MANAGERIAL AND DECISION ECONOMICS
Manage. Decis. Econ. **32**: 1–15 (2011)
Published online 2 August 2010 Online Library
(wileyonlinelibrary.com) DOI: 10.1002/mde.1513

Confessions of an Internet Monopolist: Demand Estimation for a Versioned Information Good

Henry W. Chappell Jr*, Paulo Guimarães and Orgül Demet Öztürk

Department of Economics, University of South Carolina, Columbia, SC, USA

We develop and apply a method for estimating demand system parameters for versioned information goods. Our analysis uses data collected from a web-based field experiment in which prices and versions of an information good were exogenously varied. Using a maximum simulated likelihood (MSL) procedure, we estimate parameters characterizing distributions of utility functions over a population of potential buyers. We then construct profit-maximizing versioning and pricing plans for the seller and assess the welfare implications of those plans. Because firms increasingly have opportunities to collect information by tracking behavior of customers, methods similar to ours could be useful in future commercial applications.

1. INTRODUCTION

Firms with market power require knowledge of demand to determine profit-maximizing prices. Although demand estimation provides a classic problem in applied econometrics, the rise of Internet markets and the increasing importance of information goods change the nature of this problem in two important ways. First, because firms selling on the Internet can often track buyer behavior, they have increased opportunities to gather data characterizing the preferences of buyers.[1] Second, because information goods often have near-zero marginal production costs, pricing tends to be 'value-based' rather than 'cost-based.' Specifically, second degree price discrimination in the form of product versioning has become commonplace. In this setting, detailed information about the distribution of demands across buyers is available to sellers, and use of this information is important in the design of pricing and versioning choices. In this article, we describe a field experiment undertaken to collect information from potential buyers of an information good; we then develop and apply a method for estimating demand system parameters. Using the resulting estimates, we devise a profit-maximizing versioning and pricing scheme.[2]

In Section 2, we review the theory of product versioning as presented by Varian (2000).[3] In Section 3, we describe an econometric method for estimating parameters of demand systems for versioned information goods. To apply this method, we carried out a field experiment using a 'real-world' Internet business; the design of those experiments is described in Section 4.[4] The use of experimental data enables us to avoid issues of simultaneity and identification that are problematic for most demand estimation applications. In Section 5, econometric estimates are presented

*Correspondence to: Department of Economics, University of South Carolina, Economics, Columbia, SC 29208, USA. E-mail: chappell@moore.sc.edu

and discussed. Using these estimates, in Section 6, we present profit-maximizing versioning plans for the good, and we calculate how versioning would affect seller profits and market efficiency. Conclusions follow in Section 7, where we argue that our methodology has the potential to be a useful managerial tool in real-world commercial settings.

2. THEORETICAL BACKGROUND

This section presents a simple diagrammatic illustration of the versioning problem that follows the example of Varian (2000). We initially assume that there are two groups of customers (potential buyers) who differ in their willingness to pay for quality embodied in a single unit of an information good. We refer to these groups as 'type 1' (low-demand) and 'type 2' (high-demand) customers. Figure 1 illustrates the marginal valuation (of quality) schedules for individual members of these groups. Individuals' total valuations are measured by the usual areas beneath the marginal valuation schedules. We assume that the marginal cost of production of the good is zero, regardless of quality level.

Initially suppose that the single seller of the good can identify customers by type. For this case, type 1 customers would each be charged a price P_1 equal to area A for a version of the good with quality level S_1 and type 2 customers would be charged a price P_2 equal to area A+B+C for a version with quality level S_2 (in each case, customers are charged a price equal to total willingness to pay).

Now suppose that the seller cannot observe a customer's type. The seller might again consider producing versions with quality levels S_1 and S_2 to be sold at prices $P_1 = \mathrm{A}$ and $P_2 = \mathrm{A+B+C}$, hoping to extract all surplus from buyers. However, type 2 customers would not choose the bundle intended for them. By choosing the low-quality bundle, each high-demand customer can earn surplus equal to area B rather than a surplus of zero. Therefore, under this plan, the seller would earn only A from each customer.

However, the seller can do better. Suppose that the seller again charges $P_1 = \mathrm{A}$ for a low-quality good, but now charges a price $P_2 = \mathrm{A+C}$ for the high-quality good. Type 2 buyers are just willing to buy the high quality version, since they now retain surplus B from the purchase of either version. The seller receives A from each type 1 buyer and A+C from each type 2 buyer.

From the seller's viewpoint, this is an improvement, but is still not optimal. Suppose that the seller slightly reduces the quality of the low-quality version, as depicted in Figure 2. This reduces the revenue that can be obtained from a type 1 customer by the amount of the shaded triangle. However, because it makes the low-quality version less attractive, it increases the willingness of type 2 customers to pay for the high-quality version. The amount of the increase is measured by the area of the black trapezoid and, as the diagram shows, the gain from type 2 customers can exceed the loss from type 1 customers. To select an optimal version, the seller would reduce the quality of the low-quality version to the point where the marginal reduction in revenue from type 1 customers just equals the marginal gain in revenue from type 2 customers. For the case where there are equal numbers of the two customer types, that point occurs at quality level S_1^* in Figure 2.

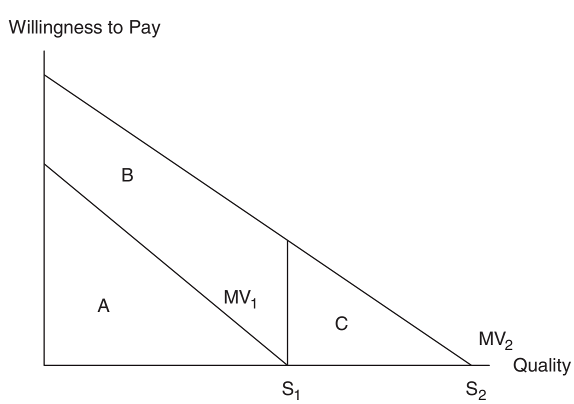

Figure 1. Optimal versioning.

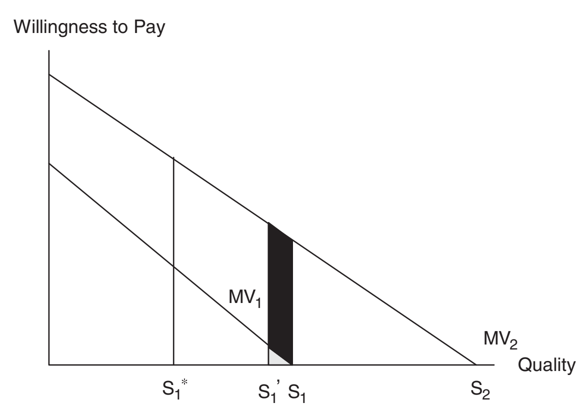

Figure 2. Optimal versioning.

DOI: 10.1002/mde

The preceding discussion demonstrates that versioning can be profitable when customers differ in their valuations of quality. However, the example relies on a special case where there are two distinct customer types with widely differing preferences. Moreover, the example demonstrates that, to construct a profit-maximizing versioning scheme, a monopolist needs to have considerable knowledge about the distribution of preferences across potential buyers. In the following section, we develop a model of consumer choice that can be estimated with experimental data and that can provide the information needed to solve a firm's versioning problem.

