First-Degree Price Discrimination Using Big Data

Benjamin Reed Shiller
Brandeis University*

January 30, 2014

Abstract

First-degree price discrimination, or person-specific pricing, had until recently rarely been observed. The reason, that reservation values were unobtainable, may no longer be true now that massive datasets tracking detailed individual behavior exist. Hence, a fundamental change in the way goods are priced may be underway. I investigate this claim in one context. I show demographics, which in the past could be used to personalize prices, poorly predict which consumers subscribe to Netflix. By contrast, modern web-browsing data, with variables such as visits to Amazon.com and internet use on Tuesdays - variables which reflect behavior - do substantially better. I then present a model to estimate demand and simulate outcomes had 1^{st} degree PD been implemented. The model is structural, derived from canonical theory models, but resembles an ordered Probit, allowing a new method for handling massive datasets and addressing overfitting and high dimensionality. Simulations show using demographics alone to tailor prices raises profits by 0.8%. Including nearly 5000 potential website browsing explanatory variables increases profits by much more, 12.2%, increasingly the appeal of tailored pricing, and resulting in some consumers paying double the price others do for the exact same product. Implications for the overall economy and its structure are discussed. (JEL: D42, L130)

(Keywords: First-Degree, Price Discrimination, Pricing, Big Data, Model Averaging)

*Shiller: Brandeis University, MS01, 415 South St., Waltham, MA 02453, shiller@brandeis.edu. I would like to thank Kathryn Graddy, Shane Greenstein, Adam Jaffe, and Joel Waldfogel for very useful feedback, suggestions, and advice. I would also like to thank Erik Brynjolfsson, Eric Clemons, Iain Cockburn, Jeremey Howard, Gary Jefferson, Blake LeBaron, Rachel McCulloch, 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 instilling in me some of their extensive knowledge of marketing in practice. Lastly, I would like to thank Steven Knowles, for great suggestions.

First-degree Price Discrimination (PD), setting a different price schedule to each individual, in theory allows firms to extract more profits than with other forms of pricing.[1] Yet, its use has been extremely rare in practice, because the requisite information on individuals' reservation values was simply unavailable. Times may be changing. While not widely recognized, because its use is often intentionally inconspicuous, first-degree PD in the form of personalized deals has recently gained traction, for example at some major grocers, pharmacies, and department stores.[2] I argue that large datasets on individual behavior, popularly referred to as "big data," are available and may reveal information such as latent demographics and interest in related products, which can be used to form a hedonic estimate of individuals' reservation values.[3] Therefore, instead of the exception, profitable personalized pricing may become the norm. In this paper, I investigate this claim. Specifically, I estimate the profit gained from first-degree PD when nearly 5000 web-browsing variables are used to estimate individuals' reservation values for Netflix. However, this alone cannot prove a break from the past. So, as a comparison, I compute the analogous profit gained from first-degree PD when only demographics, which have long been available, are used as explanatory variables. Estimation is non-trivial. A new econometric method is introduced, Ordered-choice Model Averaging (OMA), blending economic modeling with advances in machine learning to estimate optimal prices and overcome problems from over-fitting and high dimensionality.

Widespread use of first-degree PD may have large effects on the overall economy and its structure. By increasing profits of firms with monopoly power, it increases the incentives to innovate and differentiate.[4] Its use also changes the impacts of mergers in concentrated industries, and the impacts of privacy and data property rights laws. Such pricing may also cause consumers to waste effort masking themselves as low valuation consumers. Or, in an extreme, albeit unlikely scenario, consumers could reduce labor effort, knowing that earning higher wages would result in being charged higher prices. These are some of the effects that might occur if first-degree PD is widely used. And widely it might - Graddy [1995] shows another form of PD is used even in a seemingly perfectly competitive market. Later, I use BEA statistics to yield a rough but conservative estimate that suggests over a third of all consumption may be ripe for personalized pricing. One goal of this paper is to demonstrate that first-degree PD is now feasible, suggesting we need research on general equilibrium models assuming near-universal first-degree PD.

Empirical realities guide the research framework. Since firms often disguise personalized

[1] Shiller and Waldfogel [2011] find bundling, nonlinear pricing, and third-degree PD cannot extract more than about a third of surplus as profits. First-degree PD in theory can extract all surplus.

[2] See [Elliot [2013], Gross [2012], Thau [2014]].

[3] Madrigal [2012], Mayer [2011] note web browsing behavior is collected (and sold) by hundreds of firms.

[4] First-degree PD, done jointly by oligopolists, may lower profits [Spulber [1979], Thisse and Vives [1988]].

pricing, and its use is growing, estimation of the current or future impact using aggregated data is infeasible. Moreover, before/after comparisons for single firms may be biased, since the decision to first-degree PD is endogenous. So instead, this paper uses an estimated demand model to simulate the impact personalized pricing would have had, if used in one context.

Netflix provides an auspicious context for study. First, since purchases occur online, Netflix could easily implement tailored pricing based on web data. Second, Netflix can effectively price discriminate, as evident from its use of second-degree PD. Third, unlike in most contexts, hypothetical first-degree PD 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. However, Netflix subscription can easily be imputed from web-browsing histories.

The first part of the paper examines the extent to which web browsing behavior improves predictions of which consumers subscribe, compared with predictions based on demographic data which may have been used in the past to tailor prices in face-to-face transactions. Specifically, I run binary choice regression models using different sets of explanatory variables. The larger the spread of predicted probabilities, the better the model predicts which consumers are likely to subscribe at observed prices. In these analyses, the fit of predictions can be evaluated using a holdout sample.

In these and later analyses, model selection and overfitting are major concerns, particularly when selecting from nearly 5000 web browsing explanatory variables. The problems arise because it is not clear, from the estimation sample alone, which variables are significant by chance and which contain information about the true data-generating process. Model averaging - averaging predictions from an ensemble of models - has been shown to yield better predictions than do single models, both in theory and practice.[5] I find that after including 17 website variables via stepwise regression, adding further variables tends to worsen out-of-sample predictive fit. However, when instead using model averaging, out-of-sample predictive fit generally continues to improve as the next 50 variables are added.

I provide a method which adapts model averaging to an ordered-choice framework so it can be used in empirical economic modeling, while simultaneously correcting for the tendency of overfit models to underestimate prediction error. This method, which I refer to as Ordered-choice Model Averaging (OMA), is simple to implement in nearly all statistical packages. It thus solves a difficult problem for practitioners. A naive practitioner personalizing prices on spurious predictors would in effect be randomly assigning prices to individuals, which yields lower expected profits than optimized *non*-personalized prices. Using the OMA method addresses this problem, making personalized pricing profitable.

[5]See, for example, Madigan and Raftery [1994] and Bell et al. [2010]

After correcting for overfitting, the web browsing data indeed help predict which consumers subscribe to Netflix. Without any information, each individual's probability of subscribing is the same, about 16%. Including standard demographics, such as race, age, income, children, population density of residence, etc., in a Probit model improves prediction modestly - individual predicted probabilities of subscribing range from 6% to 30%. Adding the full set of variables in the OMA method, including web browsing histories and variables derived from them, substantially improves prediction - predicted probabilities range from close to zero to 99.8%.

Next, an empirical model is used to translate the increased precision from web browsing data into key outcome variables. Specifically, a demand estimation model derived from canonical quality discrimination theory models is used to estimate individual-level demand for Netflix in the observed environment, in which Netflix employed second-degree PD, but not first-degree PD. The model is then used to simulate pricing and profits in the hypothetical counterfactual occurring had Netflix implemented first-degree PD.

I find that web browsing behavior substantially raises the amount by which person-specific pricing increases profits relative to second-degree PD - 12.18% if using all data to tailor prices, but only 0.79% using demographics alone. Web browsing data hence make first-degree PD more appealing to firms and likely to be implemented, thus impacting consumers. Aggregate consumer surplus is estimated to fall by 8%. Moreover, substantial equity concerns may arise - I find some consumers may be charged about double the price some others are charged for the same product.

To my knowledge, no previous empirical papers study true first-degree PD. A concurrent paper, Waldfogel [2014], studies first-degree PD in university tuition, utilizing data from applications. Previously, Shiller and Waldfogel [2011] investigate the effectiveness of one form of first-degree PD, but assume the firm knows each individual's reservation value exactly.

The closest prior literature, a series of papers in marketing starting with Rossi et al. [1996], estimate the revenue gained from tailored pricing based on past purchase history of the same product.[6] However, they assumed that consumers were myopic. Anecdotal evidence following Amazon's pricing experiment in the early 2000's suggests otherwise [Streitfeld [2000]]. Acquisti and Varian [2005] and Fudenberg and Villas-Boas [2007] show theoretically that first-degree PD based on past purchase history actually reduces monopolist profits when consumers are forward-looking, using arguments quite similar to Coase [1972]. Consumers can avoid being charged high prices using simple heuristics such as "don't buy early at high prices," or by

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

obscuring prior purchases by deleting cookies and masking IP addresses. Such pricing is more similar to second-degree PD or intertemporal pricing than it is explicitly setting prices to different individuals.

