Personalized Price Discrimination Using Big Data

Benjamin Reed Shiller, Economics Department, Brandeis University

Working Paper Series

2016| 108

Personalized Price Discrimination Using Big Data

Benjamin Reed Shiller*

July 29, 2016

Abstract

Person-specific pricing was rarely observed in the past because reservation prices were unobtainable. I investigate whether this still holds now that detailed individual behaviors are tracked. Individuals' expected demand functions are estimated by combining a classic economic model with machine learning techniques to address overfitting and high dimensionality. I find that tailoring prices based on web browsing histories increases profits by 14.55%, and results in some consumers paying nearly double the price others do for the same product. Using only demographics to personalize prices raises profits by only 0.30%, suggesting the percent profit gain from personalized pricing has increased 48-fold.

True personalized pricing had for a very long time been limited to a very small set of venues, because it requires information on individuals' reservation prices, which, simply put, were unavailable. Now, however, large datasets on individual behavior, popularly referred to as "big data," are available and can be used to form a hedonic estimate of individuals' reservation

*Brandeis University, MS01, 415 South St., Waltham, MA 02453, shiller@brandeis.edu. I would like to thank Kathryn Graddy, Shane Greenstein, Adam Jaffe, Joel Waldfogel, and Phil Leslie for very useful feedback, suggestions, and advice. I would also like to thank Erik Brynjolfsson, Iain Cockburn, Michael Grubb, George Hall, Jeremey Howard, Gary Jefferson, Blake LeBaron, Rachel McCulloch, Julie Mortimer, Robert Pindyck, Scott Stern, Hal Varian, and seminar participants, for helpful suggestions which improved the paper. I would like to thank Steven Dupree and Neil Rothstein for sharing with me some of their extensive knowledge of marketing in practice. Lastly, I would like to thank Steven Knowles, for imparting some knowledge about websites of dubious nature.

prices. This paper develops a method for determining optimal individual-level prices from observational data, and uses the method to estimate, in one context, the profit that would be gained from personalized pricing if nearly 5000 web-browsing variables were used to estimate individuals' reservation prices. Of course, this alone cannot prove a break from the past. So, as a comparison, I compute the analogous profit gained from personalized pricing when only demographics, which have long been available, are used as explanatory variables.

While setting prices that vary across large groups is somewhat trivial on the Internet, since such prices can be fine-tuned using simple A/B testing or by varying prices and estimating demand, these same commonly-used techniques are unable to fulfill the promise of personalized pricing. The reason is that these techniques rely on the premise of comparing sales across subsamples of similar groups of consumers who are offered different prices. Unless able to make different take it or leave it offers to the same irrational/myopic individual, it is not possible to experiment at the individual level.[1]

This paper provides a solution. A probit-like model to estimate how individual-level demand varies across consumers based on their browsing histories, and ex-post we estimate price sensitivities using an optimal pricing condition. Since the findings may depend on the model's ability to extract information from the data, I present an adaptation to the probit model that can be combined with powerful machine learning techniques. This obviates the need to rely on a researcher's judgment to select a smaller set of variables, as was typically done previously (e.g. Ghose et al. [2012]).

Personalized pricing is of current interest not only to businesses, but also to policy makers. In his address to the FTC in January 2015, President Obama announced plans to introduce new legislation, a *Consumer Privacy Bill of Rights*, which institutes a framework for promoting transparent use of data that is limited by consumers' consent.[2] Personalized pricing was

[1] Acquisti and Varian [2005] and Fudenberg and Villas-Boas [2007] show that basing prices on a consumers' willingness to buy previously at higher prices will be gamed by rational forward-looking consumers, eliminating any profit gain from the strategy.

[2] http://www.whitehouse.gov/the-press-office/2015/01/12/remarks-president-federal-trade-commission

identified as a major motivating concern in a White House Report that same month (Office of the President [2015]). This paper shows how dramatic the impact of tracking might be.

Few empirical papers exist to guide firm pricing or policy-makers. While many papers focus on other forms of price discrimination, such as 2^{nd} degree, 3^{rd} degree, and bundling, only two empirical papers, Waldfogel [2014] and Shiller and Waldfogel [2011] consider personalized pricing, and neither considers consumer tracking.[3,4]

The closest prior literature, beginning with Rossi et al. [1995, 1996], considers personalized pricing based on past purchase history of the same product.[5] This old strategy, however, has one major flaw. A rational customer should realize that buying at a high price will reveal their type, implying they will pay a higher price henceforth. Acquisti and Varian [2005] and Fudenberg and Villas-Boas [2007] show that forward-looking consumers will change their behavior under such pricing, eliminating any profit gain.[6]

By contrast, personalized pricing based on many unrelated tracking variables is not subject to the same criticism. First, with bounded rationality consumers may be unable to avoid being charged high prices. I find, for example, that Netflix should charge higher prices to those that visit Amazon.com or Wikipedia.org, patterns consumers may not recognize and for which no simple heuristics exist. Even if consumers knew which behaviors result in low prices, they might prefer to ignore them rather than change potentially thousands of behaviors just to receive a lower quoted price for one product.[7] Moreover, these exact data are available to firms from multiple sources - Economist [2014] notes that over 1300 firms are tracking consumers at the 100 most popular websites.

[3]Recent examples of papers focusing on 2nd and 3rd degree price discrimination include Chu et al. [2011], McManus [2008], Miravete et al. [2014]

[4]Hannak et al. [2014] and Mikians et al. [2012] found some evidence of personalized pricing online, but did not study its effectiveness.

[5]Personalized marketing, including pricing, is referred to in the marketing literature as "customer addressability."

[6]Chevalier and Goolsbee [2009] show consumers are not myopic, but rather forward-looking with high discount factors.

[7]Firms could also charge high prices to any consumers not revealing their web browsing data, providing the incentive for consumers to reveal them.

Personalized pricing may have large impacts. While the direct impact of price discrimination on consumers is theoretically ambiguous (Aguirre et al. [2010], Robinson [1933], Schmalensee [1981]), it may be large, and there are important indirect effects from personalized pricing. By increasing profits of firms with monopoly power, it increases the incentives to innovate, differentiate, and merge.[8] Such pricing may also cause consumers to waste effort masking themselves as low valuation consumers, leading to welfare losses. Lastly, consumers may receive disutility from the invasion of privacy. These impacts are not limited to monopoly industries - Graddy [1995] and Shepard [1991] show another form of PD is used even in a seemingly perfectly competitive market, suggesting personalized pricing can be used as well.

A key question is the effectiveness of personalized pricing based on consumer tracking. But there are several challenges to studying it. First, firms are often discreet, hiding its use and framing it as a *customized coupon* which is automatically applied and requires no effort to use, and thus different from personalized pricing in name alone.[9] Second, simulating its effectiveness requires estimating the extent to which tracking variables, of which there are many thousands, predict willingness to pay. Overfitting is an immediate obvious concern.

These challenges are overcome by estimating demand as follows. First, machine learning techniques, e.g. classification trees or neural networks, can be used to estimate individuals' purchase probabilities at observed, non-personalized prices. Next, assuming the model structure from Bresnahan [1987] and Mussa and Rosen [1978] , I show one can infer individual-level estimates of a cardinal measure of perceived quality, and then use these estimates to form individual-level estimates of *expected* demand for vertically differentiated products. A sample held out from estimation is used to ensure that machine learning techniques avoid problems arising from type 1 errors which could bias the paper towards strong findings. One can then simulate the counterfactual - optimal personalized prices and corresponding profits - and compare to the status quo environment.

[8]Personalized pricing, done jointly by oligopolists, may lower profits [Corts [1998], Spulber [1979], Thisse and Vives [1988], Choudhary et al. [2005]].

[9]Pik [2014]

Netflix provides an auspicious context for study. First, since purchases occur online, Netflix could easily implement personalized pricing based on web data. Second, unlike in most contexts, hypothetical personalized pricing can be empirically studied. Doing so requires individual-level data on both web-browsing histories and *all* purchases of a particular item, data which rarely appear together in data available to academic researchers.[10] However, Netflix subscription can easily be imputed from web-browsing histories.

Simulations reveal that incorporating web-browsing behaviors substantially raises the amount by which person-specific pricing increases profits relative to constant markup pricing. Profits are only 0.30% higher if using demographics alone to personalize prices, but 14.55% higher if using all data - a 48-fold increase. As a comparison, I find second degree PD raises profits by only 8.87%. Moreover, personalized pricing can augment second degree price discrimination; personalized second degree PD raises profits by 25.50%. Thus web browsing data make personalized pricing more appealing to firms.

It is an empirical question as to whether personalized pricing raises or lowers consumer surplus (Aguirre et al. [2010], Robinson [1933], Schmalensee [1981]). We find aggregate consumer surplus is estimated to fall by 1.79%, but joint surplus rises by 0.50%. Moreover, substantial equity concerns arise, since some consumers pay almost double the price that some others do for the same product.

The remainder of the paper is organized as follows. Section 1 describes the context and industry background. Section 2 describes the data. Section 3 presents the model, and Section 4 provides estimation details. The main results of the paper are then presented 5. A brief conclusion follows.

[10]These data are, however, easily accessible by the firm.