3. AN ECONOMETRIC MODEL

We begin by considering the choice problem facing a potential buyer, customer i, selected randomly from a population. Customer i has a utility function of the form $U_i = V_i S^{b_i}$, which associates a gross valuation, U_i (measured in 'dollars worth' units) with a versioned information good that has quality level S.[5] We assume that the utility function parameters satisfy the conditions $V_i \geqslant 0$ and $0 \leqslant b_i \leqslant 1$.[6]

Customers differ in their utility function parameters, V_i and b_i, which are randomly distributed over the population. The assumption that both utility function parameters are random is an important one—under plausible assumptions, if V_i is random, but b_i is not, then no profitable two-product versioning plans exist (Appendix A provides a proof).[7] Let $G(V_i,b_i,\boldsymbol{\theta})$ be the joint distribution function from which each customer's utility function parameters are drawn, and let $\boldsymbol{\theta}$ be a vector of parameters characterizing that distribution. Further, assume that the purchase options that customers face are individual-specific. Each customer is presented with two versions of a good, with prices and quality levels given by P_{1i}, P_{2i}, S_{1i}, and S_{2i}. We assume that version 2 is the higher quality product, and so $S_{2i} > S_{1i}$. The consumer can select version 1, version 2, or neither, with the utility-maximizing purchase choice determined according to conditions (1):

If $V_i S_1^{b_i} - P_1 \leqslant 0$ and $V_i S_2^{b_i} - P_2 \leqslant 0$, do not purchase (1a)

If $V_i S_1^{b_i} - P_1 > 0$ and $V_i S_1^{b_i} - P_1 \geqslant V_i S_2^{b_i} - P_2$, purchase version 1 (1b)

If $V_i S_2^{b_i} - P_2 > 0$ and $V_i S_2^{b_i} - P_2 > V_i S_1^{b_i} - P_1$, purchase version 2 (1c)

Let R_{0i} be the probability that an arbitrary customer i chooses *not* to buy either version, given his/her choice opportunities (described by P_{1i}, P_{2i}, S_{1i} and S_{2i}) and given the distribution from which the utility function is drawn. Similarly, let R_{1i} and R_{2i} be the probabilities that customer i purchases versions 1 and 2, respectively. These probabilities depend upon the functional form of the distribution G, the parameters of the distribution, and the prices and qualities of the goods available for purchase by customer i. In equation form:

$$R_{0i} = R_0(P_{1i}, P_{2i}, S_{1i}, S_{2i}, \boldsymbol{\theta}) \tag{2a}$$

$$R_{1i} = R_1(P_{1i}, P_{2i}, S_{1i}, S_{2i}, \boldsymbol{\theta}) \tag{2b}$$

$$R_{2i} = R_2(P_{1i}, P_{2i}, S_{1i}, S_{2i}, \boldsymbol{\theta}) \tag{2c}$$

Assuming that we observe the actual choices made by buyers in the sample and the options from which they choose, our task is to estimate the parameters $\boldsymbol{\theta}$. Denoting the observed purchase choices for buyer i by the dummy variables C_{0i}, C_{1i}, and C_{2i}, respectively indicating no purchase ($C_{0i} = 1$), purchase of version 1 ($C_{1i} = 1$), and purchase of version 2 ($C_{2i} = 1$), the likelihood function for an individual observation is given by:

$$L_i = R_{0i}^{C_{0i}} R_{1i}^{C_{1i}} R_{2i}^{C_{2i}} \tag{3}$$

and the likelihood for the sample of observations is the product of the individual likelihoods:

$$L = \prod_{i=1}^{N} L_i \tag{4}$$

The maximum likelihood method requires that we find values for $\boldsymbol{\theta}$ that maximize L.

A difficulty with the estimation procedure outlined above is that it is not possible to derive closed-form solutions for the probability functions R_0, R_1, and R_2 in conditions (4) for all possible specifications for G, nor is it desirable to limit forms for G to those for which analytical solutions for the probabilities are possible.[8] However, it is possible to use simulation methods to obtain numerical rather than analytical calculations of the required probabilities. Our method is that of maximum simulated likelihood (MSL) as described by Train (2003) and others.[9] A more

DOI: 10.1002/mde

detailed discussion of estimation is provided in Appendix B.

4. A MARKET EXPERIMENT

This section describes a market experiment we have undertaken to investigate the demand for a versioned information good. The experiment was specifically designed to generate data suitable for the estimation method described in Section 3. Because the MSL estimation method we have described is simulation-based, there are essentially no restrictions imposed on permitted utility function forms or on the distributions from which utility function parameters are drawn. However, data requirements are potentially demanding.

Our econometric model assumes that we observe a sample of individual consumers' choices and that these consumers have been presented with varied price and quality options. For information goods sold via the Internet, such data may not be difficult for sellers to gather. For example, Amazon identifies its regular customers when they visit the site, keeps track of their purchase histories, and offers personalized suggestions for shopping. Under these conditions, it would be feasible to offer customers personalized prices (perhaps in the form of discount offers) in order to learn about demand characteristics.[10]

For academic researchers, obtaining such information is more difficult. Firms may be willing to experiment with prices and versions, but they are not likely to publicize their experiments or release the data they produce. The data employed in this paper were generated in a field experiment carried out by an existing Internet business, Homework Hero (www.homeworkhero.com). Homework Hero is operated by one of the authors of this paper, who undertook the market experiment for the express purpose of obtaining data for academic use.[11]

Homework Hero is a service offered to K-12 schools that permits teachers to post homework assignments for their students on the web.[12] Each teacher using the service maintains a web page on the Homework Hero site and can update assignments posted there through a web-based form. The service provided by Homework Hero is normally purchased by schools rather than by individual teachers. Our market experiment involved selling a complementary product, a collection of personalized digital images, directly to teachers.

The Homework Hero web site is organized so that each subscribing school has a set of school-specific pages and a database that holds the content of assignment pages for the school's teachers. Within a school, each teacher's assignment page uses a common template that determines the style and layout of the page, but individual teachers provide page content. Teachers have options that permit them to post images and use HTML (Hypertext Markup Language) to create a personalized appearance. In practice, many teachers go to some effort to customize the appearance of their pages with both HTML formatting and the display of images.

In our experiment, we sold bundles of images for teachers to display on their assignment pages.[13] Each of the images in a bundle included a graphic rendering of the teacher name, with the style, color scheme, animation, and theme of the images varying. The image collections offered opportunities for personalization of a teacher page, and having larger collections would permit teachers to vary the appearance of the page over time. Some of the offered bundles included images with seasonal themes, and so it would be natural to display them in sequence over the academic year. Hereafter, we assume that the quality of a bundle of images is indicated by the number of images it contains.[14]

Homework Hero had significant market power in the sale of these image collections. The images sold were customized for use on Homework Hero and were personalized for buyers. Further, Homework Hero provided storage space for the images sold. While it is possible for users to display other images on assignment pages, it would have been difficult for a teacher to replicate the product sold in the experiment.

