Approximating Purchase Propensities and Reservation Prices from Broad Consumer Tracking

Benjamin Reed Shiller*
Brandeis University

November 5, 2018

Abstract

A consumer’s web-browsing history, now readily available, may be much more useful than demographics for both behaviorally targeting advertisements and personalizing prices. Using a method that combines economic modeling and powerful machine learning techniques, I find a striking difference. Using demographics yields purchase probabilities at observed prices ranging across individuals from about 8% to about 30%. Adding consumers’ web-browsing histories increases this range to about 5% to 90%, allowing more precise behavioral targeting. I further find that personalizing prices based on web-browsing histories increases profits by 12.99% and results in some consumers paying substantially more than others for the same product. Using only demographics to personalize prices raises profits by only 0.25%, suggesting the percent profit gain from personalized pricing has increased 50-fold.

*I would like to thank Kathryn Graddy, Shane Greenstein, Phil Leslie, Imke Reimers, Katja Seim, and Joel Waldfogel for very useful feedback, suggestions, and advice. I would also like to thank Erik Brynjolfsson, Iain Cockburn, Michael Grubb, George Hall, Jeremy Howard, Adam Jaffe, 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.

Electronic copy available at: https://ssrn.com/abstract=3503079
1 Introduction

Consumers are now tracked on the web by thousands of tracking and telecom companies.¹ These tracking data may be sold for the purpose of behaviorally targeted advertising and personalized pricing. In this paper, I investigate whether web-browsing data improve predictions of purchase propensities, and whether they increase the profits from personalized pricing compared to the scenario where only demographics are available for such targeting.

Historically, advertisers and sellers have had a limited ability to target advertisements and prices due to the nature of advertising and the high costs of acquiring data beyond basic demographics. For example, television advertisers would place ads on shows that catered primarily to particular demographic groups based on age, gender, and race, but individual viewers of a given show could not be separately targeted and shown different ads. Similarly, firms might offer discounts off the regular price to seniors, women, students, or other easily verifiable demographic groups. In recent years, however, robust markets for large datasets on individual behavior, popularly referred to as “big data,” have developed. These data can be used to form a hedonic estimate of individuals’ purchase propensities and reservation prices. Thus, on the Internet, ads and prices can be targeted to individuals, rather than targeted only to broad demographic groups.

Concerns related to data collection and targeting, such as privacy concerns or concerns about the equity of personalized pricing, are influencing policy. For example, the European Union recently introduced substantial data use regulations, the General Data Protection Regulation (GDPR), which requires websites to inform consumers of the specific uses of their data (e.g., personalizing prices) and to obtain consent. In the U.S., President Obama used his address to the Federal Trade Commission (FTC) in January 2015 to express his intent to introduce new legislation, a Consumer Privacy Bill of Rights, which would institute a framework

¹Economist (2014) notes that over 1300 firms are tracking consumers at the 100 most popular websites. Telecom companies may also sell browsing data: https://tinyurl.com/mfup9ph
for promoting transparent use of data that is limited by consumers’ consent. Excessive privacy invasions and personalized pricing were identified as two major motivating concerns in a White House Report that same month (Office of the President, 2015). Bergemann et al. (2015) validates these concerns by proving a wide range of potential impacts of improved price discrimination on consumer welfare, ranging from eliminating consumer surplus entirely, to passing the entire surplus gained from price discrimination to consumers. Intuitively, firms may extract a greater share of surplus when setting personalized prices close to reservation prices. But consumer surplus can rise if previously unserved customers are offered lower prices when prices are personalized. This paper empirically investigates the impacts of price personalization.

The most similar article on behavioral targeting using web-browsing data is Goldfarb and Tucker (2011), which investigates the impact of a law limiting web-tracking on stated purchase intentions among banner advertising viewers. They find a 65% reduction in advertising effectiveness. Using a revealed preference approach, I find evidence suggesting web-browsing histories are effective for identifying some consumers that are almost certainly willing to buy.

The prior literature on personalized pricing, beginning with Rossi et al. (1995, 1996) and Chintagunta et al. (2005), among many others, typically considers personalized pricing based on past purchase history of the same product. The basic idea is intuitive: if a consumer buys a product frequently, or was previously willing to pay a high price, they likely have a high reservation price for the product and can be charged higher prices in future interactions. One can thus use a small set of variables that are intuitively excellent predictors of willingness to pay to set prices. The models described in their papers were designed for contexts like pharmacies and

---

2https://tinyurl.com/y8w9nr7h. Subsequent proposed bills have not yet been enacted, as of October 2018.
3Robinson (1933) notes that price discrimination can increase aggregate surplus only if sales rise. If sales remain the same (or fall), then aggregate surplus falls because some goods are (inevitably) inefficiently allocated to consumers that have a lower willingness to pay than some unserved consumers offered higher prices.
4Others have noted that measuring online advertising effectiveness is inherently difficult (Lewis and Rao, 2015; Johnson et al., 2016).
5Hannak et al. (2014) and Mikians et al. (2012) found some evidence of personalized pricing online but did not study its effectiveness. Rossi et al. (1995, 1996) and Waldfogel (2014) do investigate personalized prices but do not use detailed web-browsing histories as explanatory variables.
6Personalized marketing, including pricing, is often referred to in the marketing literature as “customer addressability.”
7Acquisti and Varian (2005) and Fudenberg and Villas-Boas (2007) suggest that strategic consumers may
grocery stores, where their methods are widely used today to generate personalized coupons, a less conspicuous form of personalized pricing. A concurrent paper, Dube and Misra (2017), also incorporates machine learning techniques and validates counterfactual predictions of the profit gained from personalized pricing via a field experiment. Their paper is complementary, although it has a somewhat different focus: it investigates the impact of price discrimination in business-to-business transactions, uses a much smaller set of explanatory variables (excluding online browsing histories), and uses an estimation algorithm too burdensome to be applied to datasets with very large numbers of explanatory variables.\footnote{Dube and Misra (2017) show that profits are slightly higher when personalized prices are chosen to maximize profits accounting for the uncertainty in individualized parameter estimates (using Hierarchical Bayesian methods), due to Jensen’s inequality.} Lastly, recent papers have examined the impact of telematics-based monitoring on price discrimination in insurance markets (Reimers and Shiller, 2018) and digital disintermediation on negotiated prices (Peukert and Reimers, 2018).

This paper presents a method for determining purchase propensities and optimal individual-level prices from big observational datasets, and uses the method to estimate, in one context, individual purchase propensities and aggregate profits from personalized pricing if nearly 5,000 web-browsing variables are available. Of course, these analyses cannot, by themselves, prove that price discrimination has recently become more effective. Therefore, I compare these results to the analogous outcomes when only demographics — which have long been available — can be used for such targeting.

I employ a two-step estimation procedure. First, I use machine learning techniques to estimate individuals’ purchase probabilities at observed, non-personalized prices. Next, I estimate individual-specific utility parameters in a logit framework by matching the purchase propensities implied by the logit model to the purchase propensities implied by the machine learning model. Due to limitations of available data, not all parameters are identified without additional moment conditions, so I introduce a supply-side optimal pricing condition and an aggregate

\footnote{alter their buying behavior to mask their level of interest in the product in order to receive lower prices in later periods. If they do so, it may limit the explanatory power of past purchases.}
tier-choice moment condition, to address this issue.

The estimated individual-specific model parameters govern the relationship between prices and expected individual-level demand (e.g. the probability a particular consumer chooses each of Netflix’s products). The slope of expected demand reflects the level of statistical uncertainty when predicting a consumer’s choice. A sample held out from estimation is used to ensure that machine learning techniques avoid problems arising from overfitting, which could bias estimates of statistical uncertainty and thus also bias estimated effectiveness of targeting. I then calculate individual purchase propensities at observed prices, simulate counterfactual prices and profits under personalized pricing, and evaluate a possible regulatory intervention.

Netflix provides an auspicious context for study. First, analyzing this question requires individual-level data on both web-browsing histories and all purchases of a particular item, information that rarely appears together in datasets available to academic researchers. However, Netflix subscription status can easily be imputed from web-browsing histories, implying that all needed data reside in available online browsing datasets. Second, because purchases occur online, Netflix could implement personalized pricing based on web data. Third, because interactions with consumers take place online, Netflix could preempt consumer resentment by framing personalized pricing as automatically applied customized coupons — a strategy currently employed by some large online retailers — since coupons are already widely used and tolerated by consumers (Venkatesan and Farris, 2012).

