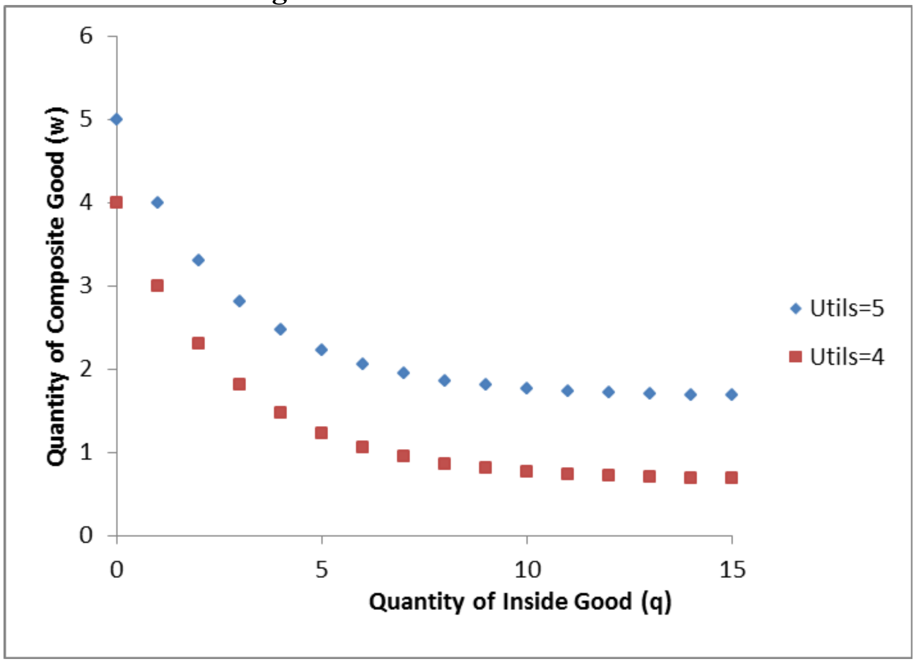
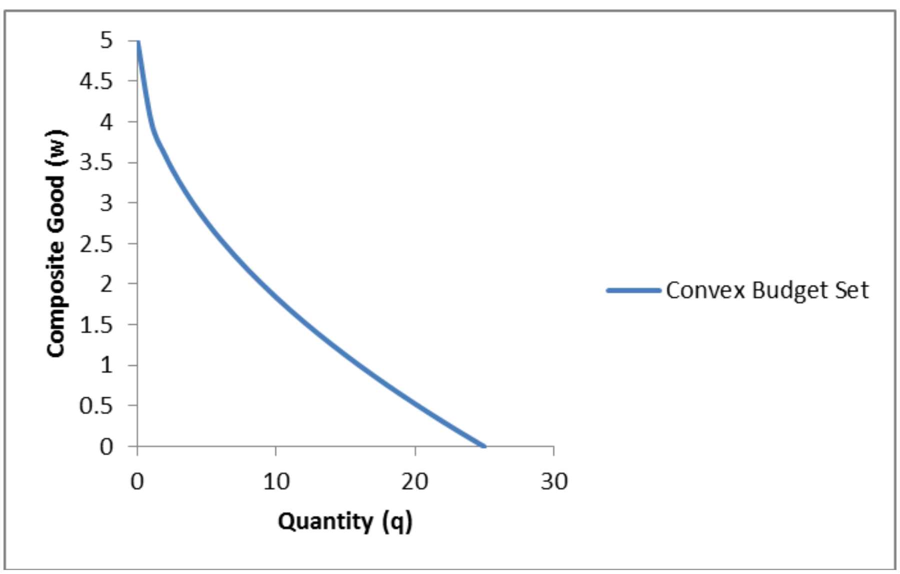


Figure A2 shows that the indifference curves implied by our utility function are nonlinear.