A total of 70 subscribing schools were selected for market experiments.[15] The selected schools typically had large numbers of teachers actively maintaining assignment pages, but in other respects were similar to subscribing schools that were not selected. The specific image bundles and prices offered to teachers were common across teachers at a school, but varied across schools. There were two reasons for this. First, given the design of the Homework Hero web site, it was technically easy to customize offers to schools, but

DOI: 10.1002/mde

would have been difficult to vary offers to individual teachers.[16] Second, we considered it undesirable to have teachers know that prices and options were being varied across buyers, and this would have been more apparent if offers were varied across teachers within a school.[17]

At each school, the sale of images took place over an eight-day period, starting at 9:00 pm on a Sunday and ending at 9:00 pm on a Monday.[18] The school experiments took place either in the fall of 2004 or the fall of 2006, with experiments running at several schools in every week in a time period extending from late August through October.[19] At each school, teachers could choose to purchase either of two image bundles, or neither. The larger bundle always included all of the images in the smaller bundle, plus additional ones. The larger bundle was also always offered at a higher price than the smaller, however, per-image prices varied across versions and schools. Bundle sizes ranged from 1 to 12 images, and prices ranged from $0 to $10 across experiments; Table 1 provides additional summary data on the bundle sizes and prices offered, the number of teachers receiving offers, and the numbers buying high- and low-quality versions.[20]

Apart from the specific offers, the mechanics of the experiments were the same across schools. Normally, when a teacher updates assignments on Homework Hero, he or she sees a screen that confirms a successful update and offers several options (view the updated page, update again, or follow links to other pages). When experiments were in progress, this screen also displayed an ad describing the image sales offer. The ad showed an example of a personalized graphic offered for sale and briefly described the sales options available (including numbers of images and prices for the offered collections). The ad also included hyperlinks to a more elaborate sales information page where teachers could browse through the images offered for sale and learn details of the product offer. Both the initial ad and the more detailed offer page made it clear that the offer would expire on a specified date. The detailed offer page indicated that group orders could not be accepted; only individual teachers were permitted to buy. The ad was altered to display a different sample image twice during each experiment. On the last day of the experiment, a bold red message emphasized that the sale would end at 9:00 pm that night. Throughout the experiment the sales offer was presented as ordinary commercial activity, without any indication of an experimental context.[21] Teachers who purchased images retained the rights to employ those images

Table 1. Bundle and Price Configurations

S_1	S_2	P_1	P_2	Number of teachers	Number buying #1	Number buying #2
1	5	0.00	3.00	20	0	0
1	5	1.00	5.00	21	0	0
1	6	2.00	5.00	21	0	0
1	7	1.00	5.00	22	0	0
1	8	0.00	4.00	26	2	1
1	9	0.00	7.00	20	1	0
1	10	1.00	3.00	31	0	0
1	11	1.00	4.00	34	0	0
1	12	1.00	6.00	19	0	0
2	5	1.00	5.00	21	0	0
2	6	0.00	1.00	59	0	1
2	7	0.00	5.00	26	0	0
2	7	1.00	3.00	19	0	1
2	8	2.00	4.00	59	0	0
2	9	1.00	3.00	155	0	2
2	10	0.00	3.00	60	3	1
2	10	1.00	8.00	34	0	0
2	11	1.00	3.00	77	0	0
2	12	1.00	4.00	40	1	0
2	12	3.00	10.00	42	0	0
3	6	2.00	4.00	25	0	0
3	6	3.00	6.00	26	0	0
3	7	2.00	4.00	30	0	0
3	8	1.00	2.00	25	0	0
3	9	0.00	4.00	19	0	0
3	10	0.00	1.00	20	2	0
3	10	0.00	2.00	21	2	1
3	10	0.00	3.00	20	0	0
3	11	1.00	2.00	20	1	0
3	11	2.00	4.00	57	0	3
3	12	0.00	3.00	27	3	1
3	12	2.00	8.00	13	0	0
4	7	3.00	5.00	52	0	0
4	8	1.00	4.00	46	2	0
4	8	3.00	5.00	21	0	0
4	9	2.00	5.00	86	1	0
4	10	0.00	2.00	44	1	0
4	10	0.00	1.00	20	1	2
4	11	1.00	2.00	99	1	0
4	11	2.00	4.00	34	0	0
4	12	1.00	3.00	57	0	0
4	12	4.00	8.00	10	0	0
5	8	4.00	5.00	16	0	0
5	8	5.00	8.00	35	0	0
5	9	1.00	2.00	24	0	1
5	10	2.00	3.00	58	0	0
5	10	5.00	9.00	22	0	0
5	11	3.00	6.00	62	0	1
5	12	2.00	4.00	21	0	1
5	12	2.00	5.00	18	0	2
5	12	3.00	5.00	197	1	2
6	9	3.00	4.00	32	0	0
6	10	1.00	4.00	34	0	0
6	11	3.00	6.00	35	0	0
6	12	1.00	5.00	42	1	1
7	10	1.00	2.00	31	0	0
7	12	0.00	3.00	38	1	3
7	12	1.00	3.00	28	0	0

DOI: 10.1002/mde

on Homework Hero as long as their schools continued to subscribe to the service.

When a teacher decided to buy, the order was placed by submitting a form on the web. Teachers could either pay immediately via credit card (using the PayPal transaction service) or pay later after we sent a bill in the mail.[22] All teachers who ordered images eventually paid for them. Teachers who chose to purchase the smaller of the two available bundles were told that they could upgrade to the larger version at anytime within the experimental period; however, only one buyer chose to do so.

Once an order was completed, the authors produced the image bundles by editing the displayed name in previously designed image templates. The images were stored on the Homework Hero server and purchasers were notified of the URLs (web addresses) for the images. By pasting a URL into the appropriate entry box on the usual Homework Hero assignment update form, a teacher could display an image on his/her assignment page. Later, by changing a number in the URL pasted in the form, a teacher could display an alternative image from the collection.

Across all participating schools, a total of 2283 teacher assignment pages were updated in the course of the experiments.[23] We could identify the owner of each of those pages and, therefore, could determine the identities of all teachers who viewed the initial ad describing the sales offer.[24] Of these 2283 teachers, ultimately 48 teachers purchased image collections, including 16 who 'purchased' bundles that were priced at zero. Total revenue from the experiment's sales was $93.00.

Obviously, purchases were infrequent, but the low sales frequency should not be surprising. Teachers do not come to Homework Hero intending to make purchases i.e., they are not shoppers, *a priori*. The offer they see is in the form of a banner ad, augmented with a brief textual message. Click-through rates on banner ads displayed on the web are known to be low—a 2007 report indicated that the average click-through rate across the web was only 0.20%.[25] Sales rates are necessarily lower than click-through rates, so with sales at about 2% of the audience, our results outperform the Internet average for banner advertising.[26] This also is not surprising—the product was targeted to an audience that should have been receptive (since all ad recipients were already Homework Hero users).

Table 2. Conditional Logit Estimates

Variable	Coefficient
Constant	−4.4566
	(0.2478)
S	0.1795
	(0.0557)
P	−0.5515
	(0.1413)
Log L	−254.87

Standard errors in parentheses.

Nevertheless, the paucity of sales limits the scope of the empirical investigation that can be undertaken.

Using the data obtained from experiments, our purpose is to devise a profit-maximizing versioning and pricing plan. It is a characteristic of versioning schemes that prices *not* vary across identifiable groups or individuals. Instead, all individuals are presented with identical options and they sort themselves through their purchase decisions. Given our intentions, we purposely refrain from using individual- and group-specific data in the estimation of the demand system.