Estimates from the first step are used to investigate the effectiveness of browsing data for targeting advertisements. I find that demographics alone yield predicted purchase probabilities that range across consumers from about 8% to about 30%. When web-browsing histories are used for prediction, 8%, or about half as many people as subscribed, have estimated purchase probabilities exceeding 30%, and often far exceeding 30%. These predictions are validated using 5,000 households held-out from the estimation sample. Web-browsing data are thus substantially more effective than demographics for finding consumers that are reasonably likely

---

9These data are, however, easily accessible by the firm.
to make a purchase. This suggests that targeted advertising can be much more effective than traditional forms of advertising, which, for example, placed ads in newspapers or on television shows based on the audience’s demographic composition.

Simulations further 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 0.25% higher if using demographics alone to personalize markups, but 12.99% higher if using all data. This finding suggests web-browsing data make personalized pricing substantially more appealing to firms.

The remainder of the paper is organized as follows. Section 2 describes the context and industry background. Section 3 describes the data. Section 4 presents the model, and Section 5 provides estimation details. The main results of the paper are then presented in Section 6. A brief conclusion then discusses concerns over perceived unfairness of personalized prices, and explains how firms have circumvented these concerns and begun personalizing prices.

2 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.\(^\text{10}\)

Netflix was differentiated from competing alternatives in at least three important ways. First, consumers’ main alternative was to travel to a brick-and-mortar store (incurring a travel cost), rather than having movies arrive at their residences.\(^\text{11}\) Second, by operating a few large

\(^{10}\)The 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 \times 114.384] \times 100 \approx 16.7.

\(^{11}\)Blockbuster did offer a DVD-by-mail service starting in 2004 (https://tinyurl.com/yc9fmwpx). But the service was marginalized because it competed with the core business that had invested in local brick-and-mortar outlets. The program only grew after incorporating in-store exchanges starting in November 2006. Subscriptions increased quickly, reaching 2 million in total by January 2007 (Netflix, 2006).
warehouses outside of cities on cheaper land, Netflix was able to stock a much larger variety of videos than local brick-and-mortar rental outlets. Third, Netflix developed a well-regarded recommendation algorithm, which reduced consumers’ search costs on the platform. A recent study (Gomez-Uribe and Hunt, 2016) suggests that 80% of viewing choices are attributable to the recommendation algorithm, whereas only 20% of viewing is attributable to consumer search. Given these advantages, Netflix is expected to have had some pricing power, at least during the period studied.

Netflix’s subscription 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.12 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 unlimited plans were estimated using industry statistics and expert guidance. They are assumed to equal $6.28 for the one DVD at-a-time plan, $9.43 for the two DVDs at-a-time plan, and $11.32 for the three DVDs at-a-time plan.13

---

12 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.

13 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 $x \cdot P_j$, 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.
3 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 of pages viewed on that website, the referring website, and details on any transactions.\textsuperscript{14} 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, I assume that a computer user is a subscriber if and only if he or she \textit{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 queues, rate movies, and so forth. Non-subscribers cannot access as many pages because 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 2. This small difference may be attributed to approximation errors in this latter estimate, and ComScore’s sampling methods.

For a small subset (a few hundred) of consumers, their Netflix transactions are recorded in the ComScore data, and hence their choice of Netflix tier (one, two, or three DVDs at a time) is observed. I use their tier choices to infer the aggregate share choosing each Netflix tier, assuming this group is representative of the larger population of Netflix subscribers. Specifically, the aggregate share choosing each tier is inferred by multiplying the fraction of consumers choosing any of Netflix’s offerings (in the entire sample) by the fraction of consumers selecting each tier among the sample of consumers whose specific tier choice is recorded (in the small sample of consumers observed purchasing Netflix).

\textsuperscript{14}In our correspondence, ComScore representatives stated that demographics were captured for individual household members as they complete “a detailed opt-in process to participate,” for which they were incentivized.
consumers, who all subscribed).

I derived several web behavior variables from the raw data. These include the percent of 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 is broken into five 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).

I then cleaned the data by removing websites associated with malware, third-party cookies, video rental chains, and pornography, leaving 4,600 popular websites to calculate additional variables. 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 the 61,312 computer users. 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, census region (North, South, East, West), and area and population density of their ZIP code tabulation area from the 2010 Census. Variables classified as basic web behavior include the percent of online browsing by time of day and by day of week, and a broadband indicator. Summary statistics for the demographics and basic browsing variables are shown in Table 1. Note the substantial variation in demographics that could be used to personalize prices.

Variables classified as detailed web-behavior indicate number of visits to a particular web-

---

15yoyo.org provides a user-supplied list of some websites of dubious nature. Merging this list with the ComScore data reveals 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 that these rules were very effective in screening out dubious websites. In addition, video rental websites were dropped.

16Pornography might contain valuable information, but might also require listing perverse website names in publication.

17ComScore’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.
site, with one variable for each of the 4,600 websites, as well as the household’s total website visits and transactions on the web during 2006. The yearly frequency of visits to individual websites, averaged across consumers, are shown in Figure 1. The most popular website in 2006, msn.com, was visited about 200 times a year by the average person, slightly more than once every other day. But visits decline quickly with rank. The 50th ranked website (usps.com) was visited by the average customer about 2.2 times per year, partly because patrons visit it infrequently, and partly because 61% of households in the sample never visited the site in 2006. Subsequent sections analyze whether such variations in behaviors across consumers, presumably indicative of preferences and habits, are useful for segmenting consumers and personalizing prices.

Prior to estimation, each explanatory variable is normalized to have a mean of zero and a standard deviation of one. This is common when using machine learning techniques that employ regularization.

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 sample, is used for estimating model parameters. The model’s fit is then tested using the sample of 5,000 consumers held out from estimation.

4 Model and Identification

This section describes the strategy for estimating individual-level expected demand for Netflix’s vertically differentiated products. I first present a demand-side model, which can be used alone for estimation if one has ideal data (i.e., individual-level purchase data [including tier choice] at experimentally varied prices, along with individuals’ web-browsing histories). With such data, one can first estimate the relationship between purchase probabilities (for each tier) and browsing histories using machine learning techniques, separately for multiple experimental price sets, and then find the individual-consumer model-parameters that match the implied
purchase probabilities from the model with machine learning predictions.\footnote{18If all price variation is attributed to random experimentation, endogeneity concerns are eliminated.}

Note, however, that the ideal data are difficult for academic researchers to obtain (although the data are reasonably easy for firms to obtain). Furthermore, without an explicit pricing experiment, price endogeneity is an important concern, one difficult to address when using machine learning techniques. However, the chosen context (Netflix in 2006) mitigates such concerns, because prices were sticky and did not vary. But, without price variation, price sensitivities are not identified from the demand-side model alone. Furthermore, without data on each individual’s choice of quality tier, individual-specific marginal utilities for higher tiers are not identified.

To address these issues, I introduce a supply-side model and aggregate tier-choice moment conditions (and model restrictions) that are used to augment the demand-side model when the data are less than ideal (e.g., when the choices of specific inside tiers are not observed and prices are sticky and do not vary). The augmented estimation model assumes that the researcher has ex-ante knowledge of marginal costs. The supply-side model yields an optimal-pricing first-order condition used to estimate consumers’ mean price sensitivity (assuming marginal costs are known). The augmented model also assumes that consumers agree on the marginal intrinsic utilities provided by higher quality tiers. Aggregate tier choice shares are then used to estimate the mean marginal intrinsic utilities for higher quality tiers when individual-level tier choice is not consistently observed. Details are below.

\section{4.1 Demand}

The aim of the model is to estimate the relationship between prices and expected demand (i.e., probability of purchase) at the individual-consumer level. As in McManus (2008), a heterogeneous agent logit framework is used to model demand for vertically differentiated
Specifically, consumer $i$'s conditional indirect utility for product $j$ equals

$$u_{ij} = \alpha_i P_j + \xi_{ij} + \sigma \epsilon_{ij} = \bar{u}_{ij} + \sigma \epsilon_{ij},$$

(1)

where $P_j$ denotes product $j$’s price, and $\alpha_i$ and $\xi_{ij}$ denote individual $i$’s price sensitivity and intrinsic valuation for product $j$, respectively. $\epsilon_{ij}$ is an iid error term following the type 1 extreme value distribution. Lastly, $\sigma \epsilon$ denotes the standard deviation of the error term. The mean utility of the outside good is normalized to zero: $u_{i0} = 0 + \sigma \epsilon_{i0}$.