1 Background

Netflix, a DVD rentals-by-mail provider, was very popular in the year studied, 2006. Over the course of the year, 11.57 million U.S. households subscribed at some point [Netflix [2006]]. This implies that about 16.7% of internet-connected households consumed Netflix during 2006.[11]

Netflix services appear differentiated from competitors offerings, implying they had some pricing power. Except for Blockbuster's unpopular *Total Access* plan, no other competitor offered DVD rentals by mail.[12] Moreover, Netflix's customer acquisition algorithm was well-regarded, further differentiating their services.

Netflix's subscriptions plans can be broken into two categories. Unlimited plans allow consumers to receive an unlimited number of DVDs by mail each month, but restrict the number of DVDs in a consumer's possession at one time. Limited plans set both a maximum number of DVDs the consumer can possess at one time and the maximum number sent in one month.

In 2006, there were seven plans to choose from. Three plans were limited. Consumers could receive 1 DVD per month for $3.99 monthly, 2 DVDs per month, one at a time, for $5.99, or 4 per month, two at a time, for $11.99. The unlimited plan rates, for $1-4$ DVDs at a time, were priced at $9.99, $14.99, $17.99, and $23.99, respectively.[13] None of the plans allowed video streaming, since Netflix did not launch that service until 2007 [Netflix [2006]].

Key statistics for later analyses are the marginal costs of each plan. The marginal costs for the one to three DVD at-a-time unlimited plans were estimated using industry statistics and

[11] Total number of U.S. households in 2006, according to Census.gov, was 114.384 million (http://www.census.gov/hhes/families/data/households.html). About 60.6% were internet-connected, according to linear interpolation from the respective numbers of connected homes in 2003 and 2007, according to the CPS Computer and Internet Use supplements. $11.57/(0.606 * 114.384) * 100 \approx 16.7$.

[12] Blockbuster's mail rentals were unpopular until they offered in-store exchanges starting in November 2006. Subscriptions increased quickly, reaching 2 million in total by January 2007 [Netflix [2006]].

[13] A very small number of buyers were observed paying $16.99 per month for the 3 DVDs at-a-time unlimited plans. These observations were interspersed over time, suggesting it was not due to a change in the posted price.

expert guidance. They are assumed to equal \$6.28, \$9.43, and \$11.32, respectively.[14]

2 Data

The data for this study were obtained from ComScore, through the WRDS interface. The microdata contain, for a large panel of computer users, demographic variables and the following variables for each website visit: the top level domain name, time visit initiated and duration of visit, number pages viewed on that website, the referring website, and details on any transactions.[15] For further details on this dataset, refer to previous research using this dataset (Huang et al. [2009], Moe and Fader [2004], Montgomery et al. [2004]).

Netflix subscription status can be imputed in these data. For a small sample of computer users observed purchasing Netflix on the tracked computer during 2006, subscription status is known. For the rest, it is assumed that a computer user is a subscriber if and only if they *averaged* more than two sub-page views within Netflix's website per visit. The reasoning behind this rule is that subscribers have reason to visit more subpages within Netflix.com to search for movies, visit their queue, rate movies, etc. Non-subscribers do not, nor can they access as many pages since they cannot sign in. According to this rule, 15.75% of households in the sample subscribe. This figure is within a single percentage point of the estimated share of U.S. internet-connected households subscribing, found in Section 1. This small difference may be attributed to approximation errors in this latter estimate, and ComScore's sampling methods.

Several web behavior variables were derived from the data. These included the percent of

[14] A former Netflix employee recalled that the marginal costs of each plan were roughly proportional to the plan prices, i.e. the marginal cost for plan j approximately equaled $x * P_j$, where x is a constant. I further assume that the marginal cost of a plan is unchanging, and thus equal to the average variable cost. With this assumption, one can find x by dividing total variable costs by revenues. According to Netflix's financial statement, the costs of subscription and fulfillment, a rough approximation to total variable costs, were 62.9 percent of revenues, implying $x = 0.629$. Subscription and fulfillment include costs of postage, packaging, cost of content (DVDs), receiving and inspecting returned DVDs, and customer service. See Netflix [2006] for further details.

[15] ComScore stated that demographics were captured for individual household members as they complete "a detailed opt-in process to participate," for which they were incentivized.

a computer user's visits to all websites that occur at each time of day, and on each day of the week. Time of day was broken into 5 categories, early morning (midnight to 6AM), mid morning (6AM to 9AM), late morning (9AM to noon), afternoon (noon to 5PM), and evening (5pm to midnight).

The data were then cleaned by removing websites associated with malware, third-party cookies, and pornography, leaving $4,788$ popular websites to calculate additional variables.[16,17] The total number of visits to all websites and to each single website were computed for each computer user.

The cross-sectional dataset resulting from the above steps contains Netflix subscription status and a large number of variables for each of $61,312$ computer users.[18] These variables can be classified into three types: standard demographics, basic web behavior, and detailed web behavior. Variables classified as standard demographics were: race/ethnicity, children (Y/N), household income ranges, oldest household member's age range, household size ranges, population density of zipcode from the Census, and Census region. Variables classified as basic web behavior included: total website visits, total unique transactions (excluding Netflix), percent of online browsing by time of day and by day of week, and broadband indicator. Variables classified as detailed web-behavior indicate number of visits to a particular website, with one variable for each of the $4,788$ websites. Each variable in these three sets is normalized prior to estimation.

The data were randomly split into two samples of individuals, a training sample of $56,312$ individuals, and a holdout sample of the remaining 5,000 individuals. The first, an estimation

[16] yoyo.org provides a user-supplied list of some websites of dubious nature. Merging this list with the ComScore data reveal that such websites tend to have very high (≥ 0.9) or very low (≤ 0.1) rates of visits that were referred visits from another website, relative to sites not on the list, and rarely appear on Quantcast's top $10,000$ website rankings. Websites were removed from the data accordingly, dropping sites with low or high rates referred to or not appearing in Quantcast's top $10,000$. Manual inspection revealed these rules were very effective in screening out dubious websites. In addition, Netflix.com and Blockbuster.com were dropped.

[17] Pornography might contain valuable information, but might also require listing perverse website names in publication.

[18] ComScore's dataset was a rolling panel. Computers not observed for the full year were dropped. A couple hundred computer users with missing demographic information were also dropped.

sample, is used for estimating model parameters. The second, a holdout sample, is used to test for and implement a correction for overfitting, in order to yield an unbiased estimate of the model's precision.

3 Model

In Section 3.1, a formal explanation of the model is presented. Section 3.2 provides an intuitive explanation of the model and identification - more formal proofs of identification are shown in Appendix B. Lastly, Section 3.3 explains a technique that avoids overfitting, which otherwise would yield a bias towards strong findings.

Behavior in the model is as follows. Consumers in the model either choose one of Netflix's vertically differentiated goods or the outside good. Consumers agree on the quality levels of each tier, but may differ in how much they value quality. The firm sets prices of the tiers of service, but not qualities.[19]

To be congruent with the context studied, the model presented is designed for data in which prices do not vary over time, which may happen when prices are sticky. Sticky prices substantially mitigate price endogeneity concerns, but require additional information in order for the model to be identified, as is explained later.

3.1 Model

The specification for the conditional indirect utility received from product j follows the theory model from Mussa and Rosen [1978], also adapted to an empirical setting in Bresnahan [1987].[20,21] It is:

[19]In the canonical second-degree PD model, e.g. Mussa and Rosen [1978], firms set both prices and qualities. In this context, however, qualities are not easily set to arbitrary levels, e.g. consumers cannot rent half a DVD.

[20]This paper, unlike Bresnahan [1987], predicts differences in the distribution of willingness to pay across individuals, based on consumer observables.

[21]Other empirical implementations of quality-based price discrimination include Mortimer [2007], McManus [2008].

$$u_{ij} = y_i q_j + \alpha \left(I_i - P_j \right) \tag{1}$$

where y_i is a cardinal measure of individual i's value for Netflix, q_j is an cardinal variable capturing the quality of product j, α is the marginal utility of income, and $(I_i - P_j)$ is the remaining income after paying price P_j. Products j are indexed in increasing order of quality, from 1 to J.

The incentive compatibility constraints can be rearranged to show that a consumer i chooses product j iff:

$$\alpha \frac{P_j - P_{j-1}}{q_j - q_{j-1}} \leq y_i < \alpha \frac{P_{j+1} - P_j}{q_{j+1} - q_j} \tag{2}$$

where the outside good has $(P_0, q_0) = (0, 0)$, and P_{J+1}, the price of the hypothetical better than existing good, is infinite.[22] The marginal utility of income, α, which is not separately identified from the scaling of quality, will henceforth be normalized to one, as usual in these models.[23]

The vertical structure of the model implies that a consumer will choose one of the inside products, as opposed to the outside good, whenever:

$$y_i \geq \frac{P_1 - P_0}{q_1 - q_0} = \psi_1 \tag{3}$$

where ψ_j has been introduced to succinctly represent $\frac{P_j - P_{j-1}}{q_j - q_{j-1}}$

Suppose the firm has imperfect information on individual values of y_i based on individual-specific data X_i, which could for example be web-browsing data. Specifically, assume $y_i = \bar{y}_i + \epsilon_i$, where $\bar{y}_i = E[y_i | X_i]$, and ϵ_i captures the firm's uncertainty.