In the following section, we present econometric results that employ the methods described in Section 3. However, as a preliminary empirical exercise, we have estimated a simple conditional logit model using our data. This model specifies that customer choices depend upon the qualities and prices of the offered bundles. The results in Table 2 reveal that both price and quality (number of images) are related to customer choices in an expected fashion—a lower price and a higher quality level make a bundle more attractive and, therefore, increase the probability that it will be purchased. Coefficients for both price and quality differ significantly from zero in the appropriate directions. While the logit specification is inappropriate for our purposes, it is reassuring to see that price and quality exhibit expected impacts on customer choices when this well-known empirical model is estimated.[27]

5. ECONOMETRIC RESULTS

Using the method described in Section 3, we will estimate parameters characterizing the distribution of individual teachers' utility functions. We continue to assume that each teacher has a utility function of the form $U_i = V_i S^{b_i}$, where V_i and b_i are random

DOI: 10.1002/mde

variables drawn from a distribution $G(V_i, b_i, \boldsymbol{\theta})$, and S is the quality (size) of the image collection. Although it would be possible to employ any of a wide range of distributions for G, we begin by using a specific simple distribution. We assume that V_i is uniformly distributed from 0 to $V_{\max}$ ($V_{\max} > 0$) but with a mass of probability at $V_i = 0$. The probability that $V_i = 0$ is indicated by parameter M_0. We also assume that b_i is uniformly distributed from b_{lo} to b_{hi} (where $0 \leqslant b_{\text{lo}} < b_{\text{hi}} \leqslant 1$) and that the distributions for V_i and b_i are independent. Our problem is to find MSL estimates of the parameters $V_{\max}$, M_0, b_{lo}, and b_{hi}.[28] To insure accuracy of our estimates and smoothness of the simulated likelihood function, we set $N_s = 1{,}000{,}000$ (in evaluating the likelihood function, N_s is the number of simulated observations used in calculating a probability for each observation in the original data). We employed the BHHH method of Berndt *et al.* (1974) as implemented in TSP's MLPROC procedure to perform the optimization. We have also replicated our results with alternative optimization routines.[29]

Two sets of model estimates are presented in Table 3. The estimations differ only in terms of restrictions imposed on the coefficients. Model 1 imposes no restrictions on coefficient values but yields estimates of b_{lo} and b_{hi} that are outside of theoretically specified limits (the estimated value for b_{lo} is less than zero and that for b_{hi} is greater than one). Although the point estimates are outside the bounds suggested by theory, neither differs significantly from its theoretically prescribed limit.[30] Model 2 imposes the theoretical restrictions, setting $b_{\text{lo}} = 0$ and $b_{\text{hi}} = 1$. In principle, it would be possible to consider a further restriction of the model requiring that $b_{\text{lo}} = b_{\text{hi}}$, implying that b_i is nonrandom (and, as we have noted, that no versioning scheme is possible). However, imposition of this restriction leads our model to imply that some observed choice patterns should occur with probability zero, and we therefore rule out this case.[31]

Although the uniform distributions assumed for V_i and b_i are plausible, it would be desirable to adopt a more general specification. The Beta distribution is an attractive option for this purpose for two reasons. First, it is flexible. Depending on parameter values, the Beta probability density function can have a variety of shapes. It can be unimodal or bimodal, and it may be symmetric or nonsymmetric. Second, the uniform is a special case of the Beta distribution, and so the hypothesis that a distribution is uniform can be tested as a restriction in a more general model.

Uniform and Beta distributions share parameters that determine upper and lower bounds; however, the Beta distribution has two additional 'shape' parameters, denoted α and β. Permitting both V_i and b_i to have Beta rather than uniform distributions therefore requires that four additional parameters be estimated. When we attempted to estimate this more general specification, our maximization algorithms failed to converge. This is probably a result of the paucity of observed purchases in the data—our data simply do not permit us to distinguish well among the refined alternatives offered by the more general specification. However, we were able to search a grid of Beta distribution shape parameters, letting each of the shape parameters vary between 0.25 and 2.0 at intervals of 0.25. Table 4 reports estimates of the model for the best values

Table 3. Model Estimates

Parameter	Model 1	Model 2
M_0	0.9537	0.9612
	(0.0069)	(0.0068)
$V_{\max}$	1.5231	1.6832
	(0.5002)	(0.3221)
b_{lo}	−0.9024	0.0000[a]
	(1.0203)	
b_{hi}	1.4036	1.0000[a]
	(0.2411)	
Log L	−251.532	−254.295

Standard errors in parentheses.
[a]Coefficient constrained to equal the indicated value.

Table 4. Model Estimates (Beta Distributions)

Parameter	Coefficient
M_0	0.9506
	(0.0071)
$V_{\max}$	3.8546
	(1.1985)
b_{lo}	0.00[a]
b_{hi}	1.00[a]
Shape parameters for V_i	
α	0.50[b]
β	2.00[b]
Shape parameters for b_i	
α	0.75[b]
β	0.50[b]
Log L	−249.875

Standard errors in parentheses.
[a]Coefficient constrained to equal the indicated value.
[b]Coefficient obtained via grid search.

DOI: 10.1002/mde

of the shape parameters. On the basis of these results, a likelihood ratio test fails to reject the restrictions imposed by the uniform distribution for V_i and b_i; consequently, we maintain the assumption of uniform distributions in the remainder of our work.[32] We now focus on the parameter estimates for Model 2 in Table 3, where uniform distributions are assumed and the restrictions $b_{\mathrm{lo}} = 0$ and $b_{\mathrm{hi}} = 1.0$ are imposed.[33] In this estimation, the mass point coefficient, M_0, has a value of 0.963, implying that 96.3% of the teachers had zero valuations for image bundles. Given that only about 2% of the sample actually purchased a bundle, this seems to be a reasonable estimate.[34] A buyer with median values of V_i and b_i (for those with V_i positive) would have a total valuation of \$2.91 for a 12-image bundle. Given that all bundle sales took place at prices between \$0 and \$6, this result also appears to be reasonable.

6. PROFIT-MAXIMIZING VERSIONING

We next consider the problem of determining a profit-maximizing versioning and pricing plan, given the demand system estimates for Model 2. To maintain comparability with the experiments, we continue to consider plans that offer just two versions of the good. In the optimal plan the large bundle necessarily includes all available images, given that we continue to assume that the marginal production cost is zero. As in the experiments, we assume that the size of the largest feasible bundle is exogenously set at 12 images.[35] We wish to determine the optimal size of the smaller bundle, S_1, and prices for both small and large bundles, P_1 and P_2.

To calculate the profitability of any plan, we again employ simulation methods. For given S_1, P_1, and P_2, we draw a large number (10,000) of customers from the estimated distribution of utility functions. For each simulated customer, we determine a purchase choice (no purchase, purchase the small bundle, purchase the large bundle) and the profit generated by that purchase choice. We then average profits across all 10,000 simulated customers and repeat this procedure for different values of S_1, P_1, and P_2. After examining all feasible plans, we determine which one yields the highest profit per customer.[36] In this exercise, we simulate only customers with positive V_i values (those with zero values for V_i would not purchase at any positive price and are irrelevant for the profit maximization calculation).