In the model, the parameters $\xi_{ij}$ and $\alpha_i$ represent estimable components of consumer $i$’s preferences. Together, they determine consumer $i$’s expected reservation price, based on information available to the firm. The error term ($\epsilon_{ij}$) reflects remaining uncertainty. As the error’s scale ($\sigma \epsilon$) declines, prediction of which product provides highest utility (product choice) improves (i.e., product choice probabilities move closer to the extremes [0,1]). Analogously, as the scale of all remaining parameters ($\xi_{ij}, \alpha_i$) increases, holding error scale ($\sigma \epsilon$) fixed, predictions of consumers’ choices improve. In practice, the latter (scale of remaining parameters [$\xi_{ij}, \alpha_i$]) determines predictive ability because the error scale ($\sigma \epsilon$) is typically normalized (to one) for the model to be identified. Following convention, I subsequently assume $\sigma \epsilon = 1$. Hence, if machine learning predicted-choice probabilities are near the extremes, zero and one, then the model will reflect these strong predictions by estimating a large scale for observable components ($\alpha_i, \xi_{ij}$).

This would imply that the components of an individual consumer’s demand inferred from their

---

19 Similar models are often referred to as “random coefficient” models. However, coefficients are not random in this context; the model explicitly estimates separate coefficients for each observed individual.

20 I assume $\epsilon$ follows the type 1 extreme value distribution with location parameter equal to the negative of Euler’s constant and scale parameter equal to one.

21 Suppose that on estimable parameters (excluding error $\epsilon$), product $j$ is preferred to product $k$ (i.e., $\alpha_i P_j + \xi_{ij} > \alpha_i P_k + \xi_{ik}$). Product $j$ is chosen if $\alpha_i P_j + \xi_{ij} + \epsilon_{ij} > \alpha_i P_k + \xi_{ik} + \epsilon_{ik}$, which can be arranged to yield $\epsilon_{ik} < \alpha_i P_j + \xi_{ij} - \alpha_i P_k - \xi_{ik} + \epsilon_{ij}$. Thus, the probability of product $j$ being chosen is $F(\alpha_i P_j + \xi_{ij} - \alpha_i P_k - \xi_{ik} + \epsilon_{ij})$, where $F()$ denotes the cdf of error term $\epsilon$. Increasing the scale of estimable parameters $\alpha$ and $\xi$ by multiplying them by the same constant ($>1$) increases the positive difference between $(\alpha_i P_j + \xi_{ij})$ and $(\alpha_i P_k + \xi_{ik})$, which increases $F(\alpha_i P_j + \xi_{ij} - \alpha_i P_k - \xi_{ik} + \epsilon_{ij})$ (i.e., probability product $j$ is chosen) and decreases the probability $k$ is chosen. Hence, increasing the scale of estimable parameters $\alpha$ and $\xi$ raises the probability of choosing the product providing higher utility based on estimable measures towards one, and reduces the probability that the other product(s) are chosen towards zero, implying better predictions.
browsing history are large compared to the unobserved components of their demand, implying that precise targeting is feasible.

The probability consumer \( i \) chooses product \( j \) is

\[
s_{ij}(\alpha_i, \xi_i, P) = \frac{\exp(\alpha_i P_j + \xi_{ij})}{1 + \sum_{k \in J} \exp(\alpha_i P_k + \xi_{ik})},
\]

(2)

where \( J \) denotes the set of inside products, and \( \xi_i \) and \( P \) denote the product-tier vectors \([\xi_{i1}, \xi_{i2}, \xi_{i3}]\) and \([P_1, P_2, P_3]\), respectively.

Correspondingly, the probability that consumer \( i \) chooses one of the inside products (Netflix’s products), as opposed to the outside good (no Netflix product), is

\[
s_{ij \neq 0}(\alpha_i, \xi_i, P) = 1 - s_{i0}(\alpha_i, \xi_i, P) = 1 - \frac{1}{1 + \sum_{k \in J} \exp(\alpha_i P_k + \xi_{ik})},
\]

(3)

The demand model is used to construct two sets of moment conditions: (1) machine learning estimated probabilities each individual \( i \) subscribes less the corresponding logit model’s predictions \((\hat{s}_{ij \neq 0}(X_i) - s_{ij \neq 0}(\alpha_i, \xi_i, P))\) and (2) aggregate product tier shares less model predictions \((\hat{s}_j - s_j\), where \( s_j = \int s_{ij}(\alpha_i, \xi_i, P) f(\alpha_i, \xi_i) d\alpha_i d\xi_i\). Note that if one had ideal data then one could instead use only one set of moment conditions, machine learning predicted probabilities of subscribing to each tier less logit model predicted probabilities \((\hat{s}_{ij}(X_i) - s_{ij}(\alpha_i, \xi_i, P))\).

4.2 Supply

The supply-side model closely follows supply-side models in typical random coefficient models, with one important difference. Typically, price sensitivity is identified from the demand-side, and the supply-side optimal-pricing first-order condition is used to recover estimates of marginal costs. But because prices are sticky in the studied context, mean price sensitivity is not identified from the demand-side. Instead, ex-ante information on marginal cost is assumed.
and $11 \theta$ annual report implies

\[
\pi = \sum_{k \in J} (P_k(\theta) - c_k) M s_k - \Gamma = \sum_{k \in J} \theta c_k M s_k - \Gamma, \tag{4}
\]

where $M$ is the mass of consumers, $\Gamma$ is the fixed cost, and $s_j$ is the aggregate share selecting tier $j$.

The first-order condition yields

\[
d\pi = \sum_{k \in J} c_k s_k + \theta \sum_{k \in J} c_k d s_k = 0, \tag{5}
\]

where $d s_j$ equals

\[
d s_j = \int d s_{ij}(\alpha_i, \xi_i, P(\theta)) \frac{d \alpha_i d \xi_i}{d \theta}, \tag{6}
\]

where $P(\theta)$ denotes the price vector as a function of markup ($P(\theta) = (1 + \theta) \times [c_1, c_2, c_3]$), $f(\alpha_i, \xi_i)$ denotes the density of the individual-specific parameters, and $d s_{ij}(\alpha_i, \xi_i, P(\theta))$ follows the logit-model framework.  

---

22 See Section 2 for details.

23 $s_{ij}(\alpha_i, \xi_i, P(\theta)) = \frac{e^{\alpha_i (1 + \theta) c_j + \xi_j}}{1 + \sum_{k \in J} e^{\alpha_k (1 + \theta) c_k + \xi_k}}$.  Hence, $d s_{ij}(\alpha_i, \xi_i, P(\theta)) = \alpha_i c_j s_{ij}(\alpha_i, \xi_i, P(\theta)) - s_{ij}(\alpha_i, \xi_i, P(\theta)) \left( \sum_{k \in J} \alpha_i c_k s_{ik}(\alpha_i, \xi_i, P(\theta)) \right)$.  

$\frac{d s_{ij}(\alpha_i, \xi_i, P(\theta))}{d \theta} = \frac{\alpha_i c_j s_{ij}(\alpha_i, \xi_i, P(\theta)) \left( 1 + \sum_{k \in J} e^{\alpha_k (1 + \theta) c_k + \xi_k} \right)}{(1 + \sum_{k \in J} e^{\alpha_k (1 + \theta) c_k + \xi_k})} = \frac{\alpha_i c_k e^{\alpha_i (1 + \theta) c_k + \xi_k}}{(1 + \sum_{k \in J} e^{\alpha_k (1 + \theta) c_k + \xi_k})}$.  

$\frac{d s_{ij}(\alpha_i, \xi_i, P(\theta))}{d \theta} = \alpha_i c_j s_{ij}(\alpha_i, \xi_i, P(\theta)) \left( 1 + \sum_{k \in J} e^{\alpha_k (1 + \theta) c_k + \xi_k} \right) = \alpha_i c_j s_{ij}(\alpha_i, \xi_i, P(\theta)) \left( 1 + \sum_{k \in J} e^{\alpha_k (1 + \theta) c_k + \xi_k} \right)$.  

$\frac{d s_{ij}(\alpha_i, \xi_i, P(\theta))}{d \theta} = \alpha_i c_j s_{ij}(\alpha_i, \xi_i, P(\theta)) \left( 1 + \sum_{k \in J} e^{\alpha_k (1 + \theta) c_k + \xi_k} \right)$.
\[
\frac{ds_{ij} (\alpha_i, \xi_i, P(\theta))}{d\theta} = \alpha_i s_{ij} (\alpha_i, \xi_i, P(\theta)) \left( c_j - \sum_{k \in J} c_k s_{ik} (\alpha_i, \xi_i, P(\theta)) \right).
\] (7)

The third moment condition used in estimation is the derivative of profits with respect to markup \(\frac{d\pi}{d\theta}\), from Equation 5. It should equal zero, assuming that the firm is choosing a markup to maximize profits.