[22] If the ratios of price differences to quality differences of adjacent-in-quality products are not strictly increasing in quality, then some products will be strictly dominated and never chosen.

[23] See Bresnahan [1987].

Equation 3 implies that if ϵ_i is independent of $\bar{y}_i$ then the probability that a consumer i chooses one of the inside products equals:

$$s_{i,j\geq 1} = 1 - G\left(\psi_1 - \bar{y}_i\right) \tag{4}$$

where $G()$ is the CDF of ϵ_i.

By rearranging Equation 4, one can infer $\bar{y}_i|\psi_1$ from the estimated probability individual i subscribes, $s_{i,j\geq 1}$, where $s_{i,j\geq 1}$ can be estimated beforehand using any method, including complex machine learning methods:

$$\bar{y}_i(\psi_1) = \psi_1 - G^{-1}(1 - s_{i,j\geq 1}) \tag{5}$$

For now assume ψ_1 is known. Later I explain how to estimate ψ_1 using supply side conditions.

One can then estimate the probability density function for y, denoted $f()$:

$$f(y) = \int g(y - \bar{y})h(\bar{y})d\bar{y} \tag{6}$$

where $g(\cdot)$ gives the density of ϵ and $h(\bar{y})$ the estimated density of $\bar{y} = E[y_i|X_i]$. Similarly, the cumulative distribution function for y_i, denoted $F()$, equals:

$$F(y) = \int G(y - \bar{y})h(\bar{y})d\bar{y} \tag{7}$$

Equation 2 implies that the probability a random consumer chooses product j equals:

$$s_j = F\left(\frac{P_{j+1} - P_j}{q_{j+1} - q_j}\right) - F\left(\frac{P_j - P_{j-1}}{q_j - q_{j-1}}\right) = F\left(\psi_{j+1}\right) - F\left(\psi_j\right) \tag{8}$$

Hence, assuming tier choice is missing at random, one can compute the values of $F(\psi_j)$ which match the share choosing a lower tier or not subscribing at all, based on a small sample of individuals for whom tier choice is observed.[24] For $j > 1$:

$$\psi_j = F^{-1}\left(1 - \left(\sum_{k=j}^{J} s_k\right)\right) \tag{9}$$

where F^{-1} is the inverse function of $F()$.

Note, from Equation 5, ψ_1 influences the distribution of y, and hence $F^{-1}(\cdot)$. As a result, ψ_j, for $j > 1$, also depend on ψ_1. While ψ_1 is not separately identified from consumer choice data alone when prices are unvarying, it can however be estimated using optimal pricing conditions.

Conversations with the former vice president of marketing at Netflix revealed that Netflix's prices for each tier had roughly the same percent markup over marginal cost. This implies Netflix used a somewhat simplistic pricing rule, $P_j = \theta c_j$, where c_j is the marginal cost and θ a markup parameter, rather than using full 2nd degree price discrimination. Under this pricing strategy, the expression for profits is:

$$\pi = \sum_j (P_j - c_j) M s_j = \sum_j (\theta - 1) c_j M s_j \tag{10}$$

where M is the mass of consumers.

Assuming a profit-maximizing firm, the markup term θ should satisfy the following optimal pricing first order condition:

$$\frac{d\pi}{d\theta} = M\left(\sum_j c_j s_j + (\theta - 1) c_j \frac{ds_j}{d\theta}\right) = 0 \tag{11}$$

where $\frac{ds_j}{d\theta}$ can be found by by plugging $P_j = \theta c_j$ into Equation 8 and taking the derivative:

[24] The share subscribing to a given tier, s_j, can be rewritten in terms of observable moments: within subscriber share of consumers choosing tier j, $\left(\frac{s_j}{1-s_0}\right)$, and share non-subscribers, s_0: $s_j = \left(\frac{s_j}{1-s_0}\right)(1 - s_0)$.

$$\frac{ds_j}{d\theta} = f(\psi_{j+1})\frac{\psi_{j+1}}{\theta} - f(\psi_j)\frac{\psi_j}{\theta} \tag{12}$$

since $\frac{dF(\psi_j)}{d\theta} = \frac{dF\left(\theta\frac{(c_j - c_{j-1})}{\Delta q_j}\right)}{d\theta} = f(\psi_j)\frac{\psi_j}{\theta}$

Note the supply conditions depend on marginal costs c_j, markup θ, and indirectly on ψ_1, since $f(y)$ and ψ_j for $j > 1$ depend on its value. If marginal costs are known, implying markup θ can be computed directly a priori, then the one remaining unknown, ψ_1, can be estimated from the optimal pricing condition.

3.2 Intuition for Model and Identification

Since the model is rather technical, an intuitive explanation is provided before formally establishing identification in Appendix B.

Figure 1 helps provide intuition for the model's mechanics. On the x-axis is the distribution for individual i's value for quality (affinity for renting movies by mail) y_i, reflecting the inherent uncertainty in our estimate (due to the ϵ error term). Locations further to the right correspond to higher affinity for movies by mail.

If the true value of y_i for an individual is large enough, larger than ψ_1, then the individual's valuation exceeds the threshold for the lowest quality tier, the one DVD at-a-time plan implying the individual buys some plan, as opposed to no plan. If y_i is at least ψ_2, the threshold for the second tier, then the individual prefers the two DVDs at-a-time plan to the one DVD at-a-time plan. Similarly, the consumer prefers three to two DVDs at-a-time when $y_i \geq \psi_3$. Hence, the probability that individual i chooses a given tier j equals the area of the PDF between ψ_j and the next highest threshold ψ_{j+1}. For $j = 1$, the one DVD at-a-time plan, this probability is given by area A in the figure. The probability of subscribing to any plan, the sum of areas A, B, and C, increases as the distribution shifts right relative to the thresholds. The mean of the distribution, relative to the thresholds, varies across individuals according to their web browsing histories.

In the model, machine learning techniques are used to flexibly estimate the probability a given consumer subscribes to some plan based on their web-browsing histories. Note that the probability of choosing any one of the plans, the sum of areas A, B, and C, depends only on the relative locations of the threshold ψ_1 and the mean of the distribution for individual i, denoted $E[y_i]$.[25] Hence, from the probability of subscribing one can infer, for each individual, the implied difference between the mean of the individual's distribution, $E[y_i]$, and the threshold ψ_1. Note, while one could alternatively estimate the difference between $E[y_i]$ and ψ_1 using a simple probit model, the above technique allows one to instead use powerful machine learning techniques to much more precisely estimate individual's propensity to consume, allowing for greater profits under personalized pricing.

The difference between ψ_1 and higher thresholds $\psi_{j>1}$ can later be estimated to match the predicted and actual fraction of consumers subscribing to each plan. Note, this straightforward calculation remains feasible when only aggregate shares consuming each tier is known, which is fortunate because the data in this paper do not contain tier choice for every consumer.

Note that the value of ψ_1 and the average across consumers of $E[y_i]$ are not separately identified. This has direct analogies to the simple probit model. In a probit model, it is assumed the outcome variable equals 1 (rather than zero) if the underlying latent variable exceeds some threshold. This latent variable is approximated as a regression expression, $y_i = \alpha + X_i\beta + \epsilon_i$. It is well known that the intercept α and the threshold are not separately identified, necessitating one of them be normalized, typically to zero.

In the context of this model, however, such normalizations would yield spurious findings. This stems from the fact that the rate at which demand changes with prices depends on the estimated threshold magnitudes. Specifically, $\frac{ds_j}{dP_k}$ is increasing in the threshold magnitudes. Hence, while increasing the cardinal measure y_i and threshold ψ_1 by the same amount would

[25]The distributions standard deviation is normalized to one, since it is not separately identified from other parameters. This assumption is discussed further in Section B

yield the same predictions at observed prices, predictions at counterfactual prices would change.

This problem can be overcome by estimating the magnitude of ψ_1 to satisfy optimal pricing conditions. A formal proof establishing identification is provided in Appendix B.

This same figure, Figure 1, also provides intuition for computing counterfactual demand at unobserved prices. Note the thresholds, $\psi_j = \frac{P_j - P_{j-1}}{q_j - q_{j-1}}$, depend on the prices. Hence, any given set of prices implies some probability that individual i consumes each tier. The expected revenues from the individual in Figure 1 equals $P_1 * Area\ A + P_2 * Area\ B + P_3 * Area\ C$, where the areas depend on prices and estimates of y_i. Total expected revenues are then found by summing expected revenues across individuals.

3.3 Addressing Overfitting

Using thousands of potential explanatory variables to estimate the probabilities individuals subscribe, $s_{i,j\geq 1}$, immediately raises concerns about overfitting. Conceptually, overfitting causes two different but related problems in this context.

First, naive model selection may yield a sub-optimal model, with too many explanatory variables included, and poor out-of-sample predictions. For example, naively selecting all variables with p-values below 0.05 may yield many type 1 errors, adding noise to predictions in fresh samples. This noise may be large. Hence, a less complex model chosen using more stringent conditions for variable selection may offer better out-of-sample predictions, even if also excluding a few variables which are helpful for predictions.[26] But too strict a model might throw out many variables useful for predictions just to avoid a few more type 1 errors. So, even after following well established machine learning methods for model selection, a good model will likely still include some false positives as predictors.