Table 5. Profit and Welfare Consequences

	Optimal versioning	Single price monopoly	Welfare maximum
N_1	2	NA	NA
P_1	\$1.24	NA	NA
N_2	12	12	12
P_2	\$4.99	\$4.30	\$0.00
% Purchasing Version 1[a]	24.9%	NA	NA
% Purchasing Version 2[a]	26.1%	33.0%	100.0%
Profit per customer[a]	\$1.48	\$1.33	\$0.00
Consumer surplus per customer[a]	\$1.12	\$1.26	\$3.99
Total surplus per customer[a]	\$2.60	\$2.59	\$3.74

[a]Averaged only over customers with nonzero valuations.

Results of this exercise, summarized in Table 5, show that the optimal versioning plan has $S_1 = 2$, $P_1 = \$1.24$ and $P_2 = \$4.99$, and yields profits per customer (for customers with non-zero valuations) of \$1.48. To compare these outcomes to those that would prevail in the absence of versioning, we repeated the simulation described above, but assumed that only the large bundle was sold at a single 'monopoly' price. Results in Table 5 show that the profit-maximizing single-version price of \$4.07 produces a profit per customer of \$1.33. Consequently, optimal versioning increases profit per customer by 11.5% over single-version monopoly pricing. In our experiment, actual revenues obtained from 2283 teachers who updated pages totaled \$93.00. Our results imply that if we had instead set an optimal price for a single version (the 12-image bundle), our expected revenue would have been \$115.38. Had we imposed the optimal two-version package and pricing scheme, revenue would have risen further to \$128.65.[37] These are small amounts, but similar percentage changes would imply huge benefits for large firms.

Our results also have implications regarding the welfare consequences of versioning. Table 5 reports that total surplus per customer shows a negligible increase from \$2.59 under single-product monopoly to \$2.60 under versioning. Profits rise, but consumer surplus falls by almost the same amount. As Table 5 shows, there is a large increase in the number of customers served under versioning. With versioned products, 48.4% of customers with non-zero valuations make a

DOI: 10.1002/mde

purchase, while only 32.7% buy under the single-product monopoly regime. However, many who buy under the versioning scenario receive a product of less than optimal quality.

Results from versioning and nonversioning monopoly scenarios can also be compared to a welfare-maximizing outcome. Assuming marginal cost of zero, the welfare maximum would provide each customer the maximum sized bundle of 12 images, resulting in a total surplus of \$3.74 per customer, a welfare gain of 43.8% over optimal versioning. With or without versioning, the welfare cost associated with monopoly pricing is apparently substantial. Of course, given the assumed cost structure for information goods, a competitive market structure is not feasible, and the first-best optimum does not provide a practical alternative.

7. CONCLUSIONS

Information goods are commonly sold in quality-differentiated versions. With appropriately designed versions and prices, this practice can generate profits as customers sort themselves by willingness to pay when choosing which versions to purchase. However, to exploit profitable versioning opportunities, producers must have knowledge of the distribution of preferences across the population of potential buyers.

We have proposed and applied an econometric method for estimating parameters describing distributions of customer preferences for an information good. To implement the method, we collected data on customer behavior in web-based field experiments in which we varied qualities and prices of digital image collections. With the resulting demand system estimates, we constructed profit-maximizing versioning and pricing schemes. Profit-maximizing versioning increased expected profits to the seller by 11.5% relative to a comparable single-version monopoly plan. Total welfare was essentially unchanged under versioning; under versioning more customers were served, but a large fraction of buyers got a lower quality product. Although welfare impacts associated with versioning by a monopolist were small, the exercise of monopoly power (with or without versioning) was associated with large losses relative to a social optimum. The extent to which such findings might generalize to other goods is an issue of interest for future research.

Versioning is a common practice, and firms who engage in versioning often have opportunities to experiment with prices in order to learn about demand. Information services and software products provide many examples of versioning, but the practice is common in other product markets as well. For example, wireless phone and cable television services come in alternative versions (i.e. service plans). Fees for plans can be varied over locations or over time and detailed information on usage patterns can be obtained from customers. The problem of how to use this information to devise service options and set prices is one that might be analyzed using methods similar to those described in this article.

APPENDIX A: SUFFICIENT CONDITIONS FOR PROFITABLE VERSIONING

We consider the choice problem facing a potential buyer, as described in Section 3. To simplify, we drop the subscript i that refers to an individual in the text. The buyer has a utility function, $U = VS^b$, which associates a gross valuation, U (measured in 'dollars worth' units) with a versioned information good that has quality level S. This functional form for the utility function is employed in our empirical work; however, the theoretical results presented in this section generalize to utility functions of the form $U = Vf(S)$, where $f'(S) > 0$. We assume that the utility function parameters satisfy the conditions $V \geqslant 0$ and $0 \leqslant b \leqslant 1$. These assumptions assure that the marginal utility of additional quality is nonnegative and nonincreasing. The latter condition ($0 \leqslant b \leqslant 1$) is not necessary to establish the results presented in this section, but it does rule out other implausible implications (cf. note 6). Versioning will be profitable only when customers are not identical. We initially assume that customers differ only in terms of the parameter V, which is randomly distributed across customers according to a twice differentiable cumulative distribution function, $G(V)$. The number of consumers is normalized to equal one and production costs are again assumed to equal zero.

Consider first the case where the seller offers a single version of the good for sale. This version has

DOI: 10.1002/mde

quality S and is sold at a price P selected by the seller. For given S and P, an individual will buy if $VS^b > P$, or, equivalently, $V > P/S^b$. The fraction of customers satisfying the latter condition is given by $1-G(P/S^b)$, so the seller's profit is

$$\Pi = P[1 - G(P/S^b)]$$

The first order condition for this problem requires that the profit-maximizing price, P^*, be chosen to satisfy

$$\frac{\partial \Pi}{\partial P} = \left[1 - G\left(\frac{P^*}{S^b}\right)\right] - \frac{P^*}{S^b} g\left(\frac{P^*}{S^b}\right) = 0$$

or

$$h(T^*) = \frac{1}{T^*}$$

where $T^* = P^*/S^b$, and h and g denote the hazard and density functions associated with the cumulative distribution G, respectively. The second-order conditions are satisfied provided that $2g(T^*)+T^*g'(T^*)>0$.