By contrast, tailored pricing based on many variables is not subject to the same criticism. First, with bounded rationality consumers may not be able to avoid being charged high prices. I find, for example, that Netflix should charge lower prices to individuals that use the internet during the day on Tuesdays and Thursdays, and higher prices to those that visit Wikipedia.org, patterns consumers may not recognize. Moreover, with many variables, there may not be any easy heuristics consumers can follow to avoid being charged high prices. Furthermore, heuristics for one product may not apply to other products. Even if consumers did understand 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. Finally, firms could charge high prices to any consumers not revealing their web browsing data, providing the incentive for consumers to reveal them.

The remainder of the paper is organized as follows. Section 1 describes the context and industry background. Next, Section 2 describes the data. Section 3 then shows how well various sets of variables explain propensity to purchase. Lastly, Sections 4 and 5 present a demand estimation model and estimate optimal person-specific prices.

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.[7]

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.[8] 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

[7] 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. $11.57/(0.606 * 114.384) * 100 \approx 16.7$.

[8] 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]].

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.[9] 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 1-3 DVD at-a-time unlimited plans were estimated using industry statistics and expert guidance. They are assumed to equal $6.28, $9.43, and $11.32, respectively.[10]

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.[11] 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 average more than two page views per Netflix visit. The reasoning behind this rule is that subscribers have reason to visit more pages within Netflix.com to search for movies, visit their queue, rate

[9]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.

[10]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.

[11]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.

movies, etc. Non-subscribers do not, nor can they access as many pages. 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 housedholds 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 behavioral variables were derived from the data. These included 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 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.[12] 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.[13] These variables are 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, one variable for each of the $4,788$ websites. All explanatory variables were normalized.

The data were randomly split into two samples of individuals, approximately equal in size. The first, an estimation sample, is used for estimating model parameters. The second, a holdout sample, is used to test for and address overfitting.

[12]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.

[13]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.

3 Prediction in Status Quo

This section estimates the probability that each consumer subscribes to Netflix using a Probit model in an estimation sample of half the observations, based on different sets of explanatory variables. The predictions are then contrasted as more sets are added, to inform on the incremental predictive ability of web behavior.

First, a Probit model is used to investigate which standard demographic variables are significant predictors of Netflix subscription. Variables are selected via a stepwise regression procedure, with bidirectional elimination at the 5% significance level. The results are shown in Table 1. Race, Hispanic indicator, Census region, and income are found to be significant. These are variables which might be gleaned in face-to-face transactions from observed physical appearence, accent, and attire, and hence could have been used to tailor prices in the past, before web browsing data became available.

Next, the set of basic web behavior variables are added, again using the stepwise procedure. The log likelihood increases by 448.7, indicating this group of added variables is significant with a p-value so low as to not be distinguishable from zero with standard machine precision. Note also that several demographic variables are no longer significant once basic web behavior variables are added, suggesting they are less accurate proxies for information contained in behavior, which cannot be easily observed in anonymous offline transactions.

Next, detailed web behavior variables are tested individually for their ability to predict Netflix subscription. Specifically, a Probit model is re-run 4,788 times, each time including the significant demographic and basic web behavior variables from Table 1 and exactly one website variable. Overall, 29% of websites were significant at the 5% level, and 18% at the 1% level, far more than expected by chance alone. The types of websites found to be most significant, shown in Table 2, and their positive signs, intuitively suggest that consumers' observed web browsing behavior is driven by the same innate characteristics as their Netflix choices.[14] 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.com), discount shoppers (Bizrate, Price Grabber), and internet savvy users (Wikipedia).

I next investigate the joint prediction of all website variables combined, rather than just

[14]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.

considering one variable at a time. An immediate obvious concern is 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, selecting all variables with p-values below 0.05 may include many variables significant by chance alone in the sample of data available. Since such variables and coefficient values do not reflect underlying patterns in the true data-generating process, they add 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 that do reflect underlying patterns in the true data-generating process.[15]

The second conceptual problem from overfitting follows as a result - the best-fitting model in sample implies predictions which are too good to be true. Including variables significant by chance in the sample of data available yields a better in-sample fit, with lower error magnitude. However, this lower error is fallacious - it would not apply in fresh samples. Basing predictions on the in-sample error would thus yield misleading conclusions in this paper. A lower-than-true model prediction error here implies a more accurate-than-true estimate of an individual's willingness to pay, which in turn implies higher-than-true profit from first-degree PD. It would thus bias upwards the gain from using personalized pricing. Analogous problems apply in this section as well.

Solutions to these overfitting problems can likewise be broken into two parts - first choosing the optimal level of complexity (threshold for variable inclusion), and then ex-post re-estimating the size of the error in the chosen model using a holdout sample. After the latter correction, predictions from a model with any chosen level of complexity will in expectation reflect the true level of uncertainty according to that model, i.e. the error magnitude will be unbiased for that model. Hence, after the correction, I am no longer biased towards strong findings. If anything, the opposite is true. If I, the researcher, were to poorly choose complexity in the first step, this would only imply that a more skilled statistician could extract even greater profits from tailored pricing, and better predict which consumers subscribe at observed prices.

To choose complexity, I follow techniques from machine learning. I begin with forward step-wise regression. Figure 1 shows that as more variables are added the likelihood continues to increase in the estimation sample, as expected. However, the corresponding likelihood using the same parameters and same coefficient values in a holdout sample peaks after the 17^{th} variable is added, typically declining thereafter.[16] This is a rather typical pattern in machine

[15] Over-complex model selection is analogous to "high variance" in the machine learning literature.

[16] Using a holdout sample to determine complexity level is similar to using information criterion methods (AIC, BIC), but requires few assumptions.

learning problems - too much complexity worsens out-of-sample fit.

This overfitting occurring when adding more than 17 website variables to the model obscures information useful for predicting purchase. Including the next most significant variable yields a model which fits worse in the holdout sample. However, there is still useful information contained in additional variables. To demonstrate, I again try adding each website variable in the 18^{th} position, this time recording the 50 models which fit best in the estimation sample. Using the coefficient values estimated in the estimation sample, I find that 48% of these models yield better fit in the holdout sample, compared to the model with 17 website variables. This implies many of these additional variables contain pertinent information. The crux of the problem is that, based on the estimation sample, it is not clear which of these additional variables and coefficient values capture true patterns from the data generating process, and which simply capture in-sample noise. Classic variable selection techniques perform poorly in handling this uncertainty.

If prediction rather than variable selection is the main goal, model averaging can help address this problem. Specifically, averaging predictions over an ensemble of models can yield better predictions than can any single model used in the average.[17] The story behind the winning teams of the Netflix Prize Challenge is a well-known example of its effectiveness (Bell et al. [2010]).

In the context of the binary Probit model used here, averaging proceeds by estimating models with the demographic and basic web behavior variables, the 17 website visit variables from stepwise regression, and one additional website variable which changes each time. Each Probit model yields, for each observation, an estimate of the difference between the underlying latent variable and the threshold the true underlying latent variable must exceed for the individual to subscribe.[18] This estimated difference for individual i, $\hat{y}_{i,diff}$, determines the probability individual i subscribes to Netflix. Averaging $\hat{y}_{i,diff}$ over models with different website variables in the 18^{th} spot yields the model averaging estimate of this difference, $\bar{y}_{i,diff}$.

Figure 1 shows the improvement from model averaging, as the number of models in the average increases. Specifically, models with different websites added as the 18^{th} variable are ranked according to the in-sample likelihood, and then $\hat{y}_{i,diff}$ from the n best-fitting among these are averaged. Out-of-sample prediction for different size n are shown by the dotted blue line in the figure, where the values on the x-axis correspond to the total number of websites used in prediction - the 17 included in all models plus all others included in any model used in the average. The figure shows that, in contrast to stepwise variable selection, when model

[17] Madigan and Raftery [1994] provide a proof that model averaging in a Bayesian framework yields predictions at least as good in expectation as predictions from any single model used in the average.

[18] For details on binary/ordered choice models, see Greene and Hensher [2010].

averaging is used increasing complexity further improves out-of-sample fit. Averaging using the top 50 seems to be about optimal, and is used subsequently.[19] In unreported tests, this entire process was repeated using principal components as explanatory variables, finding no improvement.