The second conceptual problem from overfitting follows as a result. Including type 1 errors as predictors yields a better in-sample fit, with lower error magnitude. However, it would not

[26] Over-complex model selection is analogous to “high variance” in the machine learning literature.

apply in fresh samples, thus implying the predictions of reservation prices are not as precise as the model suggests, and therefore profits from personalizing prices are overestimated.

This second problem, overly optimistic in-sample precision, can be corrected by ex-post re-estimating the precision using a holdout sample. Typically, precision is determined by the error size. However, in a choice model framework, the error's standard deviation is typically normalized, and the precision is determined by the scale of other parameters, i.e. $\bar{y}_i - \psi_1$ in this context, relative to the fixed scale of the error, $\sigma = 1$. One can appropriately rescale $\bar{y}_i - \psi_1$ by simply entering it as the sole explanatory variable in a probit model run on the sample of individuals held out from all previous estimation steps. The probit model's predicted difference between the underlying latent variable and the threshold then comprises the rescaled estimate of $\bar{y}_i - \psi_1$.

After this correction, predictions from any chosen model, whether well or poorly chosen, will in expectation reflect the true level of uncertainty according to that model. Therefore, following this correction, this paper is no longer biased towards strong findings. If anything, the opposite is true. If I, the researcher, were to poorly choose a model in the first step, this would imply that a more skilled statistician could choose a better model and extract even greater profits from personalized pricing.

4 Estimation

Before explaining the main model, I begin with a simple exercise to convey a sense of which websites when visited frequently raise or lower the probability a consumer subscribes to Netflix. Specifically, I run a series of probit regressions predicting Netflix subscription. All include total website visits, its square, and an indicator for a broadband connection, to control for aggregate browsing behavior. Each regression additionally includes one, and only one, website variable. The significance of the website variable from each of these regressions is recorded. Overall, 38% of websites were significant at the 5% level, and 27% at the 1% level, far more than expected by chance alone.

The types of websites found to be most significant, shown in Table 1, and their positive signs, intuitively suggest that consumers' observed web browsing behavior is driven by some innate characteristics that influence their value for Netflix.[27] They are comprised of websites which are likely used by movie lovers (IMDB, Rotten Tomatoes), those preferring mail ordering (Amazon, GameFly), those with preferences for hard-to-find content (Alibris), discount shoppers (BizRate), and internet savvy users (Wikipedia).

The next step is to use the many variables jointly to predict the probability each consumer subscribes, $s_{i,j\geq 1}$. There are multiple machine learning methods one could choose from, with different benefits. Classification trees fit well when explanatory variables have a discrete impact on the classification probabilities if a threshold value is exceeded. If the explanatory variables instead have a continuous impact, then LASSO regressions may work better.

Various methods can be combined using an adaptation of Friedman and Popescu [2008]'s RuleFit algorithm. Specifically, in this context, one can begin by considering N nonlinear models, e.g. cross-validated trees with different "tuning parameters," based on the J explanatory variables x_j. Each model "n" yields its own predicted subscription probabilities, $s^n_{i,j\geq 1}$.[28] These predicted probabilities, along with the raw variables in the data, x_j, are added as explanatory variables in a linear probability LASSO regression model:

$$I(subscribe) = \alpha + \sum_{n=1}^{N} \beta_n s^n_{i,j\geq 1} + \sum_{j=1}^{J} \gamma_j x_j + \lambda \left(\sum_{n=1}^{N} \beta_n + \sum_{j=1}^{J} \gamma_j \right) \tag{13}$$

where λ is a shrinkage term for the parameters β_n and γ_j, whose value is found by minimizing

[27] Such correlations may be partially driven by Netflix's own actions, for example Netflix may advertise more frequently on certain websites. This does not, however, pose a problem for the current analyses. As long as consumers were aware Netflix existed, which seems likely given 1 in 7 households subscribed, it does not matter why a given consumer is or is not likely to subscribe. Regardless of the reason, the firm may profit by raising the price to consumers predicted to be highly likely to purchase at a given price, and vice versa for consumers predicted unlikely to purchase.

[28] $s^n_{i,j\geq 1}$ is computed as the average predicted probability across the models for each fold, in 10-fold cross-validation.

the mean squared error in held-out folds, using 10-fold cross-validation. The LASSO regression accomplishes variable selection - it shrinks coefficients on weak variables, and partially redundant variables, to zero.

The variables selected by LASSO in a linear probability model are then entered into a probit regression model, which is slower, but addresses some shortcomings of the former.[29] In particular, a probit regression restricts predicted probabilities to lie in range [0,1], which is necessary for implementation in the model presented earlier. The model probability variables, $s^n_{i,j\geq 1}$, are transformed to variables which imply that predicted probability in a probit model $(\bar{y}_i = \psi_1)$ before being entered as explanatory variables.

The coefficient estimates from this probit model are then used to compute initial estimates of the predicted probabilities of subscribing among a sample of individuals held out from all earlier steps. The values of $\bar{y}_i - \psi_1$ which correspond to these initial probability estimates via Equation 4, are then entered as the sole explanatory variable in a new probit model. The resulting predicted probabilities from this second probit model run in a holdout sample are corrected for overfitting, as described in Section 3.3.

The above procedure was run separately using demographic variables, and the full set of variables, to calculate the predicted subscription probabilities. Classification trees were found not to offer any marginal predictive ability for the full model - including predicted probabilities from classification trees models, $s^n_{i,j\geq 1}$, as explanatory variables in the LASSO regressions did not lessen the mean squared error in the holdout sample, suggesting x_j have a continuous impact on predicted subscription probabilities. The remaining model parameters were calculated ex-post as described in Section 3.1.

In the machine learning literature, a common method for measuring model fit is a “confusion matrix,” a two-by-two matrix with each column representing the most likely class and each row representing the actual class. It hence summarizes the type I and type II errors.

[29] The variables selected in a cross-validated model may differ for each fold. All variables selected in any fold were included in the probit model.

The confusion matrix when all variables are included, shown in Table 2, shows the model fits well in some regards. Individuals with a predicted subscription probability exceeding 0.5 are in fact much more likely than other consumers to subscribe. But, because most individuals have probabilities below 0.5, most consumers are predicted not to subscribe, including many who in fact do. When only demographics are used as explanatory variables, all consumers are predicted not to subscribe.

In this context, the confusion matrix obscures relevant information. Consider, for example, a model which identifies many consumers with a 0.49 probability of subscribing, but none with a probability exceeding 0.5. Clearly, this model does a much better job at identifying consumers the firm may want to charge a higher price than a model which assigns everyone the average predicted probability of about 0.16. However, these two models would generate the same confusion matrix.

Figure 2 offers another test of the full model's performance, and a check on its fit. Specifically, individuals in the holdout sample are ordered according to the model's predicted probability they subscribe to Netflix, then split into 200 groups. The average predicted probability and observed probabilities, i.e. fraction buying, are then calculated for each group. Figure 2 shows the predicted probabilities, shown in the solid blue line, do in fact seem to follow the actual probabilities of subscription. There is also a substantial range of predicted probabilities.

One might think that geographic variation in preferences, possibly due to Tiebout sorting and preference externalities (George and Waldfogel [2003]), might proxy for the information in web browsing which predicts subscription. This does not, however, appear to be the case. To test this, the distribution of differences in predicted subscription probabilities for pairs of individuals in the same zipcode is compared with the corresponding distribution for randomly drawn pairs living in different zipcode. Figure 3, which plots overlaid histograms of absolute differences in predicted probabilities, shows there is not a meaningful difference in these groups' propensities to consume Netflix.[30] Hence, web browsing data offer mostly distinct information

[30]While not meaningful, the difference is statistically significant, according to the Kolmogorov-Smirnov test.

from that contained by geography.

5 Counterfactual Analyses

This section simulates counterfactual environments in which Netflix implements personalized pricing, proper second degree price discrimination, or both. Specifically, optimal profits under each pricing scheme are simulated separately, first using demographics alone and then all variables to explain a consumer's willingness to pay.[31] They are then compared with simulated profits under the status quo environment, where Netflix employed constant markup pricing.

Table 3 shows the percent increase in profits from personalized markups.[32] When all variables are used to set prices, the profits under personalized markups are 14.55% higher.[33] If personalizing prices are based only on demographics, the increase in total profits is much less, 0.30%. Since adding web browsing data substantially increases the amount by which personalized pricing raises profits, it increases the likelihood that firms will implement it.

Table 4 shows that the increase in profits from price personalization is relatively large compared to 2^{nd} degree Price Discrimination (PD). Changing from the status quo case, constant markups over cost, to 2^{nd} degree PD increases profits by 8.87%. Switching instead to personalized markups raises profits by more, 14.55%. Combining the two strategies raises profits by 25.50%, which is more than the sum of the gains from each strategy on its own.

Using the full set of variables to personalize markups substantially increases the range of

[31] Under second degree price discrimination, multiple sets of prices might satisfy the optimal pricing FOCs, indicating local minima are possible. To address this issue, globally profit maximizing prices are found via grid search using 5 cent intervals.

[32] Percentages rather than absolute profits were reported because simulated variable profits in the status quo case depend on the demand estimates, which can vary slightly depending on which set of variables were used in estimation. In practice, the two status quo profit estimates were quite close, within about half of a percent of each other.