We next consider the case where the monopolist offers two different versions of the good. Let S_1 be the quality of version 1 (the lower quality version), let S_2 be the quality of version 2 (the higher quality version), and let P_1 and P_2 be prices for those two versions. Each buyer maximizes utility by choosing to purchase either of two versions of the good, or neither. His/her choice may be summarized by the following set of conditions:

If $VS_1^b - P_1 \leqslant 0$ and $VS_2^b - P_2 \leqslant 0$,

do not purchase (1a)

If $VS_1^b - P_1 > 0$ and $VS_1^b - P_1 \geqslant VS_2^b - P_2$,

purchase version 1 (1b)

If $VS_2^b - P_2 > 0$ and $VS_2^b - P_2 > VS_1^b - P_1$,

purchase version 2 (1c)

These conditions imply that a customer chooses the product version that yields greatest valuation net of price or chooses not to buy if neither version offers a positive net valuation. Rewriting the above conditions in a more compact form, we have

If $V \leqslant T_1$ and $V \leqslant T_2$, do not purchase

If $V > T_1$ and $V \geqslant T_2$, purchase version 1

If $V > T_2$ and $V > T_{12}$, purchase version 2

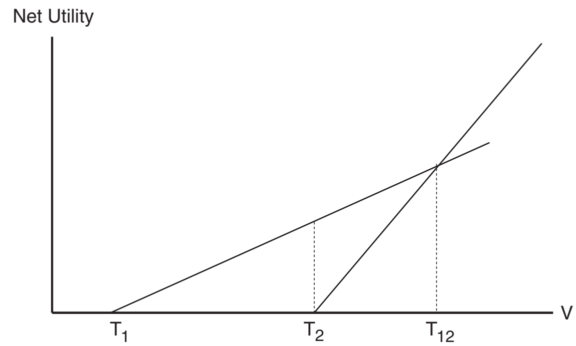

Figure A1. Net utility for two versions of a good.

where $T_1 = P_1/S_1^b$, $T_2 = P_2/S_2^b$, and $T_{12} = (P_2 - P_1)/(S_2^b - S_1^b)$.

If versioning is to be a profitable strategy for the seller, then prices and versions must be selected so that there are positive sales of both versions. This requires that $T_{12} > T_2 > T_1$ or, in terms of prices and quality, that $P_2 S_1^b > P_1 S_2^b$. The rationale for this requirement can be illustrated with a graphical argument. In Figure A1 we show the (positive) net utility gained by a buyer as a function of the parameter V. The steepest line (with slope S_2^b) represents the net valuation obtained when the buyer consumes the higher quality version of the good, while the more gently sloped line (with slope S_1^b) represents the net valuation obtained from the lower quality good. The point where the lines intersect identifies a buyer whose value of V leaves her indifferent to versions 1 and 2. In the diagram, we illustrate the case where some buyers select each version, which clearly requires that $T_{12} > T_2 > T_1$. Given this condition, the profit function for a versioning monopolist is

$$\Pi = P_2[1 - G(T_{12})] + P_1[G(T_{12}) - G(T_1)]$$

and first-order conditions are

$$\frac{\partial \Pi}{\partial P_1} = \frac{P_2}{S_2^b - S_1^b} g(T_{12}) + [G(T_{12}) - G(T_1)] - \frac{P_1}{S_2^b - S_1^b} g(T_{12}) - \frac{P_1}{S_1^b} g(T_1) = 0$$

$$\frac{\partial \Pi}{\partial P_2} = [1-G(T_{12})] - \frac{P_2}{S_2^b - S_1^b} g(T_{12}) + \frac{P_1}{S_2^b - S_1^b} g(T_{12}) = 0$$

DOI: 10.1002/mde

or after simplifying

$$\frac{\partial \Pi}{\partial P_1} = T_{12}g(T_{12}) + G(T_{12}) - G(T_1) - T_1 g(T_1) = 0 \tag{2a}$$

$$\frac{\partial \Pi}{\partial P_2} = 1 - G(T_{12}) - T_{12}g(T_{12}) = 0 \tag{2b}$$

Assuming an optimum exists, and letting asterisks denote values at the optimum, the latter condition simplifies to

$$h(T_{12}^*) = \frac{1}{T_{12}^*} \tag{3b}$$

Using condition (2b) we know that $1 = G(T_{12}^*) + T_{12}^* g(T_{12}^*)$. Substituting this expression into Equation (2a) and rearranging we obtain

$$h(T_1^*) = \frac{1}{T_1^*} \tag{3a}$$

Note that conditions (3a) and (3b) require that the hazard function associated with the distribution G intersect the curve $1/x$ at two distinct points. In addition, recall that the solution we obtained for the problem of a nonversioning seller also occurred at a point of intersection between h and $1/x$. That point, T^*, is distinct from T_1 and T_{12}. Therefore, for the case under consideration, a profitable versioning plan can exist only if the hazard function associated with G intersects the curve $1/x$ in at least three different locations.

Well-known probability distributions fail to satisfy this requirement. These include all distributions with non-decreasing hazard rates (e.g. uniform, normal, logistic, extreme-value, chi-squared, Laplace, exponential distributions, and the Gamma and Beta distributions for a range of values of the parameters) as well as some distributions with nonincreasing hazard rates (e.g. the Pareto distribution and the Gamma distribution for a range of values of the parameters). It is possible for hazard functions for other distributions, including the lognormal and Beta distributions, to intersect the function $1/x$ in three distinct points, but only for specific parameter values. We conclude that for utility functions of the form $U = VS^b$, if V is random but b is not, then two-product versioning schemes will be profitable only when very special restrictive assumptions are made about the distribution of V.

Anticipating our subsequent empirical work, we can state this result a bit differently. For utility functions of the form $U = VS^b$, if V is uniformly distributed and b is nonrandom, then no profitable two-product versioning plan exists. However, we can show by example that if b is random, then profitable versioning *is* possible. Consequently, a sufficient condition for profitable versioning is that b be random. This has an implication for empirical work—if we wish to investigate profitable two-product versioning schemes, model specifications should be sufficiently general to permit randomness in both b and V.

APPENDIX B: APPLYING THE MSL METHOD

This discussion of estimation builds on the earlier discussion provided in Section 3. Consider customer i, who accounts for a single observation in a sample of size N. In addition, consider arbitrarily selected values for $\boldsymbol{\theta}$, denoted $\boldsymbol{\theta}_0$. From the distribution $G(V_i,b_i,\boldsymbol{\theta}_0)$, draw a random sample of simulated observations, $j = 1, \ldots, N_s$, for V_j and b_j. For each V_j and b_j drawn for customer i, determine what his/her choice would be when facing prices and versions P_{1i}, P_{2i}, S_{1i} and S_{2i}. This is done by applying conditions (1) in the text for each observation j. This step yields a series of simulated choice outcomes, C_{0j}, C_{1j}, and C_{2j}, for $j = 1, \ldots, N_s$. Given these series, we can approximate the probability that customer i will make choices in each of the three categories with the frequencies observed in the simulated sample. That is,

$$R_{0i} \approx \frac{\sum_{j=1}^{N_s} C_{0j}}{N_s}$$

$$R_{1i} \approx \frac{\sum_{j=1}^{N_s} C_{1j}}{N_s}$$

and

$$R_{2i} \approx \frac{\sum_{j=1}^{N_s} C_{2j}}{N_s}$$

For sufficiently large values of N_s, these approximations approach any desired level of accuracy. Using these values for R_{0i}, R_{1i}, and R_{2i}, one can calculate the likelihood for

DOI: 10.1002/mde

observation i given parameters $\boldsymbol{\theta}_0$ using Equation (5).

Now repeat the procedure described in the preceding paragraph for each observation in the original sample, that is for $i = 1, \ldots, N$. Doing so yields a value for the likelihood for each observation, and, using Equation (6), a value of the likelihood function for the sample. We are, therefore, able to evaluate the likelihood function for arbitrary parameter values. The only remaining problem is to find parameter values that maximize the likelihood function, but conventional methods can be employed for this purpose. As we have noted in the text (cf. footnote 27), we have modified the method described here slightly to take advantage of special features of our data and our empirical model.