Next, the relative error in the model is rescaled for two reason - first to account for overfitting, which yields a downward biased estimate of the standard deviation of the model's error term, and second to account for an issue discovered when using model averaging in an ordered-choice framework, which oppositely biases it upwards.[20] Essentially, one simply re-estimates only the standard deviation of the error term in the model, finding the value which maximizes the likelihood of that model in a holdout sample. The resulting standard deviation is an unbiased estimate of that model's true standard deviation of the error, addressing both sources of bias simultaneously. In an ordered-choice framework, the standard deviation of the error term is assumed fixed, so the analogous change required is a rescaling of all other parameters. Specifically, another ordered-choice model is run in the holdout sample with $\bar{y}_{i,diff}$ as the sole explanatory variable, yielding its coefficient and a new threshold parameter. Multiplying $\bar{y}_{i,diff}$ by its coefficient, and subtracting the new threshold parameter, yields a new rescaled estimate of $\bar{y}_{i,diff}$. This rescaling in an ordered-choice model is akin to rescaling the error, and therefore removes the bias. This method is subsequently labeled the Ordered-choice Model Averaging (OMA) method.

Figure 2 shows the fit of the predictions from model averaging in the holdout sample. Specifically, individuals in the holdout sample are ordered according to their probability of subscribing to Netflix according to the model, then grouped. The average predicted probability and observed probabilities are then calculated for each group. Notice that these predicted probabilities, shown in the solid blue line, do in fact seem to follow the actual probabilities of subscription.

The main takeaways from this section are apparent in Figure 3. It shows the predicted probability each individual subscribes based on various sets of explanatory variables plotted together on one graph.[21] Including web behavior variables does in fact seem to substantially

[19] Madigan and Raftery [1994] uses the logic behind Occam's razor to advocate not using all models in the averaging.

[20] In an ordered-choice framework, as additional meaningful explanatory variables are added, the average error size (relative to magnitude of coefficients) declines. Since the standard deviation of the error term is assumed fixed, in ordered-choice models it is the scaling of $\hat{y}_{i,diff}$ that increases, relative to the error, as fit improves. In model averaging, however, by construction the scaling of $\bar{y}_{i,diff}$ does not systematically change as more models are included in the average. Hence averaging does not incorporate the increased predictive ability from jointly including the information from different models, resulting in the bias.

[21] Note the Y-axis range is larger than in Figure 2, which averages predicted probability within groups, obscuring extreme probabilities.

help prediction of which consumers are highly likely or highly unlikely to subscribe. Predicted probabilities of subscription ranged from $3.29 * 10^{-4}$ percent to 99.8% when all variables are used for prediction, but only from 6% to 30% when based on demographics alone. Without any information, each individual has a 16% likelihood of subscribing.

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, I first restrict the holdout sample to the 25,440 individuals who live in the same zipcode as at least one other individual. Then, for each pair-combination of individuals in the same zipcode, I calculate the absolute difference in their predicted probabilities of subscribing, a measure of how similar their preferences are. I then construct a comparison group as follows. For each pair in the same zipcode, each individual in the pair is matched with a randomly drawn individual from another zipcode, and the difference in probability of subscribing is likewise calculated for each newly formed pair. Comparing the range of differences in pair-probabilities for individuals in the same zipcodes with the corresponding differences for across-zipcodes pairs informs on whether individuals living in the same zipcodes are more similar in their propensity to consume Netflix. Figure 4, which plots overlaid histograms of absolute differences in predicted probabilities, separately for pairs in/not in the same zipcode, shows that there is not a meaningful difference.[22] Hence, web browsing data offer mostly distinct information from that contained by geography.

Figure 5 illustrates the information lost when only demographics are used to predict purchase. The figure plots the range of predicted probabilities, based on all variables, for two groups. The first group is the 10% of individuals with the lowest predicted probability when only demographics are used for prediction. When based on demographics, predicted probabilities of subscribing for this group range from 6.5% to 12%. When based on the full set of variables, predicted probabilities for this same group are as high as 85%. The second group contains the 10% of individuals predicted to have the highest probability of subscription when only demographics are used for prediction. A similar pattern emerges for this group.

4 Model and Estimation

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

[22]While not meaningful, the difference is statistically significant at the 0.005 level, according to the Kolmogorov-Smirnov test.

each tier, but may differ in how much they value the quality of higher tiers. The firm sets prices of the tiers of service, but not qualities.[23]

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 assumptions in order for the model to be identified. If one had time-varying prices, then one could use a more flexible model which estimates heterogeneous price sensitivities. Such a model is shown in Appendix A.

4.1 Model

The conditional indirect utility that consumer i receives from choosing product j equals:

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

where q_j and P_j are the quality and price of product j. The products are indexed in increasing order of quality. I.e., if $j > k$, then $q_j > q_k$. The parameter y_i is a person-specific parameter reflecting individual i's valuation for quality, and I_i is their income. The price sensitivity α is assumed to be the same across individuals. This utility specification is analogous to the one in Mussa and Rosen [1978].

For consumer i to weakly prefer product j to product k, the following incentive compatibility constraint must hold:

$$y_i q_j + \alpha \left(I_i - P_j \right) \geq y_i q_k + \alpha \left(I_i - P_k \right) \tag{2}$$

If q_j is greater than q_k, this reduces to:

$$y_i \geq \alpha \frac{P_j - P_k}{q_j - q_k} \tag{3}$$

If $\frac{P_j - P_k}{q_j - q_k}$ is strictly increasing in j, then no quality tier is a strictly dominated choice for all possible values of y_i. In that case, the incentive compatibility constraints only bind for products neighboring in quality, and consumer i chooses product j if and only if the following inequality condition is satisfied:[24]

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

[24]The individual rationality constraint follows the same form, with subscripts 0 (the outside good) and 1 (the

$$\alpha \frac{P_j - P_{j-1}}{q_j - q_{j-1}} \leq y_i < \alpha \frac{P_{j+1} - P_j}{q_{j+1} - q_j} \tag{4}$$

Next, y_i is replaced with a linear regression expression, $\beta_0 + X_i\beta + \sigma\epsilon_i$, and $(P_j - P_{j-1})$ and $(q_j - q_{j-1})^{-1}$ are replaced with more concise notation, $P_{\Delta j}$ and λ_j, respectively. Substituting these changes into equation 4 yields:

$$\alpha\lambda_j P_{\Delta j} \leq \beta_0 + X_i\beta + \sigma\epsilon_i < \alpha\lambda_{j+1}P_{\Delta j+1} \tag{5}$$

A couple of normalizations are required. First, σ, the standard deviation of the error term, is not separately identified from the scaling of the remaining parameters in the model. As is standard in ordered-choice models, it is normalized to 1. Second, price sensitivity α cannot be separately identified from the scaling of quality levels, λ_j, so α is also arbitrarily normalized to 1. Incorporating these changes into equation 5, and rearranging yields:

$$\theta_{i,j} \leq \epsilon_i < \theta_{i,j+1} \tag{6}$$

where

$$\theta_{i,j} = -\beta_0 + \lambda_j P_{\Delta j} - X_i\beta = \mu_j - X_i\beta \tag{7}$$

The term $\mu_j = -\beta_0 + \lambda_j P_{\Delta j}$ has been introduced to highlight the fact that β_0 and $\lambda_j P_{\Delta j}$ are not separately identified when price does not vary.

Finally, the probability that product j is consumed by individual i equals:

$$s_{i,j} = F(\theta_{i,j+1}) - F(\theta_{i,j}) \tag{8}$$

where $F()$ is the CDF of ϵ.

4.2 Model Intuition Graphically

Figure 6 helps provide intuition for the model's mechanics. On the x-axis is the uncertainty in individual i's value for quality (affinity for renting movies by mail), ϵ_i. Locations further to

lowest tier) replacing subscripts k and j in equation 3, respectively. Since quality differences determine these constraints, the assumed quality of the outside good is inconsequential. P_0, the price of not buying Netflix, equals zero.

the right correspond to higher affinity for movies by mail. The curve represents the probability density function of ϵ_i.

If the shock ϵ_i is large enough, then the individual values quality enough to be willing to buy Netflix's 1 DVD at-a-time plan, as opposed to no plan. The corresponding threshold that ϵ_i must exceed is given by $\theta_{i,1}$ from equation 7, shown by a vertical line in Figure 6. If the individual values quality (movies) even more, then the individual might prefer the 2 DVDs at-a-time plan to the 1 DVD at-a-time plan. This occurs when $\epsilon_i \geq \theta_{i,2}$. Similarly, the consumer prefers 3 to 2 DVDs at-a-time when $\epsilon_i \geq \theta_{i,3}$. Hence, the probability that individual i chooses a given tier j equals the area of the PDF of ϵ_i between $\theta_{i,j}$ and the next highest threshold $\theta_{i,j+1}$. For $j = 1$, the one DVD at-a-time plan, this probability is given by area A in the figure.