[33] In this calculation, variable costs are defined as the "cost of revenues" reported in Netflix's 2006 Annual Report Netflix [2006]. The "operating expenses" in the 2006 financial report are assumed to be fixed costs. These definitions imply the variable costs were about \$627 million, and the fixed costs were about \$305 million. Revenues in 2006 were about \$997 million, implying variable profits were about \$370 million, and total profits were about \$65 million. Multiplying variable profits by $\frac{370}{65}$ yields total profits.

prices charged to different individuals for the same product. Table 5 shows the percentiles of percent differences between personalized prices and non-personalized prices. When all variables are used, the consumer estimated to have the highest value for Netflix would face prices about 63% higher than they do under non-personalized prices. The 99.9^{th} percentile individual would face prices about 41% higher, the 99^{th} percentile about 17% higher, and the 90% percentile about 4% higher. The 75^{th} percentile consumer gets about a 1% discount, and the median consumer a 4% discount. The lowest price is nearly 10% lower. These results together imply that the highest prices offered are almost double the prices other customers are offered, for the exact same good.

Personalizing markups reduces aggregate consumer surplus by 1.79%. However, most consumers receive lower prices when prices are personalized, and hence are better off. Figure 4 shows that the impact is highly skewed. Most consumers receive a few cents of additional consumer surplus. But a few consumer are much worse off.

The firm could raise profits without harming consumers. When personalizing the markup, the firm could set a maximum equal to the non-personalized markup, and still raise profits by 3.69% percent without making any consumer worse off. Or, it could raise the maximum personalized markup from 59.1% to 81.7% percent, and raise profits by 12.24% percent while leaving aggregate consumer surplus unchanged.

5.1 Robustness Checks

Table 6 shows the profit increase from personalized pricing is robust to some modeling assumptions. The first concern is that Netflix may have under-priced in the short run to grow the business, and hence the static optimal pricing conditions may rely on a false assumption. However, I find that even if one assumes the optimal prices were double the observed prices, the increase in profits from price personalization is roughly the same, at least in percentage terms. The second concern is that movie review websites - IMDB.com, RottenTomatoes.com, and Amazon.com - might be complements for Netflix's products. If so, visiting them might not just indicate a higher intrinsic affinity for Netflix, but rather (also) indicate that a consumer already subscribes. While this may be true, their impact on the main results is small. Table 6 shows dropping these websites and re-running the model only lowers the percent gain from price personalization from 14.55% to 13.45%.

6 Discussion and Conclusion

This paper finds, in one context, that the increase in profits made feasible by personalized pricing is much higher when web browsing behaviors (14.55%), rather than just demographics (0.30%), are used to predict individuals' reservation prices. This meaningful profit increase made possible by web browsing data supports the argument that personalized pricing is evolving from merely theoretical to practical and widely employed. This will directly impact consumers, as consumer surplus is lower, and the range of prices offered to different individuals for use of the same product is quite large.

The findings in this paper raise several questions about the efficiency and equity effects of widespread personalized pricing. Most textbooks espouse efficiency of first degree price discrimination based on partial equilibrium analysis. However, when employed by multiple firms, this result may not hold. In oligopolistic [Spulber [1979]] and differentiated product [Corts [1998], Thisse and Vives [1988], Choudhary et al. [2005]] markets, first-degree PD does unilaterally raise profits, but employed jointly it may increase competition, reducing profits and hence innovation incentives. A related question is whether it is fair for consumers to pay different prices for the same product. There is no objective answer, but there appears to be a public near-consensus. Kahneman et al. [1986] find personalized pricing was viewed as unfair by 91% of respondents.

Lastly, the findings in this paper suggest a fundamental change in the way price discrimination is taught. Typically in undergraduate and MBA microeconomics classes, price personalization is taught as the theoretical pricing strategy, in order to develop intuition for PD and use as a benchmark for other forms of pricing. Now, or soon, personalized pricing may be much more than just a theoretical thought experiment.

References

http://finance.yahoo.com/news/pikato-launches-retail-mobile-marketing-130000010.html. *Yahoo Finance*, 2014.

Alessandro Acquisti and Hal R Varian. Conditioning prices on purchase history. *Marketing Science*, 24(3):367–381, 2005.

Inaki Aguirre, Simon Cowan, and John Vickers. Monopoly price discrimination and demand curvature. *The American Economic Review*, 100(4):1601–1615, 2010.

Timothy F Bresnahan. Competition and collusion in the american automobile industry: The 1955 price war. *The Journal of Industrial Economics*, pages 457–482, 1987.

Judith Chevalier and Austan Goolsbee. Are durable goods consumers forward-looking? evidence from college textbooks. *The Quarterly Journal of Economics*, 124(4):1853–1884, 2009.

Vidyanand Choudhary, Anindya Ghose, Tridas Mukhopadhyay, and Uday Rajan. Personalized pricing and quality differentiation. *Management Science*, 51(7):1120–1130, 2005.

Chenghuan Sean Chu, Phillip Leslie, and Alan Sorensen. Bundle-size pricing as an approximation to mixed bundling. *The American Economic Review*, 101(1):263–303, 2011.

Kenneth S Corts. Third-degree price discrimination in oligopoly: all-out competition and strategic commitment. *The RAND Journal of Economics*, pages 306–323, 1998.

The The Economist. Getting to know you - everything people do online is avidly followed by advertisers and third-party trackers. *The Economist Magazine*, 2014.

Jerome H. Friedman and Bogdan E. Popescu. Predictive learning via rule ensembles. *The Annals of Applied Statistics*, 2(3):pp. 916–954, 2008. ISSN 19326157.

Drew Fudenberg and J Miguel Villas-Boas. Behavior-based price discrimination and customer recognition. *Economics and Information Systems (Handbook of Information Systems)*, 1: 377–433, 2007.

Lisa George and Joel Waldfogel. Who affects whom in daily newspaper markets? *Journal of Political Economy*, 111(4):765–784, 2003.

Anindya Ghose, Panagiotis G Ipeirotis, and Beibei Li. Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, 31(3):493–520, 2012.

Kathryn Graddy. Testing for imperfect competition at the Fulton fish market. *The RAND Journal of Economics*, pages 75–92, 1995.

William H Greene and David A Hensher. Modeling ordered choices: A primer. *Cambridge University Press*, 2010.

Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 305–318. ACM, 2014.

Peng Huang, Nicholas H Lurie, and Sabyasachi Mitra. Searching for experience on the web: an empirical examination of consumer behavior for search and experience goods. *Journal of Marketing*, 73(2):55–69, 2009.

Daniel Kahneman, Jack L Knetsch, and Richard Thaler. Fairness as a constraint on profit seeking: Entitlements in the market. *The American economic review*, 76(4):728–741, 1986.

Brian McManus. Nonlinear pricing in an oligopoly market: The case of specialty coffee. *The RAND Journal of Economics*, 38(2):512–532, 2008.

Jakub Mikians, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. Detecting price and search discrimination on the internet. *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, pages 79–84, 2012.

Eugenio J Miravete, Katja Seim, and Jeff Thurk. Complexity, efficiency, and fairness in multiproduct liquor pricing. *The Wharton School Research Paper*, (74), 2014.

Wendy W Moe and Peter S Fader. Dynamic conversion behavior at e-commerce sites. *Management Science*, 50(3):326–335, 2004.

Alan L Montgomery, Shibo Li, Kannan Srinivasan, and John C Liechty. Modeling online browsing and path analysis using clickstream data. *Marketing Science*, 23(4):579–595, 2004.

Julie Holland Mortimer. Price discrimination, copyright law, and technological innovation: Evidence from the introduction of dvds. *The Quarterly journal of economics*, 122(3):1307–1350, 2007.

Michael Mussa and Sherwin Rosen. Monopoly and product quality. *Journal of Economic theory*, 18(2):301–317, 1978.

Netflix. Netflix annual report. *http://ir.netflix.com/annuals.cfm*, 2006.

Executive Office of the President. Big data and differential pricing. *Washington DC*, 2015.

Joan Robinson. Economics of imperfect competition. 1933.

Peter E Rossi, Robert E McCulloch, and Greg M Allenby. Hierarchical modelling of consumer heterogeneity: An application to target marketing. In *Case Studies in Bayesian Statistics, Volume II*, pages 323–349. Springer, 1995.

Peter E Rossi, Robert E McCulloch, and Greg M Allenby. The value of purchase history data in target marketing. *Marketing Science*, 15(4):321–340, 1996.

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

Andrea Shepard. Price discrimination and retail configuration. *Journal of Political Economy*, pages 30–53, 1991.

Benjamin Shiller and Joel Waldfogel. Music for a song: an empirical look at uniform pricing and its alternatives. *The Journal of Industrial Economics*, 59(4):630–660, 2011.

Daniel F Spulber. Non-cooperative equilibrium with price discriminating firms. *Economics Letters*, 4(3):221–227, 1979.

Jacques-Francois Thisse and Xavier Vives. On the strategic choice of spatial price policy. *The American Economic Review*, 78(1):122–137, 1988.

Joel Waldfogel. First degree price discrimination goes to school. *Working Paper*, 2014.