Budget Set

Recall that the budget constraint is defined as: $p(q) + w = y$. Figure A3 depicts this budget set for a quantity discount scheme $p(q)=q^{(1/2)}$ and y=$5. Note that the budget set is convex.

Figure A3: Budget Set from A Quantity Discount Scheme



Utility Maximization

We follow Allenby et al.'s approach (2004) to determine the optimal quantity q* of the inside good and w* of the outside good that maximize consumer utility. Allenby et al. suggest evaluating the utility function at discrete quantities to identify the optimal q* and w*. They warn against the use of Kuhn-Tucker conditions since they can lead to utility minimization instead of maximization. Specifically, in the abstract of their paper, Allenby et al. write:

"Utility maximizing solutions to economic models of choice for goods with either discrete quantities or nonlinear prices cannot always be obtained using standard first-order conditions such as Kuhn-Tucker and Roy's identity. When quantities are discrete, there is no guarantee that derivatives of the utility function are equal to derivatives of the budget constraint. Moreover, when prices are nonlinear, as in the case of quantity discounts, first-order conditions can be associated with the minimum rather than the maximum value of utility. In these cases, the utility function must be directly evaluated to determine its maximum."

Table A1 illustrates this grid search calculations when λ=0.7, $\nu_1 = 1$, y=\$5, p(q)= $q^{(1/2)}$ and $\beta = 1$. For each possible quantity value q, we calculate the utility which combines v(q) and the utility from the outside good w (i.e., w= y-p(q)).

Table A1: Overall Utility Value and Optimal Consumption

Quantity	v(q)	w= y-p(q)	Overall Utility v(q) + w
0	0.00	5.00	5.00
1	1.00	4.00	5.00
2	1.70	3.59	5.29
3	2.19	3.27	5.46
4	2.53	3.00	5.53
5	**2.77**	**2.76**	**5.54**
6	2.94	2.55	5.49
7	3.06	2.35	5.41
8	3.14	2.17	5.31
9	3.20	2.00	5.20
10	3.24	1.84	5.08
11	3.27	1.68	4.95
12	3.29	1.54	4.82
13	3.30	1.39	4.70
14	3.31	1.26	4.57
15	3.32	1.13	4.44

The results in Table A1 show that it is optimal to consume 5 units of the inside good and 2.76 units of the outside good. Additionally, the consumer spends \$2.24 for this level of consumption (p(q)=$5^{(1/2)}$) and the remaining \$2.76 (=\$5-\$2.24) on the composite good. The calculations clearly illustrate that our model can lead to interior solutions in which the consumer optimally allocate a fraction of her income to the inside good and the remaining balance to the composite good.

While our utility maximization can lead to interior solutions, it may also result in corner solutions. For example, if a consumer derives low utility from the inside good and/or if the price of the inside good is too high relative to its utility, then obviously the consumer would spend the total budget on the outside good (i.e., selects the no-choice option). Similarly, if there is no diminishing marginal utility (i.e., linear utility function with decay constant λ=1), then it is optimal for the consumer to spend the total budget on the inside good, regardless of whether the budget set is linear (no quantity discount) or convex (quantity discount). Note that the latter example is less likely to occur in our case because we impose diminishing marginal utility in our model.

Online Appendix B: Special Cases of Transaction Utility Model

In this Appendix, we show how the power and the quadratic utility functions, which are commonly used in the quantity discount literature, are special cases of our proposed transaction utility model in Equation (2). We first consider the power utility function and then the quadratic utility function.