The model estimates how the values of $\theta_{i,j}$, whose formula is shown in equation 7, vary with the explanatory variables. Suppose visits to a celebrity gossip website, a variable in set X, predicts a tendency to consume Netflix, indicating a consumer with many such visits has higher value for Netflix products on average. Then the corresponding component of β in the equation for $\theta_{i,j}$ would have a positive value. Since $X_i\beta$ enters negatively in equation 7, its impact on $\theta_{i,j}$ is negative. Hence, in Figure 6, a unit increase in the value of this X shifts all three values of $\theta_{i,j}$ left by the corresponding value of β, capturing the higher probability that individual i subscribes to Netflix.[25,26]

The values of $\theta_{i,j}$ in equation 7 are also impacted by prices. $\theta_{i,j}$ shifts to the right when there is an increase in the difference between the prices of tiers j and $j-1$, i.e. when $P_{\Delta j} = P_j - P_{j-1}$ increases. This implies the individual must have an even higher value for quality, higher value of ϵ, in order to be willing to choose tier j over tier $j-1$. A price increase in j also lowers the value of $P_{\Delta j+1} = P_{j+1} - P_j$ resulting in $\theta_{i,j+1}$ shifting to the left. Hence, when the price of tier j increases, some consumers switch to either the higher or lower adjacent tier. Note, however, that since $\frac{\partial \theta_{i,j}}{\partial P_{\Delta j}}$ and $\frac{\partial \theta_{i,j}}{\partial P_{\Delta j+1}}$ cannot be estimated in the model without price variation, their values must be calculated ex-post using auxiliary information.

Once $\theta_{i,j}$, $\frac{\partial \theta_{i,j}}{\partial P_{\Delta j}}$, and $\frac{\partial \theta_{i,j}}{\partial P_{\Delta j+1}}$ are known, one can simulate expected profits under counterfactual prices. Any given set of prices implies some probabilities that individual i consumes each

[25] One could try more flexible function forms for $\theta_{i,j}$, for example by allowing β, i.e. coefficient on X, to differ across j. However, this could cause violations to the single crossing property [Wilson [1993]], and result in odd preference orderings, such as a consumer strictly preferring the one DVD at-a-time to the two DVDs at-a-time, even when the two options are priced the same. The structure imposed by the model prevents odd outcomes like this one from occurring, using economic reasoning to presumably improve accuracy.

[26] The equations for $\theta_{i,j}$ are estimated subject to the normalized value of the standard deviation of ϵ. Note, however, that the scaling of ϵ is irrelevant in determining outcomes. If one were to instead assume, say, a higher level of the standard deviation of ϵ, σ, then the model and data would yield estimates of all other parameters exactly σ times higher as well. Due to this countervailing change, $\frac{\partial s_{i,j}}{\partial X_j}$ would be left unchanged.

tier. The expected revenues from the individual in Figure 6 equals $P_1 * Area\ A + P_2 * Area\ B + P_3 * Area\ C$, where the areas depend on prices and the individual's values of X_i. Total expected revenues are then found by summing expected revenues across individuals.

4.3 Estimation

After assuming that the ϵ error term is normally distributed, the model presented above resembles an ordered Probit model. Hence, estimation can proceed via straight-forward maximum likelihood.[27]

In this specific context, however, a couple of additional modifications are necessary before the model can be estimated. First, I assume that consumers face a choice between the 1, 2, and 3 DVDs at-a-time plans with unlimited number sent each month. There were a few Netflix subscription plans limiting the number of DVDs that could be received monthly, which do not cleanly fit into this ordered-choice setup. However, these limited subscription plans had small market shares in the data (combined shares about 10%). It is assumed that consumers of these plans would subscribe to one of the unlimited plans, had these limited plans been unavailable.[28] Second, while I can impute whether or not a given individual subscribed to Netflix, for most subscribers it is not known directly which tier he or she subscribed to. The partially-concealed tier choice requires slight modications to the likelihood function. As a result, it less well resembles the likelihood function in standard ordered probit models.[29]

The log likelihood function equals:

$$l(D;\mu,\beta) = \sum_{i(j=-1)} \log(F(\theta_{i,1})) + \sum_{i(j=0)} \log\left(1 - F(\theta_{i,1})\right) + \sum_{k=1}^{3} \sum_{i(j=k)} \log\left(F(\theta_{i,k+1}) - F(\theta_{i,k})\right) \quad (9)$$

where the data D contain subscription choice and explanatory variables, and $\theta_{i,j}$ is the function of parameters μ and β defined in equation 7. The notation $i(j = -1)$ denotes the set of indi-

[27] When including website variables, the same machine learning techniques from Section 3 are used. $\hat{y}_i$, the model's estimate of y_i, is averaged across an ensemble of models, to yield the model averaging estimate. It is then rescaled, as before, using the holdout sample, to remove any bias.

[28] A "'4 DVDS at-a-time"' unlimited plan was also available, however less than 1% of subscribers chose this plan. Owners of this plan were combined with the "'3 DVDS at-a-time"' plan owners for estimation.

[29] The ordered-choice thresholds for the 2^{nd} and 3^{rd} tiers (μ_2 and μ_3) are determined by the fraction choosing each Netflix tier among those observed purchasing, of which there are a few hundred. This method is intuitively similar to a multistage estimation procedure - first estimate a binary choice Probit model of whether subscribing at all, yielding $X\beta$ and μ_1, and afterwards find the values of the thresholds for the 2^{nd} and 3^{rd} tiers that match the model's predicted shares choosing each tier to aggregate shares, based on a random subsample of observed purchases.

viduals observed not subscribing to Netflix, $i(j = 0)$ denotes the set of individuals subscribing to Netflix, but whose subscription tier choice is unobserved, and $i(j = k)$ denote the sets of individuals observed purchasing tier $k \in (1, 2, 3)$.

5 Counterfactual Simulations

This section simulates counterfactual environments in which Netflix implements first-degree price discrimination. Specifically, optimal profits and the dispersion of prices offered to different individuals are calculated separately, first using demographics alone and then all variables to explain a consumer's willingness to pay. They are then compared with simulated profits under the status quo environment, where second-degree PD was used.

5.1 Calculating Variable Profits

For a given price schedule offered to individual i, the firm's expected variable profit from that individual are:

$$\pi_i = \sum_{j=1}^{3} (P_{i,j} - c_j)\,(F(\theta_{i,j+1}) - F(\theta_{i,j})) \tag{10}$$

where c_j is the marginal cost of providing tier j service. The marginal costs and their values were described in section 1. Recall that $\theta_{i,j}$ are a function of price. Hence, $F(\theta_{i,j+1}) - F(\theta_{i,j})$ gives the probability individual i subscribes to tier j, conditional on price.[30]

Profit maximizing prices can be found via grid search, for both the cases where the firm does and does not tailor prices to individuals.[31] Increments of 5 cents were used. Unreported tests found reducing the increment size further yields similar profit estimates.

[30]In simulations, I require that the thresholds $\theta_{i,j}$ are weakly increasing in quality of the product tier, i.e. $\mu_j \geq \mu_{j-1}, \forall j$, guaranteeing that no tier is a strictly dominated choice (i.e. probability of subscription to each tier is ≥ 0). To ensure prices are such that this requirement is met, a lower bound price is set for each tier, conditional on the next lower tier's price. The lower bound of P_{j+1} is the lowest value satisfying:

$\mu_{j+1} = (P_{j+1} - P_j) * \lambda_{j+1} - \beta_0 \geq \mu_j$

$\Rightarrow P_{j+1} \geq (\mu_j + \beta_0)/\lambda_{j+1} + P_j$

[31]Computation was sped by grouping individuals with similar values of parameters, computing the variable profits from a prototypical individual in the group, and scaling up profits for the group by the number in the group.

5.2 Assignment of Unidentified Parameter

In order to simulate scenarios with counterfactual pricing, one must specify consumers' responsiveness to price, since it is not identified in data lacking price variation. Equation 7 shows that λ determines the rate by which $\theta_{i,j}$ changes with prices, and hence the slope of demand.[32] Rearranging the equation to solve for λ_j yields:

$$\lambda_j = \frac{\mu_j + \beta_0}{P_j - P_{j-1}} \tag{11}$$

Note in equation 11 that λ_j, the price parameter, is monotonically increasing in the value of the parameter β_0, whose value is not recovered from estimation. Hence higher β_0 imply strictly higher price sensitivities. This suggests that β_0 can be determined after estimation using supply side conditions, similar to Gentzkow [2007]. Once β_0 is known, all other parameters can be recovered.

Specifically, I assume Netflix has some pricing power, and estimate the value of β_0 which implies that observed prices are the prices which maximize Netflix's static profits. Formally, I search over β_0 to find the value of β_0 which minimizes the summed square of differences between observed prices for the tiers and simulated profit maximizing prices.[33] The resulting value, 0.624, yields a set of simulated prices that are close to observed prices, $[\$10.30, \$15.00, \$17.70]$ vs. $[\$9.99, \$14.99, \$17.99]$. Since the prices of the three tiers in simulations all depend on a single parameter β_0, it was not possible to find a value of β_0 matching all three prices exactly.