### 4.3 Moment Conditions

To recap, the aim is to estimate individual-specific preference parameters \((\alpha, \xi)\) for every consumer. There are three sets of moment conditions used in estimation in the typical case where the data are less than ideal.

\[
G (\alpha, \xi; \hat{s}_{ij \neq 0} (X_i), \hat{s}_j, c_j, \theta) = \begin{bmatrix} \hat{s}_{ij \neq 0} (X_i) - s_{ij \neq 0} (\alpha_i, \xi_i, P(\theta)) \\ \hat{s}_j - \int s_{ij} (\alpha_i, \xi_i, P(\theta)) f (\alpha_i, \xi_i) d\alpha_i d\xi_i \\ \frac{d\pi}{d\theta} \end{bmatrix}
\] (8)

The first set of moment conditions is the difference between machine learning estimates of individuals’ probabilities of subscribing to any tier of Netflix’s products less the corresponding logit-model predictions. The second is the difference between the aggregate share selecting tier \(j\) (\(\hat{s}_j\)) and the corresponding model prediction (\(s_j\)), where \(s_j = \int s_{ij} (\alpha_i, \xi_i, P(\theta)) f (\alpha_i, \xi_i) d\alpha_i d\xi_i\).

The third moment condition is the derivative of profits with respect to markup. Estimated individual-specific demand parameters \((\alpha_i, \xi_i)\) minimize \(G'G\).

If one has the ideal data (i.e., complete data generated from a randomized pricing experiment), then one can instead use a single set of moment conditions. In that case, individual-specific preference parameters \((\alpha, \xi)\) can be estimated by minimizing the difference between machine learning predicted probabilities of purchase of each tier and the corresponding model predictions (\(\hat{s}_{ij} (X_i) - s_{ij} (\alpha_i, \xi_i, P))\).
4.4 Identification and Restrictions

I begin by explaining how all parameters are identified with ideal data. Ideal data include separate estimates of the probabilities that each consumer chooses each tier of service \( s_{ij}(X_i) \) under at least two different price menus from a pricing experiment. I then explain restrictions and additional moment conditions needed to estimate parameters with the data at hand, which are not as rich as the ideal data.

First note that estimates of a consumer’s probabilities of choosing each tier \( \hat{s}_{ij}(X_i) \) under a single price schedule are not sufficient on their own to identify his or her preference parameters \((\alpha_i, \xi_i)\). If one both (i) reduces the price sensitivity \( \alpha_i \) by constant \( A \), and (ii) adds a constant \( A \times P_j \) to consumer \( i \)’s intrinsic value for each product \( j \) \( \xi_{ij} \), then the mean utility for each tier \( j \) (and hence choice probability) remains unchanged.\(^{24}\) Hence, an infinite combination of \( \alpha_i \) and \( \xi_{ij} \) can match logit-predicted probabilities \( s_{ij}(\alpha_i, \xi_i, P) \) to machine learning predictions \( \hat{s}_{ij}(X_i) \).

However, additional information on demand elasticities identifies all parameters. Note, if holding choice probabilities \( s_{ij}(\alpha_i, \xi_i, P) \) fixed at machine learning predictions (i.e., holding fixed conditional indirect utility by adjusting \( \xi_i \) to compensate for any changes in \( \alpha_i \)) then demand elasticities depend on only the price sensitivity \( \alpha_i \). For example, in the logit model, an individual’s own price elasticity equals \( \left( \frac{ds_{ij}(\alpha_i, \xi_i, P)}{dP_j} \right) \frac{P}{s_{ij}(\alpha_i, \xi_i, P)} = \alpha_i P_{ij} (1 - s_{ij}(\alpha_i, \xi_i, P)) \), implying a monotonic relationship between \( \alpha_i \) and demand elasticity. Thus, only one set of \( \alpha_i \) and \( \xi_i \) can rationalize both the choice probabilities \( \hat{s}_{ij}(X_i) \) and elasticities \( \left( \frac{ds_{ij}(\alpha_i, \xi_i, P)}{dP_j} \right) \frac{P}{s_{ij}(\alpha_i, \xi_i, P)} \).

Any added information on choice elasticities, such as choice probability estimates \( \hat{s}_{ij}(X_i) \) under multiple price schedules, or an assumption that individualized-prices are optimally set (and marginal costs are known ex-ante), allows identification of all parameters.\(^{25}\)

\(^{24}\)One could run a pricing experiment, using two price menus, \( P^L_j \) and \( P^H_j \), then relate browsing histories to choice probabilities separately for each of the two price menus. Then, for each individual, one could predict the probability a consumer with a given web-browsing history buys each product \( j \) under price menu \( P^L_j \), and the corresponding probability that consumer buys each product \( j \) under price menu \( P^H_j \). Individual-level logit parameters \((\alpha_i, \xi_{ij})\) can then be found to match. Alternatively, without a pricing experiment, a researcher could...
In this context, however, the data are not ideal. First, because individual-level tier choice is not observed, there is not enough information to separately identify $\xi_{ij}$ for each individual $i$ and tier $j$. To address this issue, an identifying restriction is imposed: all consumers agree on the marginal intrinsic utilities for higher tiers ($\xi_{ij} = \xi_{i1} + \xi_j$, for $j > 1$). With this restriction, $\xi_j$ are identified by the second moment condition ($\hat{s}_j - s_j = 0$), which matches observed aggregate tier shares ($\hat{s}_j$) and model predictions ($s_j = \int s_{ij} (\alpha_i, \xi_i, P(\theta)) f (\alpha_i, \xi_i) d\alpha_i d\xi_i$), because the predicted share choosing a given tier $j$ ($s_j$) is monotonically increasing in $\xi_j$

Second, there is no information on individual $i$’s demand elasticity, because prices did not vary and were not individually set. Hence, individual-level price sensitivities ($\alpha_i$) and intrinsic values for the first tier ($\xi_{i1}$) are not separately identified. One can address this issue by assuming a representative price sensitivity ($\alpha$) and by using information from the supply side. Assume that for any change in $\alpha$ the parameters $\xi_{ij}$ are adjusted accordingly so that subscription probabilities are kept the same, and the first moment condition remains satisfied. Then, larger negative values of $\alpha$ imply more elastic demand and thus lower optimal markups (over ex-ante known marginal costs). Only one value of $\alpha$ implies observed markups are optimally set, satisfying the third moment condition ($d\pi/d\theta = 0$).

### 4.5 Estimation Routine

I search for parameter values minimizing the sum of squared moments, $G'G$. To speed computation, I search over the representative price sensitivity ($\alpha$) and marginal utilities for higher tiers ($\xi_j$, for $j > 1$), nesting within a search for the individual-level parameters ($\xi_{i1}$) equalizing the logit-model predicted probabilities of selecting an inside product ($s_{ij \neq 0}(\alpha, \xi_i, P)$) and the corresponding machine learning estimates ($\hat{s}_{ij \neq 0}(X_i)$). 

\[ \xi_{i1} = \ln \left( \frac{\hat{s}_{ij \neq 0}}{1 - \hat{s}_{ij \neq 0}} \frac{1}{\sum_{k \neq j \neq 0} \exp(\alpha_P + \xi_{i1} + \xi_{k1})} \right) \]
5 Estimation

The model is estimated in two steps. The first step — the main focus of this section — estimates purchase probabilities for individual consumers. In the second step, remaining model parameters are estimated as described in the preceding section.

A logistic regression with LASSO regularization is used to estimate the probability each consumer subscribes to any one of Netflix’s services (one, two or three DVDs at a time), as a function of individual-level observables (e.g., browsing data). The penalized likelihood function equals

\[ L = \sum_i \log \left( s_{ij \neq 0}(X_i) \times I(buy) + (1 - s_{ij \neq 0}(X_i)) \times (1 - I(buy)) \right) - \lambda \sum \text{abs} \left( \beta \right), \]  

(9)

where \( s_{ij \neq 0}(X_i) \) denotes the model-predicted probability of subscribing, \( I(buy) \) indicates whether consumer \( i \) subscribes, \( \lambda \) is the LASSO penalty parameter, \( \sum \text{abs} \left( \beta \right) \) is the absolute sum of coefficient values, and

\[ s_{ij \neq 0}(X_i) = \frac{\exp(\phi + X_i \beta)}{1 + \exp(\phi + X_i \beta)}. \]  

(10)

The estimation routine searches for parameters \((\phi, \beta)\) to maximize the penalized likelihood, for a given penalty parameter. The penalty parameter \((\lambda)\) is estimated by maximizing the out-of-sample likelihood using two-fold cross-validation.