The Power Utility Function

The power utility function specifies the utility that consumer i derives from q_j units of product j as follows:

(B1) $$\mathrm{v_{ij}(q_j)} = v_{\mathrm{ij1}} \mathrm{q_j^{\beta_i}}, \quad \mathrm{q_j} \geq 1,$$

where $v_{\mathrm{ij1}} \geq 0$ is the utility from the first unit and $0 < \beta_{\mathrm{i}} \leq 1$ is a parameter that enforces the diminishing marginal utility. For q=1, 2, 3, these utilities are $\mathrm{v_{ij}}(1) = v_{\mathrm{ij1}}$, $\mathrm{v_{ij}}(2) = v_{\mathrm{ij1}} 2^{\beta_{\mathrm{i}}}$, and $\mathrm{v_{ij}}(3) = v_{\mathrm{ij1}} 3^{\beta_{\mathrm{i}}}$ and the corresponding marginal utilities are, respectively, $\mathrm{mv_{ij}}(1) = v_{\mathrm{ij1}}$, $\mathrm{mv_{ij}}(2) = v_{\mathrm{ij1}}(2^{\beta_{\mathrm{i}}} - 1)$, and $\mathrm{mv_{ij}}(3) = v_{\mathrm{ij1}}(3^{\beta_{\mathrm{i}}} - 2^{\beta_{\mathrm{i}}})$.

Our proposed utility model delineates that the marginal utility from the qth unit (q > 1) of product j is a fraction λ_{iq} $(0 < \lambda_{\mathrm{iq}} \leq 1)$ of the marginal utility from the (q-1)th unit. Thus, represented in terms of our proposed utility function, the power utility function implies the following pattern of decay constants λ_{iq} $(\mathrm{q} > 1)$:

$$\lambda_{\mathrm{i2}} = \frac{\mathrm{mv_{ij}}(2)}{\mathrm{mv_{ij}}(1)} = \frac{v_{\mathrm{ij1}}(2^{\beta_{\mathrm{i}}} - 1)}{v_{\mathrm{ij1}}} = 2^{\beta_{\mathrm{i}}} - 1,$$

$$\text{(B2)} \quad \lambda_{i3} = \frac{mv_{ij}(3)}{mv_{ij}(2)} = \frac{v_{ij1}(3^{\beta_i} - 2^{\beta_i})}{v_{ij1}(2^{\beta_i} - 1)} = \frac{(3^{\beta_i} - 2^{\beta_i})}{(2^{\beta_i} - 1)}, \cdots$$

$$\lambda_{iq} = \frac{mv_{ij}(q)}{mv_{ij}(q\text{-}1)} = \frac{v_{ij1}(q^{\beta_i} - (q\text{-}1)^{\beta_i})}{v_{ij1}((q\text{-}1)^{\beta_i} - (q\text{-}2)^{\beta_i})} = \frac{(q)^{\beta_i} - (q-1)^{\beta_i}}{(q-1)^{\beta_i} - (q-2)^{\beta_i}}, \quad q > 1.$$

Therefore, the power utility function represents a special case of our proposed utility function where the pattern of the decay constants is given by the above Equation. Note that the decay constants are an increasing function of q. That is $\lambda_{i2} < \lambda_{i3} < \cdots$. This means that the marginal utilities for successive units decay at a slower rate.

The Quadratic Utility Function

The quadratic utility function specifies the utility that consumer i derives from q_j units of product j as follows:

$$\text{(B3)} \qquad v_{ij}(q_j) = \begin{cases} \beta_{i0} + \beta_{i1} q_j - 0.5 \beta_{i2} q_j^2, & \text{if } q_j \le \dfrac{\beta_{i1}}{\beta_{i2}}, \\ \beta_{i0} + \dfrac{(\beta_{i1})^2}{2\beta_{i2}} & \text{if } q_j > \dfrac{\beta_{i1}}{\beta_{i2}} \end{cases}$$

where $\beta_{i0} \ge 0$ is an intercept term and $\beta_{i1} \ge 0$ and $\beta_{i2} \ge 0$ are parameters whose ratio represents a utility threshold beyond which marginal utility is zero. For q=1, 2, 3, these utilities are $v_{ij}(1) = \beta_{i0} + \beta_{i1} - 0.5\beta_{i2}$, $v_{ij}(2) = \beta_{i0} + 2\beta_{i1} - 2\beta_{i2}$, $v_{ij}(3) = \beta_{i0} + 3\beta_{i1} - 4.5\beta_{i2}$ and the corresponding marginal utilities are, respectively, $mv_{ij}(1) = \beta_{i0} + \beta_{i1} - 0.5\beta_{i2}$, $mv_{ij}(2) = (\beta_{i1} - 1.5\beta_{i2})$, and $mv_{ij}(3) = (\beta_{i1} - 2.5\beta_{i2})$.