5.3 Counterfactual Results

Profits, prices, sales, and other outcome variables are simulated both under status quo pricing, i.e. second-degree PD, and under first-degree PD. This process is repeated twice, once using only demographics to predict willingness to pay, and once using the full set of variables.

Table 3 shows the percent increase in profits from tailoring prices to each individual, i.e. first-degree PD.[34] Using all variables to tailor prices, one can yield variable profits 2.14% higher

[32] Note the omission of the price sensitivity parameter α. α is not separately identified from λ_j and has been normalized to 1. This normalization is inconsequential, however, as only the ratio of these parameters matter for product choice - the ratio of α to λ_j is the coefficient on tier j price.

[33] In unreported tests, I found that the qualitative findings of the paper were not very sensitive to the supply-side estimate of the price sensitivity. Choosing another value of β_0 yielding optimal simulated prices more than twice those observed still yielded very similar results when expressed in percent changes.

[34] 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

than variable profits obtained using non-tailored second-degree PD. Using demographics alone to tailor prices raises variable profits by much less, yielding variable profits only 0.14% higher than variable profits attainable under second-degree PD.

The variable profits changes appear more striking when expressed in total profits. Netflix's annual variable profits would have increased by about \$8 million, or 12.18% of total profits, if Netflix had tailored price based on web browsing data.[35] If personalized prices were based only on demographics, the increase in total profits is much less, 0.79%. Since adding web browsing data substantially increases the amount by which first-degree PD raises profits, it increases the likelihood that firms will implement individually-tailored pricing.

Using the full set of variables to tailor prices substantially increases the range of prices charged to different individuals for the same product, and thus may impact whether the price distribution is perceived as fair. Figure 7 shows histograms of prices for the 1 DVD at-a-time tier. The figure includes overlaid histograms, one for person-specific prices when all variables are used to tailor prices, and another when only demographics are used. Clearly, a much wider range of prices results when all variables are used to individually-tailor prices.

Table 5 provides further details on the impact of tailored pricing on the distribution of prices offered to different individuals, separately for each subscription tier. When all variables are used, the consumer estimated to have the highest value for Netflix would face prices about 60% higher than prices faced when prices are not tailored to the individual. The 99.9^{th} percentile individual would face prices about 30% higher, the 99^{th} percentile about 17% higher, and the 90% percentile about 5% higher. The median consumer pays slightly less when prices are tailored, and the lowest offered prices are about 20% less than untailored prices. These results together imply that the highest prices offered would be roughly double the lowest prices offered, for the exact same good.

Since charging high prices to some individual might encourage entry of competitors or elicit a negative visceral response from consumers, firms may prefer instead to offer only targeted discounts and not raise prices to anyone. To investigate the profitability of this strategy, I re-optimize tailored prices setting an upper bound price for each tier equal to the tier's price under profit maximizing second-degree PD, $[\$10.30, \$15.00, \$17.70]$ for the three tiers respectively. The variable profits, sales, and aggregate consumer surplus under this strategy are shown in Table 4. The profit gain from tailored pricing is about three-quarters lower when the upper

of each other.

[35]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. 2.14% of \$370 million is about \$8 million, which is about 12.18% of \$65 million.

bound is imposed. As before, the profit gain from tailored pricing is much higher when prices are based on all variables rather than only on demographics (3.19% vs. 0.28%).

A pertinent question is whether first-degree PD substantially raises the fraction of surplus which is extractable by the firm. I find the answer is no - only about 42% of the theoretical maximum variable profits can be captured when prices are tailored based on web browsing history.[36]

This raises the question of how much prices would vary if the firm were better able to predict willingness to pay, which certainly may be possible with bigger and better datasets. Other data might include location by time of day, collected on smartphones via GPS, and textual variables derived from user-generated text on twitter, emails, and text-messages. Moreover, panel data may aid estimation of differential discount rates, which can be used to tailor couponing and bargaining strategies [Goldberg [1996]].

The model can be used to answer this question. Specifically, I assume that the model captures the true distribution of willingness to pay, and estimate the values of model parameters which would yield this same distribution, but would imply willingness to pay is more accurately estimated. Mechanically, I assume different values for the standard deviation of the error ϵ in the predicted value for quality y_i.[37] Lower standard deviations imply better estimates of y_i, but shrinking the error terms ϵ_i also changes the distribution of $y_i = X_i\beta + \beta_0 + \epsilon_i$, shrinking the range of willingness to pay. To offset this change, I rescale $X_i\beta + \beta_0$ about its mean until yielding approximately the same distribution of willingness to pay as the original model, but with lower standard deviation of ϵ.[38]

The results are shown in Figure 8, which plots various percentiles of prices offered to consumers for the first tier of service against the standard deviation of an individual's estimated willingness to pay for the tier. Plots for tiers 2 and 3 look similar. Price percentiles change roughly linearly in the standard deviation of willingness to pay, as least for medium-sized changes.

[36] Maximum possible profits, if willingness to pay were known exactly, is computed as follows. First, I draw a sample of the true underlying values of $y_i = \beta_0 + X\beta_i + \epsilon$. The optimal price to charge an individual for one tier in isolation sets their utility for the tier, shown in equation 1, equal to the utility of the outside good $\alpha_i I_i$. Solving yields: $P_j = y_i q_j$. Variable profits from each simulated individual are then computed as the maximum of the profits across tiers j for that individual.

[37] This can equivalently be accomplished by multiplying all model parameters by the inverse of the standard deviation of the error and leaving the scale of the error term unchanged.

[38] A wider price grid was used in these simulations. To speed computation the increments between grid points were increased as well, to $0.25.

6 Discussion and Conclusion

Many purchase categories may soon, or already do, use personalized pricing. Some firms selling food and clothing already do, and education obviously does [Elliot [2013], Gross [2012], Thau [2014]]. Recreation consumption seems an obvious candidate, and vehicle-related sales easily can more effectively first-degree PD, by linking identifying information required for vehicle purchases and registration with behavior datasets. A conservative estimate with just these categories suggests about a third of all consumption might be ripe for personalized pricing.[39]

This paper finds, in one context, that the increase in profits made feasible by first-degree PD is much higher when web browsing behaviors (12.2%), rather than just demographics (0.8%), are used to predict individuals' reservation values. This meaningful profit increase made possible by web browsing data supports the argument that first-degree PD is evolving from merely theoretical to practical and widely employed. This will directly impact consumers, as consumer surplus is lower, and the estimated 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 first-degree PD . Most textbooks espouse its efficiency based on partial equilibrium analysis. However, when employed by multiple firms, this result may not hold. In oligopolistic [Spulber [1979]] and differentiated product [Thisse and Vives [1988]] 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 first-degree PD was viewed as unfair by 91% of respondents.

Much research is left to be done. To my knowledge, research on general equilibrium models with widespread first-degree PD and rich heterogeneity is non-existent and greatly needed. There likewise is much room for more empirical work on the subject.

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

[39] BEA finds the following consumption shares: transportation (10.0%), recreation (8.7%), food (7.7%), clothing (3.3%), and education (2.5%). $http://www.bea.gov/iTable/iTableHtml.cfm?reqid = 9\&step = 3\&isuri = 1\&903 = 74$

References

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

Robert M Bell, Yehuda Koren, and Chris Volinsky. All together now: A perspective on the netflix prize. *Chance*, 23(1):24–29, 2010.

Ronald H Coase. Durability and monopoly. *Journal of Law and Economics*, 15(1):143–149, 1972.

Stuart Elliot. For cvs regulars, ads tailored just to them. *New York Times*, 2013.

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.

Matthew Gentzkow. Valuing new goods in a model with complementarity: Online newspapers. *The American Economic Review*, 97(3):713–744, 2007.

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

Pinelopi Koujianou Goldberg. Dealer price discrimination in new car purchases: Evidence from the consumer expenditure survey. *The Journal of Political Economy*, 104(3):622–654, 1996.

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.

Ashley Gross. Worth the deal? groceries get a personalized price. *NPR Online*, 2012.

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.

David Madigan and Adrian E Raftery. Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, 89(428):1535–1546, 1994.

Alexis Madrigal. I'm being followed: How google - and 104 other companies - are tracking me on the web. *The Atlantic*, 2012.

Jonathan Mayer. Tracking the trackers: Early results. *The Center for Internet and Society's Blog*, July 2011.

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.

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.

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.

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.

David Streitfeld. On the web, price tags blur: What you pay could depend on who you are. *Washington Post*, 2000.

Barbara Thau. How big data helps stores like macy's and kohl's track you like never before. *Forbes*, 2014.