NOTES

1. The problem of optimal price experimentation by a monopolist facing uncertain demand has been explored by Rothschild (1974) and Aghion *et al.* (1991). Loginova and Taylor (2008) investigate incentives for monopoly experimentation when sellers can charge personalized prices. They find that learning by the monopolist is often thwarted by strategic behavior of buyers. Acquisti and Varian (2005) report similar results and characterize circumstances in which conditioning prices on past purchase behavior can be profitable. Esteves (forthcoming) finds that under duopoly, firms may eschew learning about the loyalties of individual buyers because subsequent price discrimination can lead to a more aggressively competitive outcome.
2. Our analysis is related to other studies that have examined price discrimination for quality-varying goods. See, for example, Crawford and Shum (2006, 2007) for applications to cable television programming packages, Leslie (2004) on Broadway theatre ticket sales, McManus (2007) on retail markets for specialty coffee, Miravete (2002) on local telephone services, and Verboven (2002) on gasoline and diesel automobiles. Our approach is distinct in its use of web-based experimentation to collect data.
3. Deneckere and McAfee (1996) provide a seminal analysis of the versioning phenomenon in their analysis of 'damaged goods.' Varian and Shapiro (1999) provide a nontechnical discussion of versioning and a number of examples. Bakos and Brynjolfsson (2000) analyze 'bundling' or 'aggregation' of information goods. Versioning can be thought of as a special instance of bundling in which lower quality versions (bundles) can be formed by subtracting components from versions (bundles) of higher quality.
4. Harrison and List (2004) provide a survey of field experiments in economics and provide a comparative discussion of field and laboratory experiments.
5. See Belleflamme (2005) for an exposition of the versioning problem using an alternative utility function. The Belleflame utility function specification is a specialized and restrictive one; it not only assumes that all customers have constant marginal utility of quality but also assumes that a discrete increment to utility is gained if positive, rather than zero, output is consumed.
6. A negative value for b_i implies that the marginal utility of quality is negative, while a value of b_i greater than one implies that the marginal utility of quality is rising. In the latter case, if quality could be added to a good at a constant cost, then the profits from selling to a single customer would be unbounded.
7. Jing (2000, 2007) reports a similar result for the case of a linear utility function. Jing also shows that when b is nonrandom, versioning may be profitable if network externalities exist.
8. For the form of the utility function and distributional assumptions adopted for most of our empirical analysis, it is possible to derive an analytical form for the likelihood function. However, even for this case, the likelihood function is very complex.
9. Discussions of maximum simulated likelihood estimation are also provided by Arias and Cox (1999), Gourieroux and Monfort (1993), Lee (1995), Lerman and Manski (1981), and Stern (1997). As Train (2003, p. 242) notes, the approximation of probabilities in the MSL procedure introduces a bias in estimation; however, the bias diminishes as the number of draws used in simulation increases. Arias and Cox (1999) further note that when probabilities are approximated by frequencies, the simulated probabilities will not be continuous functions of the underlying parameters, and standard optimization algorithms for maximum likelihood estimation may fail. The use of large numbers of draws in simulation increases the smoothness of the simulated likelihood function and improves the performance of numerical optimization methods.
10. In fact, Amazon has carried out pricing experiments. In the summer of 2000, DVD prices were varied randomly to visitors in an experiment intended to provide information about customer demand. See the CNET News Website, 'Now Showing: Random DVD Prices on Amazon,' available at http://news.cnet.com/2100-1017-245326.html
11. Normally, when a firm undertakes a pricing experiment, it must weigh the value of information gained against profits lost while offering suboptimal experimental prices. In our case, this was not an issue. The product offered in the experiment was available only in the experimental setting, not in the

DOI: 10.1002/mde

normal course of business. All revenues were donated to the Economics Department at the University of South Carolina.

12. In September 2004, about 200 schools subscribed to the service, and about 5,000 teacher pages existed. The site received about 100,000 unique visits and about 1.2 million 'total hits' per week.
13. Examples of the images sold are displayed at: http://professorchappell.com/Data/MDE/index.htm.
14. Because we know precisely which images were offered in each bundle, in principle we could investigate whether specific images were especially highly valued. Because overall sales rates were low, we have not attempted to refine the empirical models in this manner.
15. Some schools subscribe to a 'noncommercial' version of the service that normally excludes all advertising from assignment pages. Because our experiments displayed ads viewed by teachers, site managers at these schools were notified in advance that ads would be displayed during a brief experiment undertaken for academic purposes. However, teachers viewing the ads would typically not have known that the sales offer was associated with a market experiment.
16. In addition, Homework Hero's privacy policy rules out the use of cookies that would identify an individual for purposes of offering an individual-specific price.
17. Knowledge that prices varied could have caused some confusion and/or resentment among customers. Since Homework Hero is a 'real' business, we wanted to avoid this possibility. In addition, the experiment should attempt to create conditions similar to those that would prevail in a post-experimental selling stage in a real business. Once an optimal versioning and pricing scheme is established by a firm, all customers would be presented with identical options.
18. Experiments involving two schools were intentionally extended by a day because the normal end date fell on the Labor Day holiday. One experiment was inadvertently extended by 2 days because of an oversight by the authors.
19. We originally planned experiments only for 2004. However, by 2006, more schools were using the service and we decided to run additional experiments. Some new images were included in bundles offered for sale in the 2006 round of experiments. Ex post, statistical tests support the decision to pool data generated in 2004 and 2006. A likelihood ratio test fails to reject the hypothesis of equal coefficients over the two samples for our preferred specification (Model 2 in Table 3): $\chi^2 = 3.66 < 5.99 = \chi^2_{.05}(2)$. Experiments in each year were staggered over a period of several months because the work associated with selling and producing the products made it difficult to manage large numbers of experiments simultaneously. The staggering of school opening dates (from early August to mid-September) also led us to vary start dates for the experiments. We conducted the experiments early in the school year because teacher usage is heavier at this time.
20. In the 2004 experiments, no bundle sold for more than \$5, suggesting that an upper limit of \$10 for the price range in experiments was reasonable. The selected bundle sizes cover most of the possible unique combinations that are possible. The number of distinct offers (58) is less than the number of schools (70) because schools were sometimes combined in identical treatments.
21. A possible exception has been noted (cf. note 15). However, most teachers would not have known about the experimental nature of the offer.
22. Because teachers often update pages in short periods of time between class periods, we felt that it would be helpful to keep the purchase process as simple as possible. Since credit card forms are sometimes cumbersome, we offered the 'pay later' option.
23. Pages for clubs, sports teams, other organizations, or pages maintained by groups of more than two teachers were excluded from the sample. None of the individuals maintaining these pages ordered image bundles during the experiment.
24. Unfortunately, we were not able to unambiguously determine which teachers clicked to proceed to the detailed sales offer page, except for those who eventually made a purchase.
25. See 'So Many Ads, So Few Clicks,' *BusinessWeek* Website, November 12, 2007. Available at http://www.businessweek.com/magazine/content/07_46/b4058053.htm.
26. On the basis of data from our 2004 experiments, we infer that roughly 10% of the target population clicked-through the initial ad at some point (based on the number of recorded hits on the ad banner and the total number of teacher pages updated).
27. The most important limitation of the logit model is that it precludes profitable versioning (we have found no profitable versioning plans and conjecture that this is a general result). In addition, the random utility specification underlying the logit model contains a single additive random error term that varies across choices as well as individuals. This assumption permits some peculiar outcomes in the context of our application. For example, a buyer might choose a smaller bundle when offered a larger bundle that includes the smaller as a subset, even when the larger bundle is offered at a lower price. The latter limitation is also a property of mixed logit models that are sometimes employed in the estimation of random utility models. See Train (2003) on the mixed logit model.
28. The estimation procedure implemented for our application takes advantage of special features of our model and data. First, when there is a mass point of probability for $V_i = 0$, it is only necessary to simulate choices over the range of positive values for V_i. Second, because all price and version options were identical across teachers in a school, and because we do not include any individual-specific variables, there are a small number of possible values for the likelihood function for any

DOI: 10.1002/mde

observation. It is not necessary to simulate for each observation separately, since many observations within a given school are identical from an econometric perspective.