A Observational equivalence of scaling σ under second degree price discrimination

Theorem 1. *Under optimized second degree price discrimiantion, the derivative of demand for any given product j with respect to its price, or the price of another product, is invariant to an arbitrary positive rescaling σ applied to both ψ_j and y_i.*

Proof. Note, from Equation 8, that the change in shares as price(s) change can be summarized by $\frac{\partial F(\psi_j)}{\partial P_j}$ and $\frac{\partial F(\psi_j)}{\partial P_{j-1}}$, and it can be shown that $\frac{\partial F(\psi_j)}{\partial P_j} = -\frac{\partial F(\psi_j)}{\partial P_{j-1}}$, where $\psi_j = \frac{P_j - P_{j-1}}{\Delta q_j}$. If one multiplies y by a positive constant σ yielding a new distribution with CDF and PDF denoted F_σ and f_σ, and also multiplies ψ_j by σ, then the change in shares under this new scaling depends only on the analogous values of $\frac{\partial F_\sigma(\sigma\psi_j)}{\partial P_j} = -\frac{\partial F_\sigma(\sigma\psi_j)}{\partial P_{j-1}}$. Thus, showing $\frac{\partial F_\sigma(\sigma\psi_j)}{\partial P_j} = \frac{\partial F(\psi_j)}{\partial P_j}$ is sufficient to complete the proof.

Expanding the derivative for the rescaled distribution, and plugging in $f_\sigma(\sigma\psi_j) = \frac{1}{\sigma} f(\psi_j)$:[34]

[34] First note that $F(y) = F_\sigma(\sigma y)$, i.e. y and σy have the same cumulative density from their respective

$$\frac{\partial F_\sigma(\sigma\psi_j)}{\partial P_j} = \frac{\sigma}{\Delta q_j} f_\sigma\left(\sigma\psi_j\right) = \frac{1}{\Delta q_j} f\left(\psi_j\right) = \frac{\partial F(\psi_j)}{\partial P_j}$$

□

B Technical Identification Explanation

For some components, identification is straightforward. Subscription probabilities, $s_{i,j\geq 1}$, are estimated using established machine learning techniques, for which identification is well understood. Furthermore, for an assumed distribution for ϵ and value of ψ_1, $\bar{y}_i$ and ψ_j can be computed directly from Equations 5 and 9.

The remainder of this subsection focuses on less obvious components: (1) identification of ψ_1 and (2) the relationship between the assumed distribution of ϵ and the estimate of y's distribution.

The following theorem shows there is a unique value of ψ_1 which satisfies the optimal pricing condition.

Theorem 2. *One can easily see from Equation 3 that adding any constant to y_i and ψ_1 yields new values which imply the same share subscribing at observed prices. However, a unique value satisfies the optimal pricing conditions, so long as $\sum_j \left(\frac{\theta-1}{\theta}c_j\Big(f(\psi_{j+1}) - f(\psi_j)\Big)\right) \neq 0$. Otherwise, all values do.*

Proof. Suppose the values of y_i and ψ_1 satisfy the optimal pricing condition. Now suppose an additive shift applied to y_i and ψ_1 denoted by a " $'$ " symbol: $y' = y + A$ and $\psi_1' = \psi_1 + A$. Note if y and ψ_1 are additively shifted by A, then ψ_j for $j > 1$ must also be shifted by A in order to continue matching predicted and actual shares of consumers choosing specific quality tiers.

distributions. Second, recall that the area under a PDF must integrate to one, which can only remain true after stretching a distributions along the x-dimension by a scale σ if the height at each corresponding cumulative density is scaled downwards by scale $\frac{1}{\sigma}$.

Denote the PDF and CDF of $y' = y + A$ as $f_A()$ and $F_A()$, respectively. Note that $f_A(\psi'_j) = f(\psi_j)$. Recall that $\frac{dF(\psi_j)}{d\theta} = \frac{\psi_j}{\theta} f(\psi_j)$. Combining these two points, one can find that: $\frac{dF_A\left(\psi'_j\right)}{d\theta} = \frac{\psi_j + A}{\theta} f(\psi_j)$. Using this result, one can show that, after changing to additively shifted variables y' and ψ'_j, the pricing condition in Equation 11 simplifies to:

$$
\begin{aligned}
&\tfrac{d\pi}{d\theta} = M\left(\textstyle\sum_j (c_j s_j) + \sum_j \left((\theta-1)c_j \tfrac{ds_j}{d\theta}\right)\right) \\
&= M\left(\textstyle\sum_j (c_j s_j) + \sum_j \left(\tfrac{\theta-1}{\theta} c_j \left(f(\psi_{j+1})(\psi_{j+1} + A) - f(\psi_j)(\psi_j + A)\right)\right)\right) \\
&= M\left(\textstyle\sum_j (c_j s_j) + \sum_j \left(\tfrac{\theta-1}{\theta} c_j \left(f(\psi_{j+1})\psi_{j+1} - f(\psi_j)\psi_j\right)\right) + A\sum_j \left(\tfrac{\theta-1}{\theta} c_j \left(f(\psi_{j+1}) - f(\psi_j)\right)\right)\right)
\end{aligned}
$$

The last component, $A\sum_j \left(\frac{\theta-1}{\theta} c_j \left(f(\psi_{j+1}) - f(\psi_j)\right)\right)$, captures the difference in the pricing condition from adding constant A to both $\bar{y}_i$ and ψ_j. So long as $\sum_j \left(\frac{\theta-1}{\theta} c_j \left(f(\psi_{j+1}) - f(\psi_j)\right)\right) \neq 0$, then the expression is monotonic in A, and if the condition is met for $A = 0$, then it will not be met for any other value of A. If $\sum_j \left(\frac{\theta-1}{\theta} c_j \left(f(\psi_{j+1}) - f(\psi_j)\right)\right) = 0$, then all values of A satisfy the optimal pricing condition. $\square$

The above proof shows that either a unique value of ψ_1 satisfies the optimal pricing conditions, or all do - an unlikely scenario which can easily be recognized during estimation.

Next, it is shown that the scale of y's distribution, which is fixed by the assumed standard deviation σ of the prediction error ϵ, is irrelevant. First note that changing the scale of ϵ in the model from 1 to some other positive value σ will result in parameter estimates which are likewise scaled up by σ - the same reasoning explains why the standard deviation of the error is normalized in a probit model.[35] Furthermore, scaling all parameters by positive σ has no impact on whether Equation 2 holds, implying it has no impact on predicted consumer choices at observed prices. Moreover, as the proof in Theorem 3 shows, it does not impact the slope of demand, not just at observed prices, but at any hypothetical markup. Since all values of σ lead to identical demand at observed prices, and demand that always changes at the same rate as price changes, they imply identical demand functions. Thus any assumed scale σ leads to equivalent outcomes and model predictions.

[35] See Greene and Hensher [2010], page 22

Theorem 3. *For any initial markup θ, the slope of demand, $\frac{ds_j}{d\theta}$, is invariant to a positive rescaling σ applied to both ψ_j and y_i.*

Proof. Note that $\psi_j = \theta\left(\frac{c_j - c_{j-1}}{\Delta q_j}\right)$ is defined for arbitrary markup θ, whether or not matching the observed markup of $\approx 58\%$. Denote with a " $'$ " symbol the variables scaled by σ. I.e. $\psi'_j = \sigma\psi_j$, and $y'_i = \sigma y_i$. Recall that the slope of demand, $\frac{ds_j}{d\theta}$, depends only on the values of $\frac{dF(\psi_j)}{d\theta}$. Thus, if the cumulative distribution of $y'_i = \sigma y_i$ is denoted $F_\sigma()$, then showing $\frac{dF_\sigma(\psi'_j)}{d\theta} = \frac{dF(\psi_j)}{d\theta}$ is sufficient to complete the proof.

Note that $f_\sigma\left(\psi'_j\right) = \frac{1}{\sigma} f\left(\psi_j\right)$.[36] Also recall that $\frac{dF(\psi_j)}{d\theta} = \frac{\psi_j}{\theta} f\left(\psi_j\right)$. It follows that:

$$\frac{\partial F_\sigma(\psi'_j)}{d\theta} = \frac{\psi'_j}{\theta} f_\sigma\left(\psi'_j\right) = \frac{\sigma\psi_j}{\theta} f_\sigma\left(\psi'_j\right) = \frac{\psi_j}{\theta} f\left(\psi_j\right) = \frac{dF(\psi_j)}{d\theta}$$

$\square$

Since demand does not depend on scaling σ, it follows that the same markup term value θ satisfies the supply side condition after a change to the scaling. Appendix A shows the reasoning in Theorem 3's proof can be extended from the constant markup case, considered here, to the case of full second degree price discrimination.

Lastly, identification of the shape of y's distribution is discussed. As Equation 6 shows, the probability that a randomly drawn consumer has a particular value of y equals the weighted average probability that each consumer has that value of y, which itself depends on the assumed shape of the ϵ's distribution. However, in the limit, as predictions approach perfection, the error shrinks and the distribution of ϵ becomes degenerate, and thus all functional form assumptions for ϵ are equivalent.[37] Thus, in the limit, all functional form assumption imply the same single value of y_i for individual i, and hence by aggregation the same shape of the distribution of

[36]First note that $F(y) = F_\sigma(\sigma y)$, i.e. y and σy have the same cumulative density from their respective distributions. Second, recall that the area under a PDF must integrate to one, which can only remain true after stretching a distributions along the x-dimension by a scale σ if the height at each corresponding cumulative density is scaled by $\frac{1}{\sigma}$.