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.

Robert Wilson. Nonlinear pricing. *Oxford University Press*, 1993.

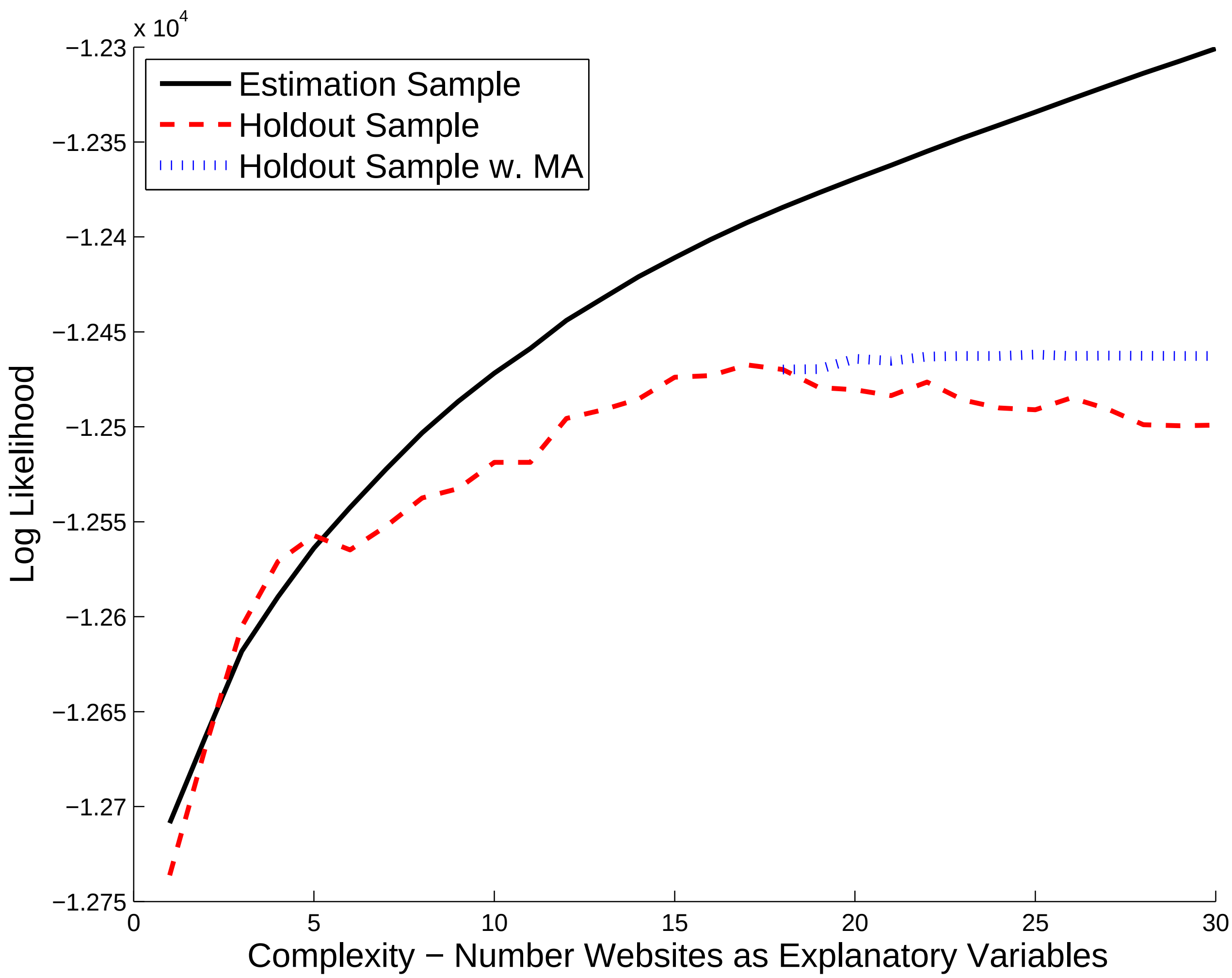

Figure 1: Model Selection - Likelihood vs. Number Websites Included as Explanatory Variables

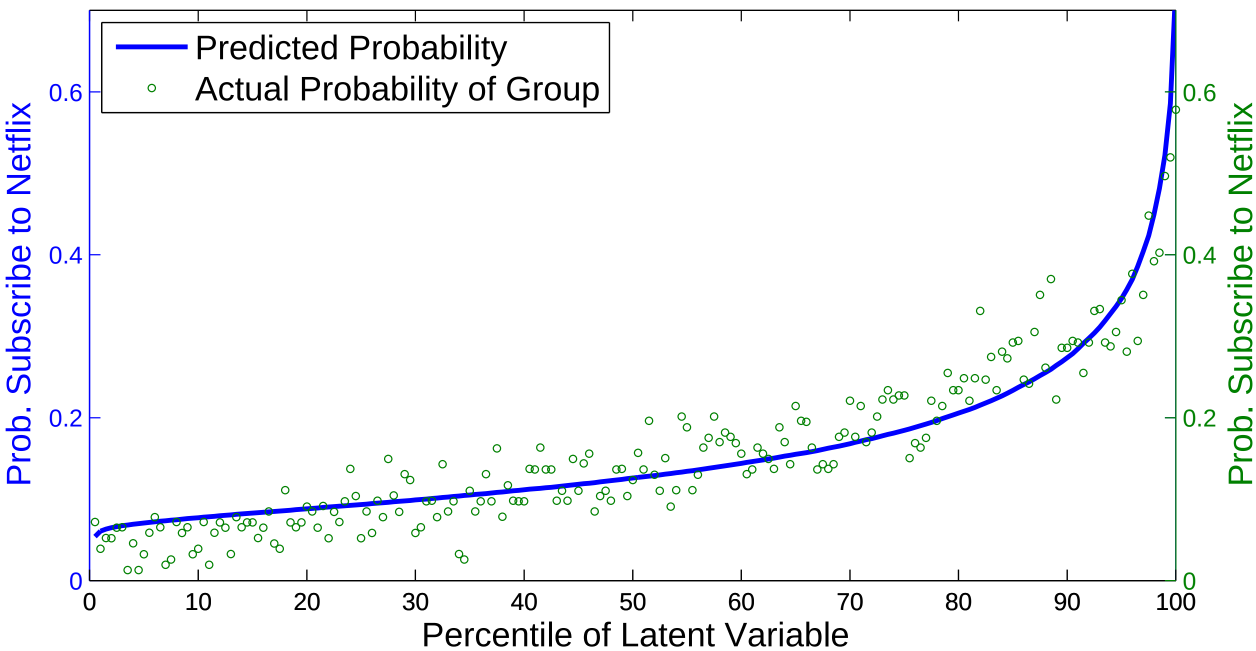

Figure 2: Model Fit - Predicted Probilities in Holdout Sample When All Variables Used