A set of 4,633 explanatory variables are considered. In all models, 18 demographic variables are considered, including (i) indicators for children, race, Hispanic ethnicity, census region (North, West, and South); (ii) ordinal group number (e.g. 18- to 20-year-olds are considered group number 1), the group number squared, and indicator for censored from above, for each of the following: age, income, and household-size groupings; and (iii) linear measures of the area.
and population density of the household’s ZIP code tabulation area. Additionally, some models include the remaining 4,615 explanatory variables summarizing individuals’ web browsing, including indicators for the browsing habits listed in Table 1, the intensity of web use (number site visits), and its square, and the number of visits at each of 4,600 websites. All explanatory variables are normalized to have mean 0 and standard deviation of 1, prior to estimation.

A useful feature of LASSO models is that the procedure selects potentially meaningful explanatory variables, setting coefficients on other variables to zero. A large number, 938 of the 4,633 considered explanatory variables, remain and have nonzero coefficients in the full model. Of these, only 5 are demographic variables, suggesting that the web-browsing data provide richer information on consumers’ tastes.

The top 30 variables, ordered by coefficient magnitude, are reported in Table 2. Because all variables are normalized prior to estimation, coefficient size provides a measure of the influence of the variable on estimated subscription probabilities. Note that the 30 most influential explanatory variables do not include any demographic variables, suggesting demographic variables provide little marginal information about subscription probabilities. Also note that the list does not include any basic browsing behaviors like timing or intensity of web use. Rather, a household’s tendency to visit particular websites seems to contain the most information about their affinity for Netflix.

The intuition linking website visits to an innate affinity for Netflix is obvious for some. GameFly (#1), Audible (#4), and Amazon (#30) — all positive predictors of Netflix subscription — indicate a preference for receiving products by mail, suggesting a higher valuation for Netflix, which at that time delivered DVDs by mail. 4chan.org (#7) is a site for anime enthusiasts interested in content that is typically available at Netflix — which operates large warehouses — but may not be available at the local brick-and-mortar outlets comprising Netflix’s competition in 2006. The top sites also include a site catering to consumers interested in

---

28 Testing revealed that regularization methods yielded a somewhat better out-of-sample fit, compared with stepwise models commonly used in the economics literature in the past.
technology startups (#26, techdirt) and a movie review website (#9, imdb.com). To address the concern that movie review websites might be complements for Netflix, raising concerns of reverse causality, the model is re-estimated, excluding all websites categorized as related to “movies” or “TV” (including imdb.com, DVDfab.com, and slysoft.com), according to Alexa web rankings. These results are reported in the robustness subsection.

The impacts of the most influential variables are large. The last column of Table 2 shows the corresponding marginal change in subscription probability, on average across consumers, occurring when visits to the website increase by one standard deviation. For example, users who visit GameFly.com one standard deviation more often than average are 11.8 percentage points more likely to subscribe to Netflix at observed prices, implying a 75% (11.8/16 × 100) higher probability of subscribing than the mean consumer, whose subscription probability is about 16%.

Figure 2 offers a check on the model’s out-of-sample fit. Individuals in the holdout sample are ordered according to the predicted probability that they subscribe to Netflix, according to the model, then split into 600 groups. The average predicted probability and observed probabilities (i.e., fraction buying) are then calculated for each group. Figure 2 shows that the predicted probabilities, shown in the solid blue line, do in fact seem to follow the actual probabilities of subscription, even in the sample of consumers held out from estimation entirely.

The logit demand model parameters ($\xi_i, \xi_j, \alpha$) are then estimated via methods of moments, as described in Section 4. The estimated parameters imply demand elasticities for the three quality tiers (1–3 DVDs at a time) of -2.1, -3.4, and -3.9, respectively.

In some contexts, a subset of sites’ visits could correlate with purchase because they are complements, not because they are a priori useful for segmenting consumers. When designing a pricing experiment, practitioners should use caution, and if possible, utilize browsing data from a period prior to the consumers’ purchase decisions.
6 Results

6.1 Efficacy of Targeting

One can analyze how well various data types identify likely consumers using only the first stage of the model, which predicts the probability a consumer subscribes from a large set of variables without relying on the assumptions of the structural model. Several tests yield the same conclusion: web-browsing data yield superior predictions of consumers’ choices.

Consider, as a measure, the log likelihood of consumers’ observed choices in a sample of 5,000 consumers entirely held out of estimation. The log likelihood with no information (i.e., assuming all consumers have the same mean probability of subscribing) equals -2,161.5. Including demographics raises the likelihood to -2148.4, a change of only 13.1. Including browsing data raises the likelihood more, to -1,989.3, a much larger increase of 172.2.

The effectiveness of various datasets for prediction can be also be illustrated. Figure 3 shows the rank-ordered probabilities individual consumers subscribe to Netflix at observed prices, separately for the cases where (1) demographics are used for prediction, and (2) all variables are used. Note from this figure that the predicted probabilities of subscribing, based on demographic data, range from 7.8% to 30.2%. No consumer, based on demographics alone, has a purchase probability exceeding 30.2%. By contrast, when browsing data are used for prediction, more than 8% of customers are identified as having purchase probabilities exceeding 30.2%. 8% of consumers is a relatively large share, considering only about 15% of consumers subscribe to Netflix. Furthermore, as Figure 3 shows, many of these consumers have predicted purchase probabilities far exceeding 30%, some close to 90%. Traditional forms of targeting (i.e., placing advertisements next to content that attracts particular demographic profiles) therefore appear far less effective than online advertisements that utilize a user’s browsing data.

The demographic data employed in this paper include many of the demographic variables
an advertiser might use for traditional forms of ad targeting. However, one key variable —
local geography, widely used to target advertisements to local markets — was too sparse in
the available data to fully utilize in the model. 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 whether the browsing data are capturing the same information that geography would
if added to other demographic predictors, I investigate whether consumers in the same ZIP
code are measurably more similar in their preferences than are consumers across ZIP codes.
Specifically, I compare the distribution of differences in predicted subscription probabilities
(using the full model) for pairs of individuals in the same ZIP code with the corresponding
distribution for randomly drawn pairs living in different ZIP codes. If ZIP codes were a good
proxy for propensity to purchase, one would expect consumers in the same ZIP code to have
much more similar purchase propensities than consumers across ZIP codes. Figure 4, which
plots overlaid histograms of absolute differences in predicted probabilities, shows there is no
meaningful difference in these groups’ propensities to consume Netflix.\footnote{While not
meaningful, the difference is statistically significant, according to the Kolmogorov-Smirnov test.}
Hence, web-browsing
data offer mostly distinct information from that contained by geography.

6.2 Personalized Pricing Counterfactuals

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 using all
variables to explain a consumer’s willingness to pay.\footnote{In principal, consumers’ observed choices provide information on the ranges of \( \epsilon_{ij} \) consistent with their choices. However, because the goal is to set prices to new consumers before their choices are observed, rather than conditioning price on past purchase history, this information is disregarded.} They are then compared with simulated
profits under the status quo environment, where Netflix employed constant percent-markup
Table 3 shows the percent increase in profits from personalizing markups \((\theta_i)\), where percent markups differ across individuals, but for a given individual are the same across products.\(^{32}\) When all variables are used to personalize markups, profits increase 12.99%.\(^{33}\) If personalized markups are based only on demographics, the increase in total profits is much less, 0.25%. The nascent availability of these data thus increases the likelihood that firms will implement price discrimination, because adding web-browsing data substantially increases the amount by which personalizing markups raises profits.

Using the full set of variables to personalize markups substantially increases the range of prices charged to different individuals for the same product. Figure 5 shows a histogram of percent changes in markups when markups are personalized. When all variables are used, the 99.9th percentile markup is about 48% higher than markup set when prices are not personalized. The 99th percentile personalized markup is about 16% higher, and the 90th percentile markup is about 2% higher. The 75th percentile consumer receives approximately a 2% discount, which is not surprising, because less than a quarter of individuals \((\approx 16\%)\) subscribed under status quo prices. The median consumer would receive a 4% discount, and the lowest price is nearly 7% lower. Figure 5 shows that when only demographics are used, the variation across consumers in prices offered is much less pronounced.

Personalizing markups reduces aggregate consumer surplus by 0.50%.\(^{34}\) However, most consumers receive lower prices when prices are personalized, and hence the majority are slightly

---

\(^{32}\)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.