[37]The analogous change for fixed standard deviation of ϵ is an increase in the scaling of all <u>other</u> parameters. The implications for the shape of y's distribution does not depend on such semantics.

y across individuals. By contrast, if the data offer no predictive power on differences across individuals, then y's distribution follows the assumed distribution of ϵ exactly.

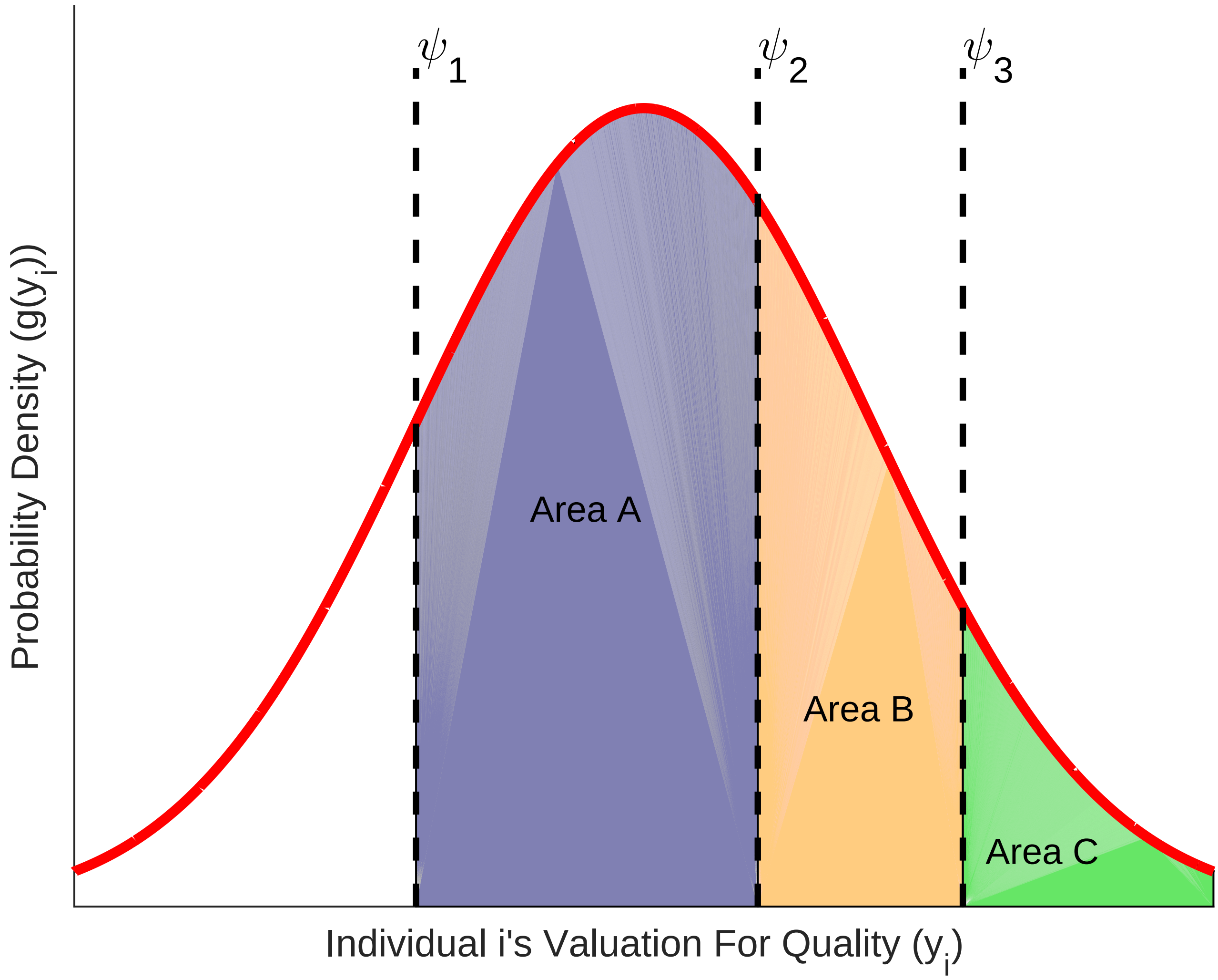

Figure 1: Graphical Intuition For Model

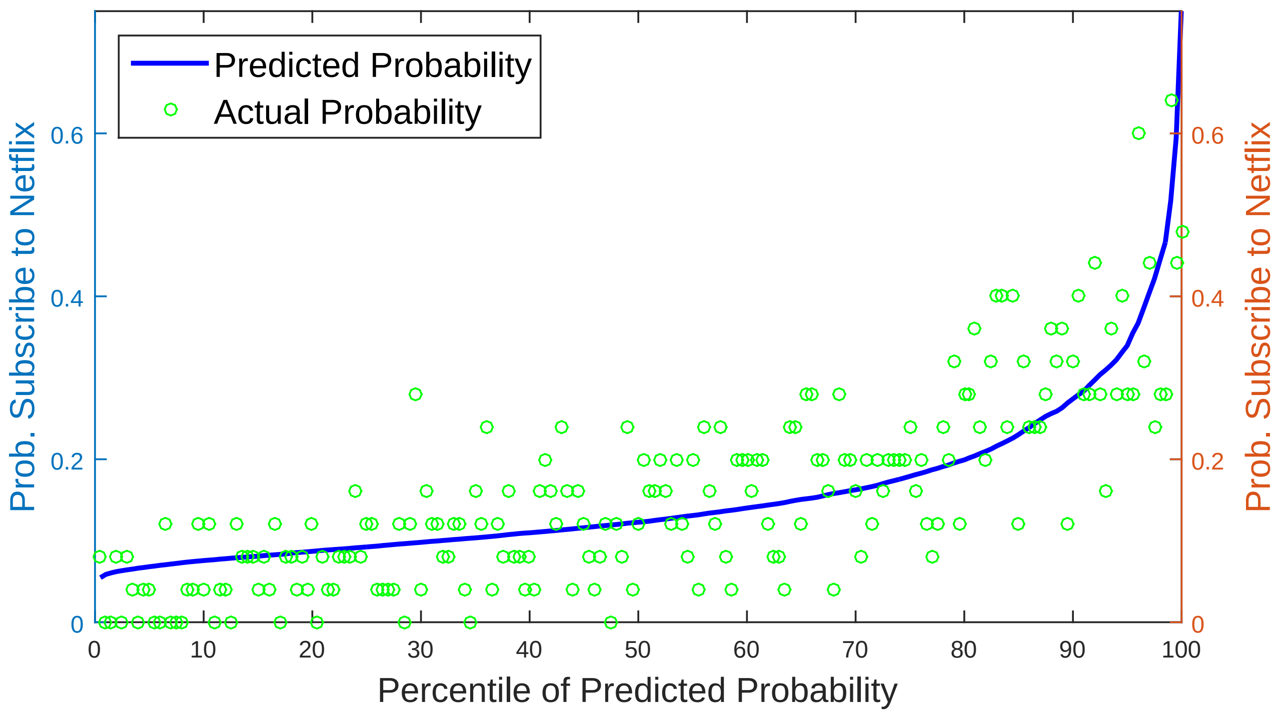

Figure 2: Model Fit - Predicted Probabilities in Holdout Sample When All Variables Used