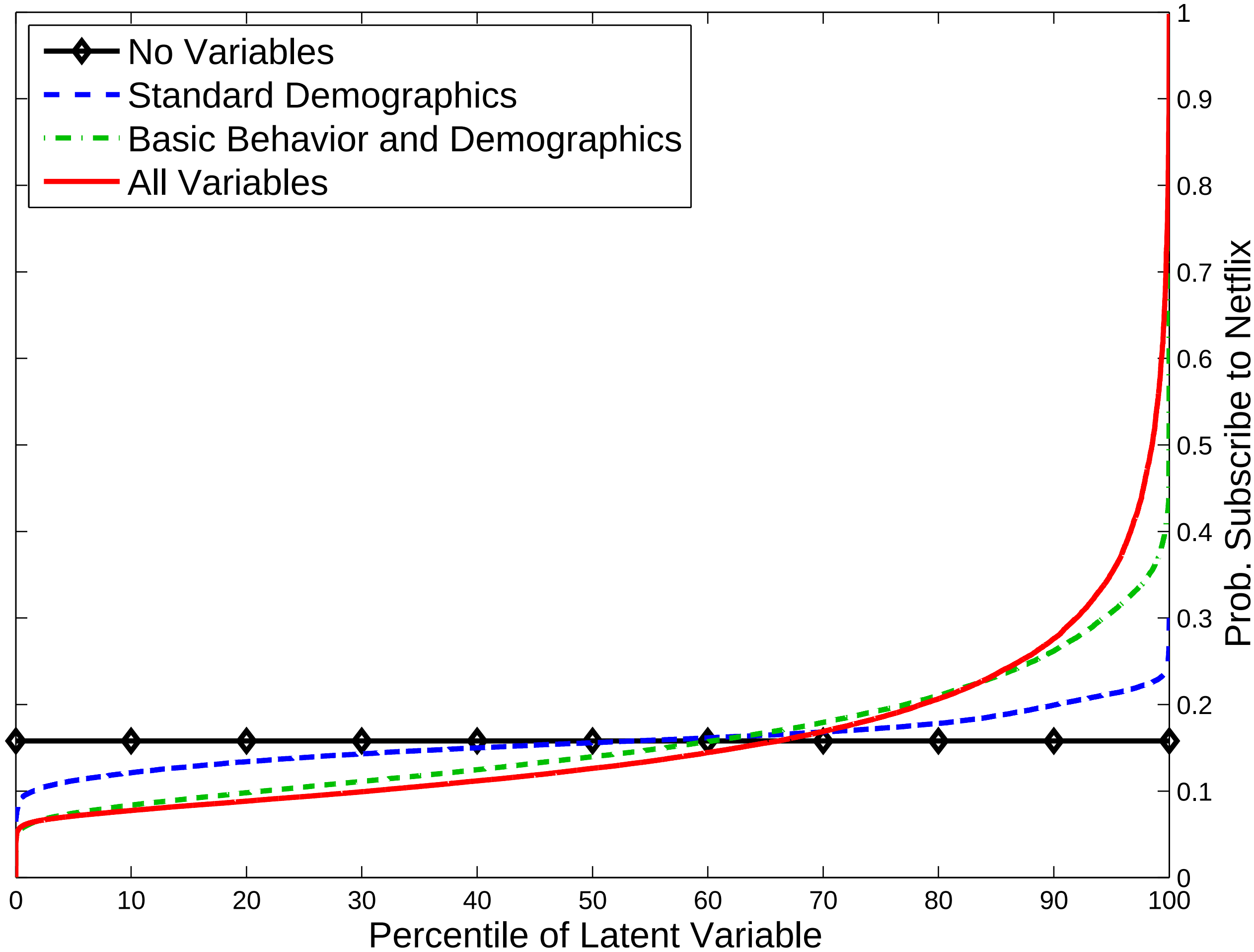

Figure 3: Range of Predicted Probabilities, Using Various Sets of Explanatory Variables

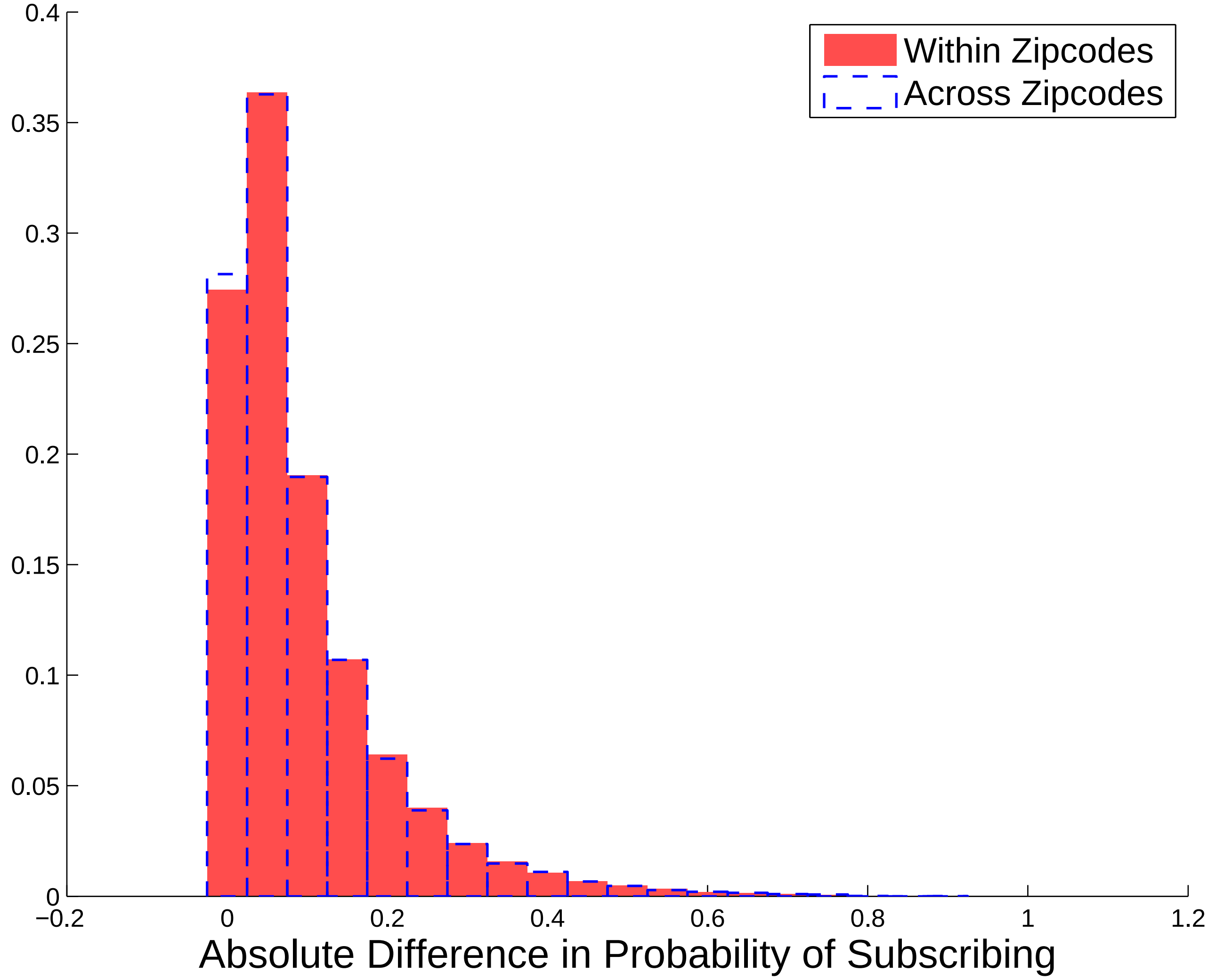

Figure 4: Histograms of Absolute Differences in Probabilities of Subscribing to Netflix, in Randomly Drawn Pairs of Individuals, Both Within and Across Zipcodes