\(^{34}\)Percent change in consumer surplus equals \(\frac{CS_{\text{personalized}} - CS_{\text{status quo}}}{CS_{\text{status quo}}} \times 100\), where under the logit modeling assumptions, consumer surplus for a given pricing scheme equals \(\sum_i \ln(1 + \sum_{k \in J} \exp(\alpha P_k + \xi_{ik}))\).
better off.

Next, as a comparison, I consider second-degree price discrimination (PD), where all consumers face the same prices, but there is a separate markup for each product tier \((\theta_j)\). Suppose that when a consumer arrives, preferences are private (i.e., the firm has no information on individual-specific preferences). Under second-degree PD, the markups for each tier are designed so that consumers self-select the tier with an incremental price about equal to their willingness to pay for the incremental quality.\(^{35}\) Despite the fact that all consumers face the same price schedule, consumers end up paying different prices per unit of quality and the firm’s expected profits earned from a randomly arriving consumer increase.

Personalization and second-degree PD are not mutually exclusive, and combining the strategies by setting a separate markup \((\theta_{ij})\) for every combination of consumer and product tier may increase profits further. When a consumer arrives, the firm may have some information on his or her preferences (e.g., based on their web-browsing history). But a consumer’s exact preferences remain private information. Based on available information and the remaining uncertainty, the firm can form an expected distribution of parameters governing a consumer’s preferences. From the firm’s perspective, an arriving consumer represents a group of heterogeneous consumers with private preferences, from which the arriving consumer’s actual realized preferences are drawn. Just as classic second-degree PD can increase expected profits from an anonymous consumer, combining personalization with the mechanics of second-degree PD can raise profits when consumers retain some private information about their preferences.

The results are shown in Table 4. Changing from the status quo case (constant percent markup over cost, \(\theta\)) to second-degree PD (separate markup \(\theta_j\) for each tier) increases profits by 22.48%. Switching instead to personalized markups \((\theta_i)\) raises profits by 12.99%. Combining the strategies by setting a separate markup \((\theta_{ij})\) for each combination of consumer \(i\) and tier \(j\) raises profits by 36.8%.

\(^{35}\)In classic second-degree PD theoretical models, if the firm chooses qualities for an arbitrarily large set of vertically differentiated goods, each consumer pays exactly their full willingness to pay for marginal quality. With a small discrete set of goods, some consumers may pay less than their full valuation for marginal quality.
Second-degree PD raises profits by raising the price of the lowest quality tier ($\approx 18\%$) and reducing the price of the highest tier ($\approx 6\%$). Increasing the price of the lowest tier reduces profits from that tier, but reduces the price differences between higher tiers and the lowest tier and thus loosens the incentive compatibility constraints. More consumers are willing to pay the resulting smaller incremental price to acquire a higher quality tier, shifting some consumers to higher tiers that have larger absolute markups. Overall, profits increase. Similar patterns occur when personalizing separate markups for each tier (personalized second-degree PD).

6.3 Regulations

In principle, price regulations could yield personalized prices that raise both profits and aggregate consumer surplus, compared with the status quo case, a uniform markup charged to all consumers. This can be demonstrated with a simple example. Suppose personalized markups are offered to consumers whose optimal personalized markup is below the flat markup observed, and all other consumers continue to receive the same markup as before. No consumers are offered higher prices. Thus, no one is worse off. But consumers offered a lower price are better off than before, so aggregate consumer surplus rises. Personalizing prices to a subset of consumers, as opposed to no consumers, increases the firm’s profit as well.

In practice, however, mutually beneficial regulations may be elusive. In a world with price discrimination, regulators may not be apprised of the uniform pricing markup, or consumer demand in general. Hence, the simplest regulation — explicitly setting a price/markup ceiling — may be challenging to implement well.

Instead, a more feasible price regulation is considered. The regulation will be referred to as a discount penetration regulation, as it limits the percentage of consumers receiving a discount off a list price chosen by the firm. If the discount penetration ceiling is set to 100%

---

36 The derivative of variable profit from tier $j$ with respect to its price equals: $\int s_j + \alpha s_{ij}(1 - s_{ij})(P_j - c_j)df_i$. It is negative for the first tier. The derivative of the profit for tier $j$, for $j > 1$, respect to the price of the first tier is $-\int s_{ij}s_k(P_k - c_k)df_i$, which is weakly positive because $\alpha < 0$. 

Electronic copy available at: https://ssrn.com/abstract=3503079
(no regulation), then the firm sets a high list price and offers each consumer a personalized discount. This is equivalent to fully personalized pricing. At the other extreme, a discount penetration ceiling of 0% prohibits the firm from offering a discount to any consumer and is therefore equivalent to banning price discrimination. If regulators set the discount penetration rate somewhere in between, at, say, 85%, then the firm could offer as many as 85% of consumers a discount off the list prices.

The impacts of price regulations capping the discount penetration at n% are simulated. The simulation consists of an inner step and an outer loop. The outer loop performs a grid search over possible markup ceilings (list prices) to find the markup ceiling that maximizes profits. The inner step selects which consumers to offer discounts off the list prices in order to maximize profits. First, one calculates the expected profits for each consumer if that consumer was offered the list prices. Consumer are then ordered according to the difference between the expected profits they generate at the list prices and at the optimal personalized markup. To satisfy the price regulation while minimizing the impact on profits, the \( (100 - n) \)% of consumers with lowest difference are offered the list prices. The remaining \( n \)% receive personalized discounts. Summing expected profits across consumers yields the total expected profits for a given guess at the markup ceiling (list prices) in the outer loop. The markup ceiling maximizing profits (in the outer loop) is chosen.

This entire simulation process is repeated for a range of discount penetration regulation limits, ranging from \( n\% = 0\% \) (same markups for all consumers) to \( n\% = 100\% \) (no regulation). The impact of these regulations on the firm is intuitive. Profits are monotonically declining in the strength of price regulations (inverse of discount penetration rate).

The simulations reveal that prohibiting price discrimination entirely maximizes consumer surplus. In fact, consumer surplus can be higher absent regulation than with limited regulation. Reducing the percent of consumers allowed to receive personalized discounts from 100% (no regulation) to 97% reduces consumer surplus by 0.13%. For discount penetration cap regulations between 1% and 96%, consumers in aggregate are better off than they are under
full price discrimination, but worse off than they are under uniform pricing. Thus, while it is theoretically possible for less stringent regulations to induce personalized prices that increase consumer surplus, such an outcome is not realized. Consumer surplus is maximized when price discrimination is banned entirely.

To understand why consumers are unable to benefit from the less stringent regulations, consider the impact at the individual consumer level. Consider the case where the discount penetration cap is set 85.43% (i.e., the percent of consumers whose profit-maximizing personalized markup is below the uniform markup observed empirically). The profit-maximizing markup ceiling (determining list prices) is 72%, which is higher than the observed uniform markup of 59%. This higher markup aims to extract surplus from the most avid customers. However, to adhere to the regulations’ quota, the firm must offer this markup to 14.53% of consumers. The list price markup (72%) thus must be applied to some consumers who would be offered lower personalized prices absent regulations. Not surprisingly, the optimal personalized markups for some of these consumers are close to the list markup. But the firm also charges list prices to some consumers who are unlikely to buy at any price, consumers with much lower optimal-personalized markups, because these consumers generate little profit for the firm regardless. The firm is better off using these low value consumers to satisfy the regulations’ list price quota, allowing the firm to personalize discounts to other (mid value) consumers. See Figure 6. Notice that limited regulations result in some consumers paying higher prices than they would absent regulations, and some paying higher prices than they would if price discrimination were banned entirely. Thus, some consumers are made worse off by limited regulations, regardless of whether the alternative is full regulation (banning PD), or no regulation.

In summary, discriminatory pricing can in theory increase both producer surplus and aggregate consumer surplus. Regulators, therefore, might attempt to use limited regulations to reach such an outcome, rather than prohibiting personalized pricing entirely. However, simulations reveal that less stringent price regulations limiting the fraction of consumers receiving a personalized discount reduce aggregate consumer surplus. Moreover, these regulations have
the unintended consequence of increasing prices for the least avid consumers, consumers that would be offered the largest discounts absent regulations. Less stringent regulations are thus unappealing in some ways.