As λ_{iq} is the ratio of the marginal utilities from the qth unit and (q-1)th unit, the quadratic utility function implies the following pattern of decay constants:

$$\text{(B4)} \quad \lambda_{iq} = \frac{mv_{ij}(q)}{mv_{ij}(q\text{-}1)} = \begin{cases} \dfrac{\beta_{i1} - 0.5\beta_{i2}((q)^2 - (q-1)^2)}{\beta_{i0} I_{q=2} + \beta_{i1} - 0.5\beta_{i2}((q-1)^2 - (q-2)^2)}, & \textit{if } 2 \le q \le \dfrac{\beta_{i1}}{\beta_{i2}} \\ 0 & \textit{if } q > \dfrac{\beta_{i1}}{\beta_{i2}} \end{cases}.$$

Clearly, the utility quadratic function represents a special case of our proposed utility function where the pattern of the decay constants is given by the above Equation. Note that the decay constants are a decreasing function of q if $\beta_{i0} = 0$. That is $\lambda_{i2} > \lambda_{i3} > \cdots$. This means that the marginal utilities for successive units decay at a faster rate.

Online Appendix C: Model Estimation

In this Appendix, we describe model estimation. We estimate the surplus model in Equation (10) using a hierarchical Bayesian, multinomial logit approach. Consider a sample of I consumers, each choosing at most one product from a set of J products, where each product j=1, …, J is associated with q_j quantity. Let t indicate a choice task or occasion. If consumer i contributes T_i such observations, then the total number of observations in the data is given by $T = \sum_{i=1}^{I} T_i$. Let $z_{ijt} = 1$ if the choice of product j is recorded for choice occasion t; otherwise, $z_{ijt} = 0$. Let j = 0 denote the index for the no-choice alternative. Thus, $z_{i0t} = 1$ if the consumer chooses none of the products in choice task t. Let ψ_i be the joint vector of model parameters. Then the conditional likelihood, $L_i|\psi_i$, of observing the choices consumer i makes across the T_i choice occasions is given by

(C1) $$L_i \mid (\psi_i) = \prod_{t=1}^{T_i}\prod_{j=0}^{J} \Pr_{ijt}^{y_{ijt}},$$

where the $\Pr_{ijt}$ are the multinomial logit choice probabilities (see Equation 13).

To capture consumer heterogeneity, we assume that the individual-level regression parameters, ψ_i, are distributed multivariate normal with mean vector $\bar{\psi}$ and covariance matrix Σ. Then, the unconditional likelihood, L, for a random sample of I consumers is given by

(C2) $$L = \prod_{i=1}^{I}\int L_i \mid \psi_i f(\psi_i \mid \bar{\psi}, \Sigma) d\psi,$$

where $f(\psi_i \mid \bar{\psi}, \Sigma)$ is the multivariate normal $N(\bar{\psi}, \Sigma)$ density function.

The likelihood function in Equation (C2) is complicated because it involves multidimensional integrals, making classical inference using maximum likelihood methods difficult. We circumvent this complexity by adopting a Bayesian framework to make inferences about the parameters and using MCMC methods, which avoid the need for numerical integration. The MCMC methods yield random draws from the joint posterior distribution and inference is based on the distribution of the drawn samples.

For the Bayesian estimation, we use the following set of proper but noninformative priors for all the population-level parameters. Suppose $\bar{\psi}$ is a p × 1 vector and Σ^{-1} is a p × p matrix. Then the prior for $\bar{\psi}$ is a multivariate normal with mean $\eta_\psi = 0$ and covariance $C_\psi = \mathrm{diag}(10)$. The prior for Σ^{-1} is a Wishart distribution, W(R, ρ) where ρ = p+1 and R is a p × p identity matrix.