29. We have replicated the results using similar routines in Stata and using the downhill simplex (Amoeba) method of Nelder and Mead (1965), in FORTRAN. The latter method is especially useful when likelihood functions are not smooth; it also makes use of global properties of the function and does not require local gradient information.
30. A likelihood ratio test fails to reject the joint hypothesis that $b_{\mathrm{lo}} = 0$ and $b_{\mathrm{hi}} = 1.0$; specifically, $\chi^2 = 1.676 < 5.99 = \chi^2_{.05}(2)$.
31. For some price-bundle configurations, the model implies that one bundle should dominate the other for all customers. However, in the data, some customers choose each bundle.
32. Treating the best shape parameters obtained in the grid search as those maximizing log L, the test results are: $\chi^2 = 8.84 < 9.49 = \chi^2_{.05}(4)$; we fail to reject the hypothesis of uniform distributions.
33. Complete results of the grid search are posted at http://professorchappell.com/Data/MDE/index.htm. This web page also provides a plot of the likelihood function for Model 2 of Table 3, revealing that the estimation did locate a maximum.
34. Many teachers who might have had positive valuations of the good nevertheless failed to notice or investigate the offer and, for purposes of empirical modeling, they are best treated as having zero valuations. This result does reveal the importance attached to getting the attention of potential customers, apart from selecting profit-maximizing versions and prices.
35. It would be possible to perform these calculations assuming a larger maximum bundle size, but extrapolating beyond ranges covered in the experiments introduces additional uncertainty for forecasts.
36. We consider only integer values for S_1.
37. Of 2283 teachers, only 86.5 (3.79%) are estimated to have had positive valuations. Multiplying 86.5 by profits per customer yields these amounts.

REFERENCES

Acquisti A, Varian HR. 2005. Conditioning prices on purchase history. *Marketing Science* **24**: 367–381.

Aghion P, Bolton P, Harris C, Julien B. 1991. Optimal learning by experimentation. *The Review of Economic Studies* **58**: 621–654.

Arias C, Cox T. 1999. Maximum simulated likelihood: a brief introduction for practitioners. Staff Paper 421, University of Wisconsin-Madison, Department of Agricultural and Applied Economics.

Bakos Y, Brynjolfsson E. 2000. Aggregation and Disaggregation of Information goods: implications for bundling, site licensing, and micropayment systems. in *Internet Publishing and Beyond*: *The Economics of Digital Information and Intellectual Property*, Varian HR, Kahin B (eds). MIT Press: Cambridge, MA; 114–137.

Belleflamme P. 2005. Versioning in the information economy: theory and applications. *CESifo Economic Studies* **51**: 329–358.

Berndt EK, Hall BH, Hall RE, Hausman JA. 1974. Estimation and inference in nonlinear structural models. *Annals of Social Measurement* **3**: 653–665.

Business Week Website. 2007. So many ads, so few clicks. 12 November 2007. Available from: http://www.businessweek.com/magazine/content/07_46/b4058053.htm (Accessed on 31 May 2010).

CNET News Website. 2000. Now showing: random DVD prices on Amazon. Available from: http://news.cnet.com/2100-1017-245326.html (Accessed 31 May 2010).

Crawford GS, Shum M. 2006. The welfare effects of endogenous quality choice: the case of cable TV. Working paper, University of Arizona.

Crawford GS, Shum M. 2007. Monopoly quality degradation and regulation in cable television. *Journal of Law and Economics* **50**: 181–219.

Deneckere RJ, McAfee RP. 1996. Damaged goods. *Journal of Economics and Management Strategy* **5**: 149–174.

Esteves RB. Pricing with customer recognition. *International Journal of Industrial Organization*. DOI: 10.1016/j.ijindorg.2010.03.008.

Gourieroux C, Monfort A. 1993. Simulation-based inference: a survey with special reference to panel data models. *Journal of Econometrics* **59**: 5–33.

Harrison G, List J. 2004. Field experiments. *Journal of Economic Literature* **42**: 1009–1055.

Jing B. 2000. Versioning information goods with network externalities. In *Proceedings of the Twenty-First International Conference on Information Systems*, Ang S, Krcmar H, Orlikowski WJ, Weill P, DeGross JI (eds). Association for Information Systems: Atlanta; 1–12.

Jing B. 2007. Network externalities and market segmentation in a monopoly. *Economics Letters* **95**: 7–13.

Lee L. 1995. Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory* **11**: 437–483.

Lerman S, Manski C. 1981. On the use of simulated frequencies to approximate choice probabilities. In *Structural Analysis of Discrete Data with Econometric Applications*, Manski C, McFadden D (eds). MIT Press: Cambridge, MA; 305–319.

Leslie P. 2004. Price discrimination in Broadway theatre. *RAND Journal of Economics* **35**: 520–541.

Loginova O, Taylor CR. 2008. Price experimentation with strategic buyers. *Review of Economic Design* **12**: 165–187.

McManus B. 2007. Nonlinear pricing in an oligopoly market: the case of specialty coffee. *RAND Journal of Economics* **38**: 512–532.

Miravete EJ. 2002. Estimating demand for local telephone service with asymmetric information and optional calling plans. *Review of Economic Studies* **69**: 943–971.

Nelder JA, Mead R. 1965. A simplex method for function minimization. *Computer Journal* **7**: 308–313.

Rothschild M. 1974. A two-armed Bandit theory of market pricing. *Journal of Economic Theory* **9**: 185–202.

Stern S. 1997. Simulation-based estimation. *Journal of Economic Literature* **35**: 2006–2039.

Train K. 2003. *Discrete Choice Methods with Simulation.* Cambridge University Press: Cambridge, New York, NY.

Varian HR. 2000. Versioning information goods. In *Internet Publishing and Beyond*: *The Economics of Digital Information and Intellectual Property*, Varian HR, Kahin B (eds). MIT Press: Cambridge, MA; 190–201.

Varian HR, Shapiro C. 1999. *Information Rules*. Harvard Business School Press: Boston, MA.

Verboven F. 2002. Quality-based price discrimination and tax incidence: evidence from gasoline and diesel cars. *RAND Journal of Economics* **33**: 275–297.

DOI: 10.1002/mde