6.4 Robustness Checks

Table 5 shows that the increase in profits from personalized pricing is robust to some modeling assumptions. The first concern is that Netflix may have underpriced in the short run to grow the business, and hence the static optimal-pricing conditions may rely on a questionable assumption. However, I find that even if one assumes that the optimal prices were double the observed prices, and re-estimates the model, the increase in profits from price personalization is roughly the same, at least in percentage terms. The second concern is that movie review websites such as IMDB.com and RottenTomatoes.com might be complements for Netflix’s products. If so, visiting them might not just indicate a higher intrinsic affinity for Netflix, but also indicate that a consumer already subscribes. While this may be true, the impact of movie review websites on the main results is relatively small. Dropping all websites categorized by Alexa Web Rankings as related to movies or TV, along with Wikipedia, and re-running the model lowers the percent gain from price personalization from 12.99% to 10.49%. Even with such sites dropped, personalized pricing based on browsing histories is much more effective than personalized pricing based on only demographics. Lastly, I repeat personalized pricing simulations excluding the most and least avid consumers, the top and bottom half percent, out of concern that functional form assumptions may cause imprecision in estimates at the extremes. Excluding these extreme consumers, the profit increase from personalized pricing is estimated to be 0.075% when only using demographics, and 6.9% when browsing data are used to personalize prices. Although the gain to personalized pricing falls when the extreme consumers are excluded, the main conclusion remains the same: web-browsing data are much more effective than demographic data for personalizing prices.
7 Discussion and Conclusion

This paper finds, in the context of Netflix, that web-browsing data substantially improve targeted advertising and increase profits from personalized pricing, relative to when only demographic data are available. Better ad targeting may benefit advertisers, firms, and even consumers, but the efficiency and equity effects of widespread personalized pricing are less well understood. Most textbooks espouse the efficiency of personalized pricing based on partial equilibrium analysis. But I find that the benefits accrue to firms, not consumers in aggregate, and feasible regulations allowing limited amounts of price discrimination do not appear to benefit consumers.37

A related question is whether it is fair for consumers to pay different prices for the same product. Kahneman et al. (1986) finds personalized pricing was viewed as unfair by 91% of respondents. Yet the prevalence of third-degree price discrimination (and coupons) suggests firms can profit by (effectively) offering different prices to different groups, and therefore consumers are willing to pay different prices than others for the exact same good.

Still, perceived fairness remains an important business consideration, and Amazon’s multiple attempts at personalized pricing are an instructive example. In the year 2000, customers who discovered they were being offered different prices on Amazon reacted with fury.38 In response, Amazon refunded price differences and stopped personalizing prices. Over the last decade, a variety of firms began employing personalized pricing but were more careful than Amazon about framing. Personalized prices were not called their true name, but rather were labeled “customized coupons” or “personalized discounts.” They are not coupons in the traditional sense; they are automatically applied when the consumer reaches the website and typically require no action on the customer’s part (or sometimes just one click). They are

37When price discrimination is employed by multiple firms, the implications are more nuanced. In homogeneous-product oligopolistic markets (Spulber, 1979) and differentiated product markets (Corts, 1998; Thisse and Vives, 1988; Choudhary et al., 2005), personalized PD does unilaterally raise profits, but employed jointly it may increase competition, reducing profits and hence innovation incentives.

38https://tinyurl.com/y76k4mp7
merely called “coupons” or “discounts” to address concerns over perceived unfairness. In 2017, Amazon followed suit. Specifically, Amazon began allowing third-party sellers to offer coupons to consumers who had viewed or purchased certain products, designated by Amazon’s internal product codes (ASINs).\(^{39}\) Figure 7 shows an example of the seller interface allowing targeted couponing. Using the coupon requires trivial effort from the targeted consumer; they merely click the coupon box, as shown in the example in Figure 8. Hence, Amazon allows third-party sellers to offer different prices to consumers based on their browsing histories at Amazon. It may not be widespread at Amazon (yet), but its existence and growing use comprises a conspicuous example of a major retailer embracing personalized pricing, but under a different name to address concerns over perceived unfairness.

References

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

Bergemann, D., B. Brooks, and S. Morris (2015). The limits of price discrimination. *The American Economic Review* 105(3), 921–957.

Chintagunta, P., J.-P. Dube, and K. Goh (2005). Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models. *Management Science* 51, 832–849.

Choudhary, V., A. Ghose, T. Mukhopadhyay, and U. Rajan (2005). Personalized pricing and quality differentiation. *Management Science* 51(7), 1120–1130.

Corts, K. S. (1998). Third-degree price discrimination in oligopoly: All-out competition and strategic commitment. *The RAND Journal of Economics*, 306–323.

Dube, J.-P. and S. Misra (2017). Scalable price targeting. *Working Paper.*

\(^{39}\)https://www.skubana.com/amazon-coupon/
Economist (2014). Getting to know you: Everything people do online is avidly followed by advertisers and third-party trackers. *The Economist Magazine*.

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

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

Goldfarb, A. and C. E. Tucker (2011). Privacy regulation and online advertising. *Management Science* 57(1), 57–71.

Gomez-Uribe, C. A. and N. Hunt (2016). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)* 6(4), 13.

Hannak, A., G. Soeller, D. Lazer, A. Mislove, and C. Wilson (2014). Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pp. 305–318. ACM.

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

Johnson, G. A., R. A. Lewis, and D. H. Reiley (2016). When less is more: Data and power in advertising experiments. *Marketing Science* 36(1), 43–53.

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

Lewis, R. A. and J. M. Rao (2015). The unfavorable economics of measuring the returns to advertising. *The Quarterly Journal of Economics* 130(4), 1941–1973.
McManus, B. (2008). Nonlinear pricing in an oligopoly market: The case of specialty coffee. *The RAND Journal of Economics* 38(2), 512–532.

Mikians, J., L. Gyarmati, V. Erramilli, and N. Laoutaris (2012). Detecting price and search discrimination on the internet. *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, 79–84.

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

Montgomery, A. L., S. Li, K. Srinivasan, and J. C. Liechty (2004). Modeling online browsing and path analysis using clickstream data. *Marketing Science* 23(4), 579–595.

Netflix (2006). Netflix annual report. [http://ir.netflix.com/annuals.cfm](http://ir.netflix.com/annuals.cfm).

Office of the President, E. (2015). Big data and differential pricing. *Washington DC*.

Peukert, C. and I. Reimers (2018). Digital disintermediation and efficiency in the market for ideas. *Working Paper*.

Reimers, I. and B. Shiller (2018). Welfare implications of proprietary data collection: An application to telematics in auto insurance. *Working Paper*.

Robinson, J. (1933). Economics of imperfect competition.

Rossi, P. E., R. E. McCulloch, and G. M. Allenby (1995). Hierarchical modelling of consumer heterogeneity: An application to target marketing. In *Case Studies in Bayesian Statistics, Volume II*, pp. 323–349. Springer.

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

Spulber, D. F. (1979). Non-cooperative equilibrium with price discriminating firms. *Economics Letters* 4(3), 221–227.
Thïsse, J.-F. and X. Vïves (1988). On the strategic choice of spatial price policy. *The American Economic Review* 78(1), 122–137.

Venkatesan, R. and P. W. Farris (2012). Measuring and managing returns from retailer-customized coupon campaigns. *Journal of marketing* 76(1), 76–94.

Waldfogel, J. (2014). First-degree price discrimination goes to school. *Working Paper.*
Figure 1: Frequency of Visits Across Websites

Notes: The figure shows the log frequency of visits to each website over the year, averaged across consumers. The websites are ordered according to this measure on the horizontal axis.
Notes: Consumers in a holdout sample are divided into 600 groups according to their predicted subscription probabilities. For each group, the averaged predicted probability, and actual share subscribing, are each calculated. The predicted probability is shown by the blue line, and the actual share of consumers subscribing to Netflix is shown by green circles.
Figure 3: Range of Predicted Probabilities

Notes: The figure plots the estimated probabilities that individual consumers subscribe, when only demographics are used to predict whether a given consumer subscribes, and when all variables including browsing histories are used. Predicted probabilities are reported for percentiles ranging from 0.1 to 99.9 (in increments of 0.1).
Figure 4: Histograms of Absolute Differences in Probabilities of Subscribing to Netflix, in Randomly Drawn Pairs of Individuals, Both Within and Across ZIP Codes

Notes: Two categories of randomly drawn consumer pairs are considered: pairs of consumers residing in different ZIP codes, and pairs of consumers residing in the same ZIP code. For each pair, the difference between their predicted probabilities of subscribing is calculated. The density of this consumer similarity measure is shown in the figure, separately for pairs in the same ZIP code (solid red rectangles) and in different ZIP codes (dashed blue line edged, empty rectangles). Consumers in the same ZIP code might be expected to be more similar to each other due to Tiebout sorting. But the evidence in the figure does not support that claim.
Figure 5: Densities of Changes in Price when Personalized

Notes: The figure shows the density of percent changes in markups faced by individual consumers when markups are personalized.
Figure 6: Prices Offered Across Consumers Under Various Regulations

Notes: The figure shows the percent markup over cost faced by individual consumers under different scenarios: (1) price discrimination allowed, (2) regulation limiting the fraction of consumers offered a discount off some list price, and (3) regulation preventing price discrimination entirely. The consumer index on the horizontal axis is an ordinal measure of consumers’ estimated affinity for Netflix.
Figure 7: Seller Interface Allowing Sellers to Personalized Coupons on Amazon

Source: https://tinyurl.com/y827fuzr.