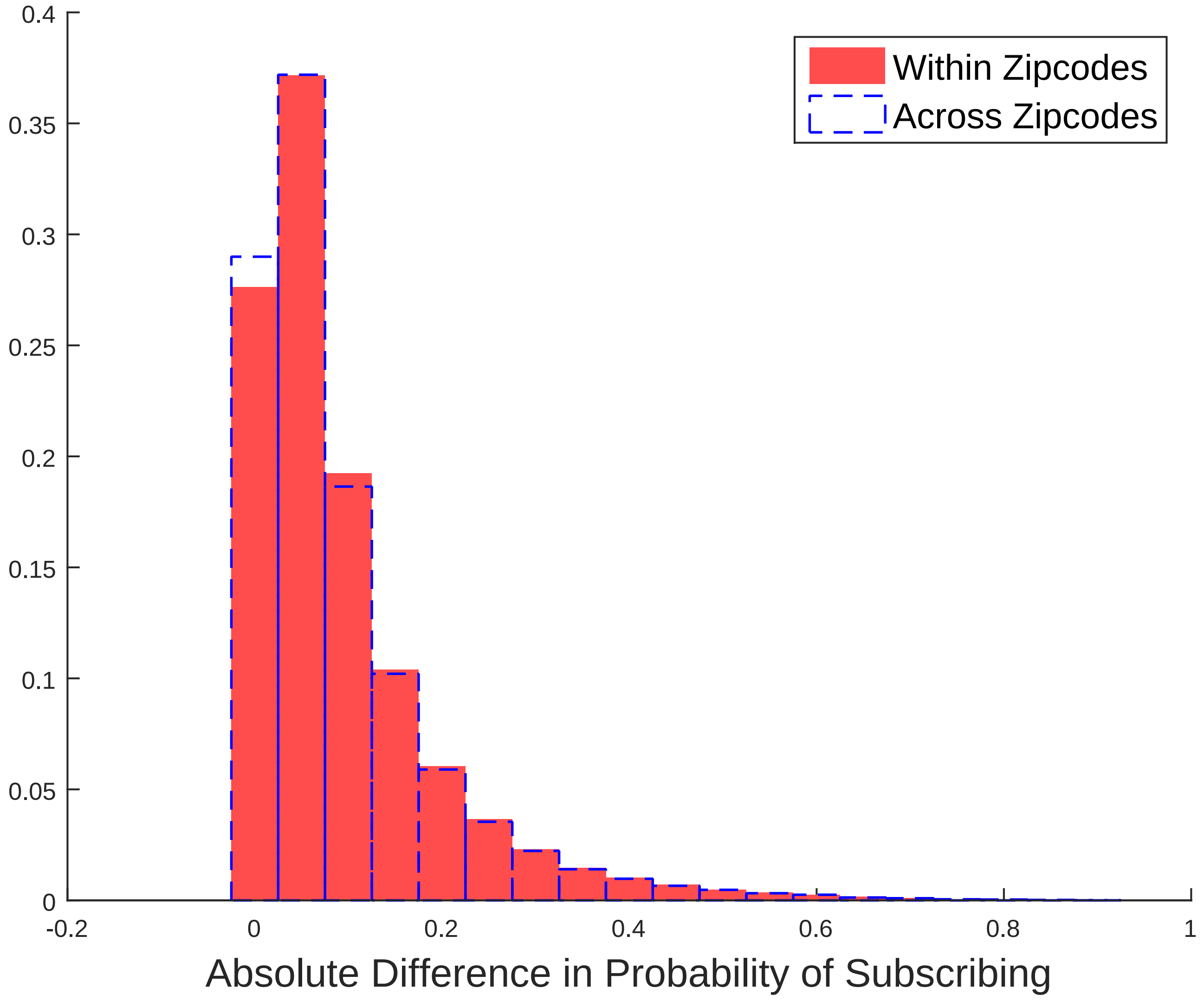

Figure 3: Histograms of Absolute Differences in Probabilities of Subscribing to Netflix, in Randomly Drawn Pairs of Individuals, Both Within and Across Zipcodes

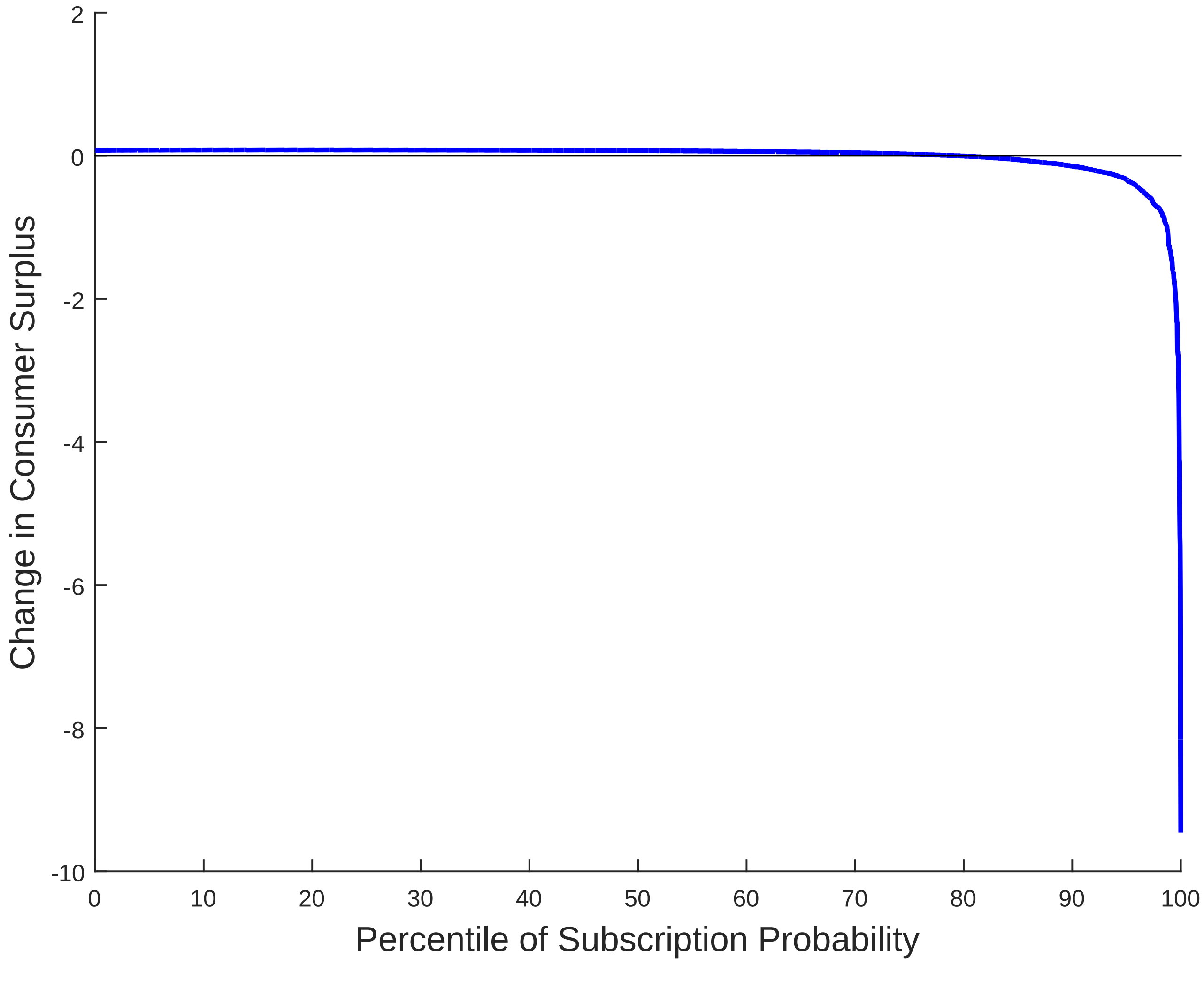

Figure 4: Distribution of the Change in Consumer Surplus (in Dollars) From Personalized Pricing

Table 1: Websites Best Predicting Netflix Subscription (All Are Positive)

Rank	Website Name	Rank	Website Name
1	amazon.com	11	become.com
2	bizrate.com	12	about.com
3	citysearch.com	13	pricegrabber.com
4	gamefly.com	14	wikipedia.org
5	shopping.com	15	epinions.com
6	imdb.com	16	barnesandnoble.com
7	rottentomatoes.com	17	msn.com
8	target.com	18	alibris.com
9	shopzilla.com	19	overstock.com
10	dealtime.com	20	smarter.com

Table 2: Confusion Matrix

	Predicted Class	
Actual Class	Subscriber	Non-Subscriber
Subscriber	33	745
	(3.29)	(3.29)
Non-Subscriber	31	4191
	(2.32)	(2.32)

Bootstrapped standard errors in parentheses.

Table 3: Simulated Changes in Various Outcomes Resulting From Personalized Markups

	Percent Change When Price Based on:	
	Demographics	All Variables
Total Profits	0.30%	14.55%
	(0.03)	(0.42)
Subscribers	0.07%	3.14%
	(0.02)	(0.11)
Sales (DVDs At-a-Time)	0.04%	2.13%
	(0.02)	(0.12)
Aggregate Consumer Surplus	−0.05%	−1.79%
	(0.02)	(0.13)
Joint Surplus	0.00%	0.55%
	(0.01)	(0.05)

Bootstrapped standard errors in parentheses.

Table 4: Percent Increase in Profits, Relative to Non-Personalized Linear Pricing

	Non-Personalized	Personalized
Linear Pricing	N/A	14.55%
		(0.42)
2nd Degree PD	8.87%	25.50%
	(0.02)	(0.45)

Bootstrapped standard errors in parentheses.

Table 5: Percent Difference Between Personalized and Non-Personalized Prices

Price Percentile	Percent Difference When Markup Based On: Demographics	All Variables
Lowest	−2.01%	−9.17%
	(0.23)	(0.40)
0.1	−2.01%	−9.05%
	(0.22)	(0.24)
1.0	−1.89%	−8.42%
	(0.17)	(0.16)
10.0	−1.26%	−7.16%
	(0.08)	(0.12)
25.0	−0.75%	−6.03%
	(0.08)	(0.10)
50.0	−0.38%	−4.15%
	(0.09)	(0.09)
75.0	0.38%	−1.01%
	(0.08)	(0.07)
90.0	1.64%	3.64%
	(0.13)	(0.12)
99.0	2.52%	17.34%
	(0.18)	(0.54)
99.9	3.77%	40.70%
	(0.32)	(2.03)
Highest	5.03%	63.94%
	(0.82)	(4.14)

Bootstrapped standard errors in parentheses.

Table 6: Robustness Check

		Percent Increase in Profits, When Price Personalization Based On:	
		Demographics	All Variables
Main Model		0.30%	14.55%
		(0.03)	(0.42)
Robustness Checks			
	Excluding Movie Review Sites	N/A	13.45%
			(0.39)
	Less Price Sensitive Model	0.29%	13.90%
		(0.03)	(0.40)

Bootstrapped standard errors in parentheses.