Online Appendix D - Quantity Discount Schedule for a New Entrant in a Competitive Market

In this Appendix, we describe how we determine the quantity discount schedule that MovieMail should offer in the presence of competition. Suppose MovieMail is a new entrant in a market in which Netflix and Blockbuster already offer online movie rental services. Presently Netflix offers four online movie rental plans whereas Blockbuster offers only three. Both Netflix and Blockbuster charge \$8.99, \$13.99, \$16.99 for the one, two, three DVDs out-at-a-time plans, respectively. For the four DVDs out at-a-time plan, Netflix charges \$23.99. At these prices, Netflix (Blockbuster) has a relative quantity-weighted share of 70% (30%).

Suppose MovieMail decides to offer four online movie rental plans. Suppose that the MovieMail service will be available to all consumers in the market and will enjoy full awareness immediately after launch. Suppose also that consumer switching costs are negligible and that Netflix and Blockbuster do not react to MovieMail pricing. Let p_1, p_2, p_3, and p_4 be MovieMail prices (decision variables) for the one, two, three, and four DVDs out-at-a-time plans. In addition, let c_1, c_2, c_3, and c_4 be the plan-specific marginal costs for the one, two, three, and four DVDs out-at-a-time plan, respectively. Please see Section 6 of the text for how we compute the marginal costs. Then, MovieMail chooses the discount pricing scheme $\mathbf{p} = (p_1, p_2, p_3, p_4)'$ to maximize the gross contribution:

(D1) $$\sum_{i=1}^{250}\sum_{q=1}^{4}(p_q - c_q)\Pr_{iq}(\mathbf{p}),$$

subject to the incentive compatibility constraint that the per-DVD price is decreasing over successive quantities:

(D2) $$\frac{p_q}{q} \leq \frac{p_{q-1}}{q-1},\ q = 2,\cdots,4,$$

where $\Pr_{iq}(\mathbf{p})$ is the choice probability of a MovieMail plan with q DVDs out-at-at-time by consumer i, i=1, …, 250 conditional on the discount pricing scheme **p** and the prices for the plans currently offered by Netflix and Blockbuster. We use the MCMC draws of the parameter to calculate these choice probabilities (see Equation 13). Note that, with MovieMail in the market, the choice set faced by consumers includes a total of eleven plan options: four new MovieMail plans, four Netflix plans, and three Blockbuster plans.

We maximize the constrained objective function in Equation (D1) using Proc NLP in SAS. The optimization results show that the optimal monthly price for the one, two, three, and four DVDs out-at-a-time plans are, respectively, \$8.22, \$12.69, \$16.40, and \$21.82.[16] These optimal prices are slightly lower than those offered by Netflix and Blockbuster which are, respectively, \$8.99, \$13.99, \$16.99, and \$23.99. At these prices, MovieMail expects to capture 15.25%, 8.37%, 8.20% and 4.75% market share for the one, two, three, and four DVDs out-at-a-time plans, respectively. The quantity-weighted shares of the online DVD rental market with MovieMail entry are as follows: MovieMail 33%, Netflix 49%, and Blockbuster 18%.

Note that these shares are derived under the assumption of full awareness, availability and no switching costs. Our optimization also assumes that Netflix and Blockbuster do not react to MovieMail entry by adjusting their prices and/or changing other attributes of their rental plans. While such an assumption is standard in conjoint simulators (Green and Srinivasan 1990), it does not lead to an equilibrium solution (e.g., Nash). A proper analysis should allow the incumbent firms to react to MovieMail entry as in Ferjani et al. (2009). However, such equilibrium analysis requires making assumptions about the nature of competition in the online DVD rental market.

[16] We checked for the global optimality of this solution using a grid search. Both Proc NLP and grid search resulted in virtually identical quantity discount schemes.