Figure 8: Example of Clickable Coupon on Amazon

Notes: Screenshot taken on Amazon.com on Sep 21, 2018, while logged into a user’s account. To redeem the coupon, the consumer merely needs to click the box. At checkout, the price is listed at $63.99, rather than the originally listed price of $79.99.
Table 1: Summary Statistics for Household Demographics and Basic Browsing Habits

| Household Demographics | # Residents |
|-------------------------|-------------|
| I(Children) Yes         | 59.5%       |
| Age (Eldest)            |             |
| Group 1 (18-20)         | 0.3%        |
| Group 2 (21-24)         | 2.0%        |
| Group 3 (25-29)         | 4.1%        |
| Group 4 (30-34)         | 14.9%       |
| Group 5 (35-39)         | 9.1%        |
| Group 6 (40-44)         | 19.7%       |
| Group 7 (45-49)         | 12.1%       |
| Group 8 (50-54)         | 12.1%       |
| Group 9 (55-59)         | 8.7%        |
| Group 10 (60-54)        | 6.8%        |
| Group 11 (65+)          | 10.4%       |
| Zipcode                 |             |
| Mean                    |             |
| StDev                   |             |

| Browsing Habits          |             |
| I(Broadband) Yes         | 78.5%       |
| Timing of internet use   |             |
| Early morning (midnight to 6am) | 8.7% | 14.4 |
| Mid morning (6am to 9am) | 11.7% | 13.5 |
| Late morning (9am to noon)| 21.1% | 13.9 |
| Afternoon (noon to 5pm)  | 34.9% | 16.1 |
| Evening (6pm to midnight)| 23.6% | 26.5 |
| Income (in thousands)    |             |
| Group 1 (< 15)           | 9.8%        |
| Group 2 (15-24.9)        | 6.5%        |
| Group 3 (25-34.9)        | 10.9%       |
| Group 4 (35-49.9)        | 20.3%       |
| Group 5 (50-74.9)        | 26.0%       |
| Group 6 (75-99.9)        | 12.2%       |
| Group 7 (> 100)          | 14.3%       |
| Monday                   | 15% | 4.0 |
| Tuesday                  | 15.6% | 4.8 |
| Wednesday                | 15.6% | 4.6 |
| Thursday                 | 15.2% | 4.7 |
| Friday                   | 14.6% | 4.4 |
| Saturday                 | 11.9% | 7.0 |
| Sunday                   | 12.1% | 7.2 |

Notes: The table reports summary statistics for the demographic and basic browsing behavior variables. The percent of households in the sample in the respective grouping are reported for binary grouping variables. For other variables, the mean and standard deviation across individuals are reported.
Table 2: Estimation Results - Prediction Whether Individual Subscribes

| Order | Visits to:        | Coef. | Value (w. one standard deviation increase in visits) | \( \Delta \Pr(\text{subscribe}) \) |
|-------|-------------------|-------|-----------------------------------------------------|----------------------------------|
| 1     | gamefly.com       | 0.78  | 11.78                                                |                                  |
| 2     | ameblo.jp         | -0.48 | -5.03                                                |                                  |
| 3     | slysoft.com       | 0.36  | 4.80                                                 |                                  |
| 4     | audible.com       | 0.33  | 4.40                                                 |                                  |
| 5     | dvdfab.com        | 0.30  | 4.01                                                 |                                  |
| 6     | sutterhealth.org  | 0.30  | 3.96                                                 |                                  |
| 7     | 4chan.org         | 0.29  | 3.84                                                 |                                  |
| 8     | jambase.com       | 0.29  | 3.83                                                 |                                  |
| 9     | imdb.com          | 0.28  | 3.72                                                 |                                  |
| 10    | houstonpress.com  | 0.27  | 3.49                                                 |                                  |
| 11    | kw.com            | -0.26 | -2.90                                                |                                  |
| 12    | somethingawful.com| -0.25 | -2.83                                                |                                  |
| 13    | jacksonville.com  | -0.24 | -2.72                                                |                                  |
| 14    | lacity.org        | -0.24 | -2.65                                                |                                  |
| 15    | jalopyjournal.com | 0.23  | 3.03                                                 |                                  |
| 16    | uhaul.com         | 0.23  | 2.99                                                 |                                  |
| 17    | smackjeeves.com   | 0.23  | 2.98                                                 |                                  |
| 18    | dailypress.com    | -0.23 | -2.58                                                |                                  |
| 19    | sonlight-email.com| -0.22 | -2.54                                                |                                  |
| 20    | fairfaxcounty.gov | 0.22  | 2.89                                                 |                                  |
| 21    | ganeshaspeaks.com | 0.22  | 2.89                                                 |                                  |
| 22    | onstation.com     | -0.22 | -2.49                                                |                                  |
| 23    | whig.com          | 0.21  | 2.74                                                 |                                  |
| 24    | techdirt.com      | 0.21  | 2.68                                                 |                                  |
| 25    | zylom.com         | -0.21 | -2.36                                                |                                  |
| 26    | npr.org           | 0.20  | 2.64                                                 |                                  |
| 27    | baseballamerica.com| 0.20 | 2.63                                                 |                                  |
| 28    | apunkachoice.com  | 0.20  | 2.63                                                 |                                  |
| 29    | elpais.com        | 0.20  | 2.61                                                 |                                  |
| 30    | amazon.com        | 0.20  | 2.56                                                 |                                  |

Notes: The table shows coefficient estimates from a logistic regression with LASSO regularization. The 30 variables with largest coefficients are reported. Prior to estimation, all variables were normalized, so the mean and standard deviation of visits across individuals are 0 and 1, respectively. Column (3) reports the coefficient estimates. Column (4) shows the corresponding percentage point increase in subscription probability arising from a one standard deviation increase in the respective variable, averaged across consumers.
Table 3: Simulated Changes in Various Outcomes Resulting from Personalized Markups

| Percent Change When Price Based on: | Demographics   | All Variables |
|------------------------------------|----------------|---------------|
| Total Profits                      | 0.27% (0.01)   | 12.99% (0.23) |
| Subscribers                        | 0.13% (0.00)   | 3.12% (0.03)  |
| Sales (DVDs At-a-Time)             | 0.14% (0.00)   | 3.21% (0.05)  |
| Aggregate Consumer Surplus         | 0.04% (0.00)   | −0.50% (0.08) |
| Joint Surplus                      | 0.05% (0.00)   | 0.91% (0.02)  |

Notes: The table reports the changes in simulated outcomes when markups are personalized, compared to when they are not. Bootstrapped standard errors in parentheses.

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

| Non-Personalized | Personalized |
|------------------|--------------|
| Linear Pricing   | N/A          | 12.99% (0.23) |
| 2nd-Degree PD    | 22.48% (0.17) | 36.86% (0.15) |

Notes: Linear pricing denotes a consistent percent markup (θ) over cost across all three tiers. Second-degree price discrimination denotes that the percent markup (θ_j) may differ across the tiers. Personalized pricing denotes consumers are charged different markups based on their browsing histories. Bootstrapped standard errors in parentheses.
Table 5: Robustness Checks

| Price Personalization Based on: | Percent Increase in Profits When |  |  |
|-------------------------------|---------------------------------|--|--|
|                               | Demographics                     | All Variables |
| Main Model                    | 0.27% (0.01)                     | 12.99% (0.23) |
| Robustness Checks             |                                 |               |
| Excluding Movie and TV Sites  | N/A                              | 10.50% (0.44) |
| Less Price Sensitive Model    | 0.27% (0.01)                     | 13.00% (0.29) |
| Excl. Extremes (1%)           | 0.07% (0.01)                     | 6.92% (0.22)  |

Notes: The first robustness check re-estimates the model excluding any site categorized as related to movies of TV, according to Alexa Web Rankings, from the set of websites used to predict subscription and affinity for Netflix. The second robustness check re-estimates the model assuming that optimal markups are twice as high as observed markups. The last robustness check re-simulated profits under both status quo and personalized markups, excluding the top and bottom half a percent of consumers with highest/lowest predicted affinity for Netflix. Bootstrapped standard errors in parentheses.