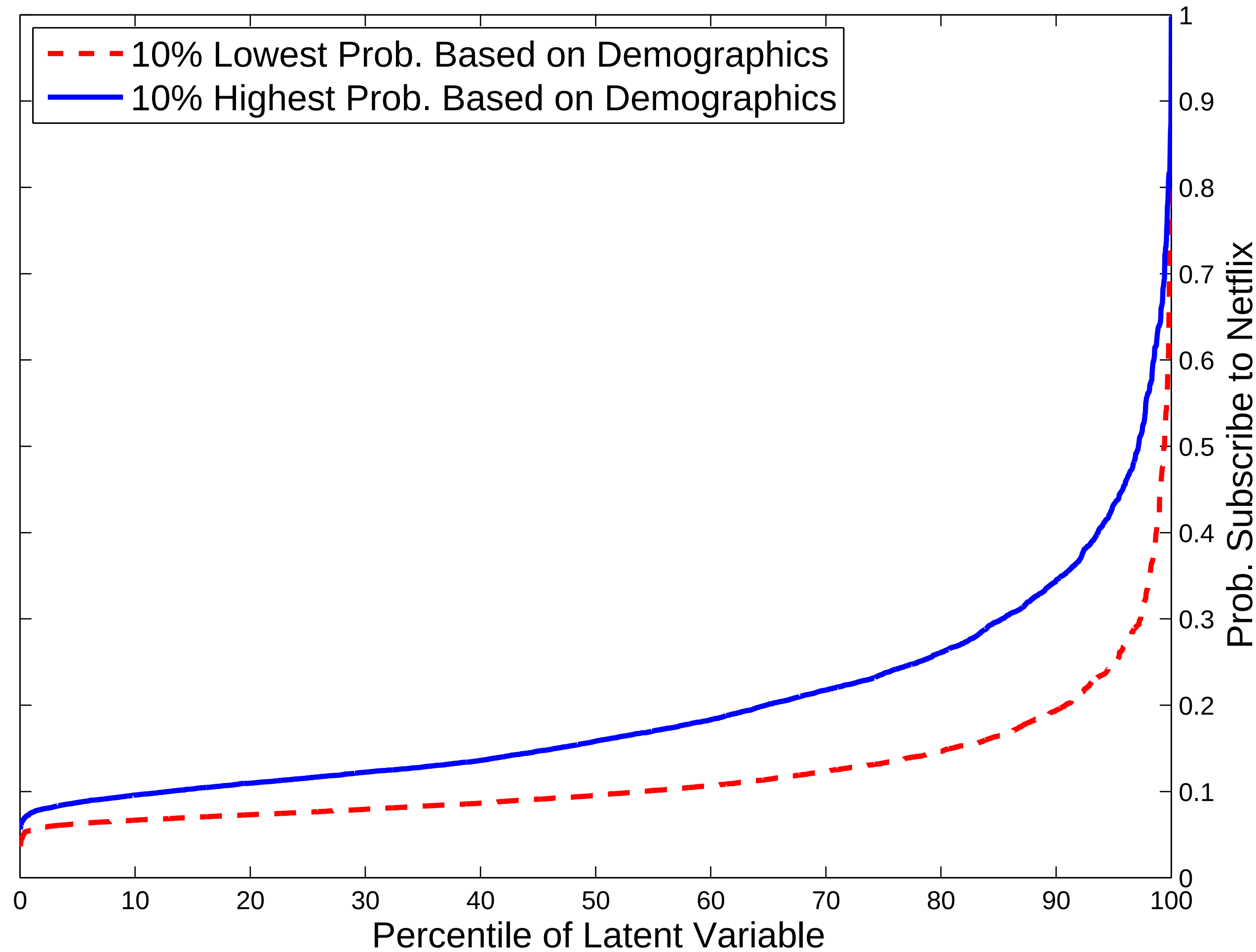

Figure 5: Range of Predicted Probabilities For Subsets of Individuals

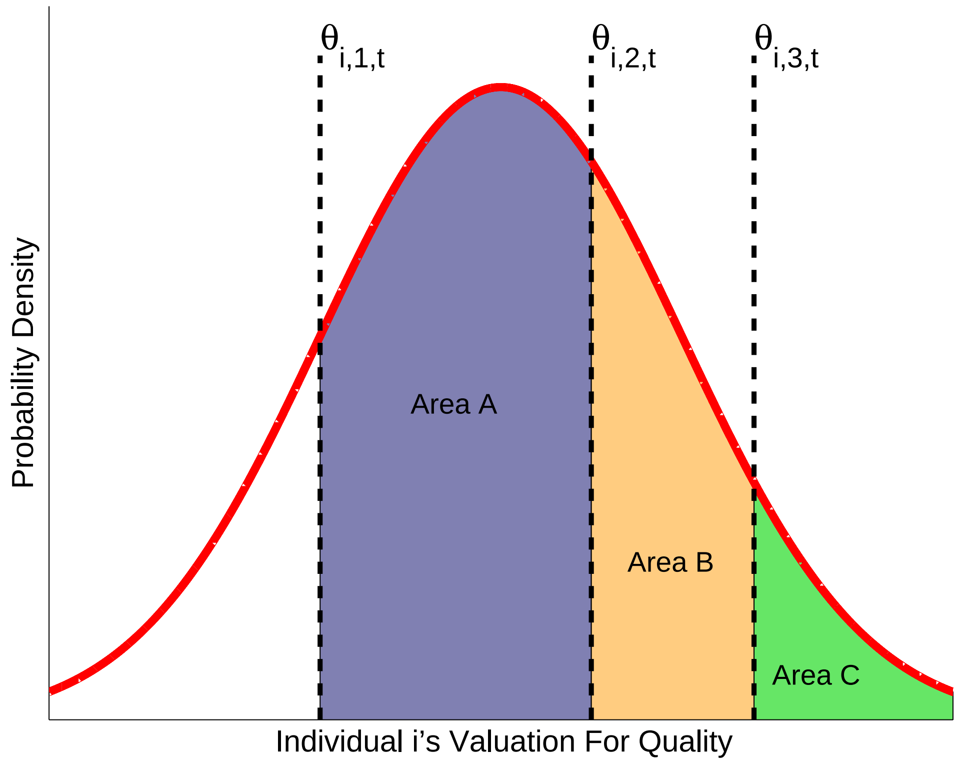

Figure 6: Graphical Depiction of Model

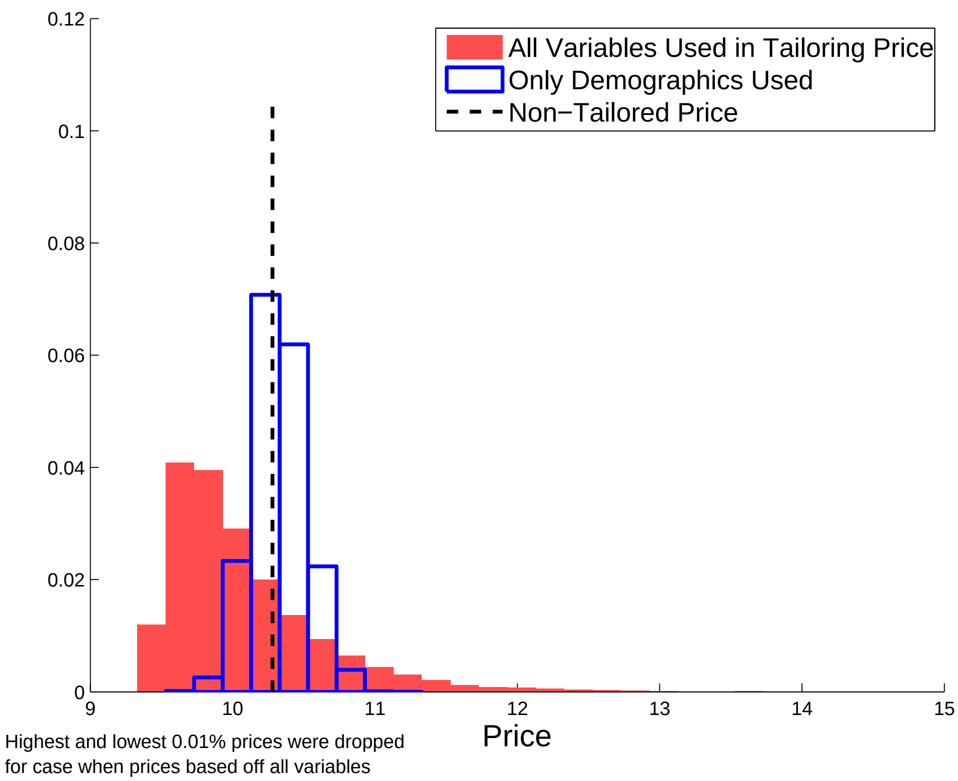

Figure 7: Histogram of Individually-Tailored Prices - 1 DVD At-a-Time Plan

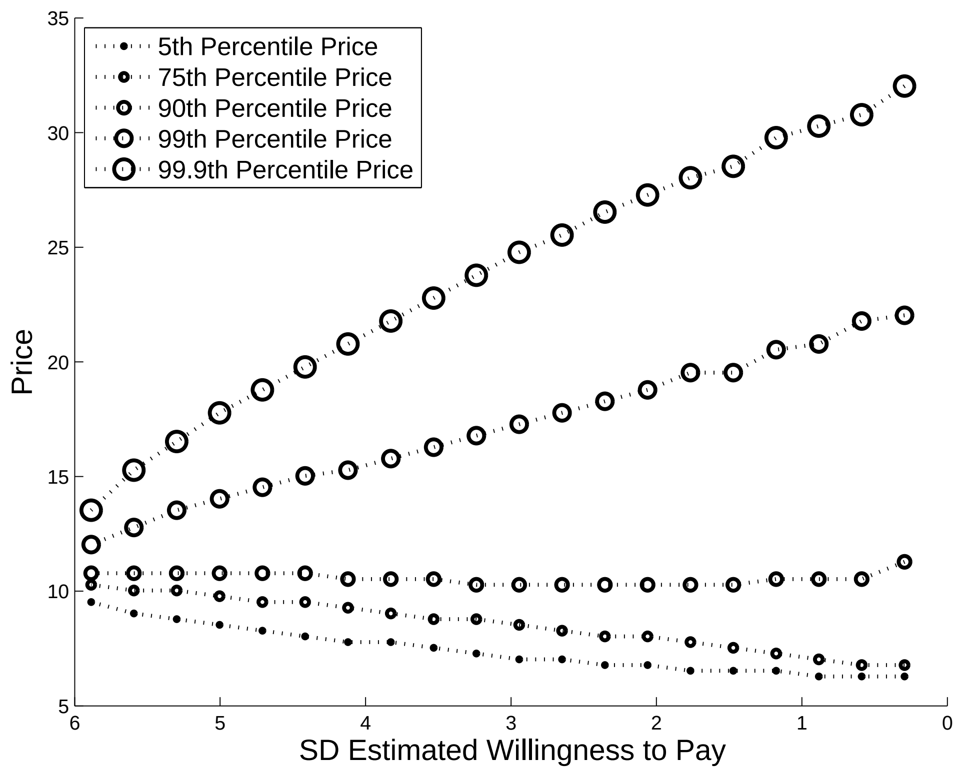

Figure 8: Graphical Depiction of Dispersion of Prices - 1 DVD At-a-Time Plan

Table 1: Binary Choice Model Results

Variable Name†	Demographics	Demographics and Basic Behavior
Age Oldest Household Member	-0.046	-0.032
	(0.000)	(0.000)
Census N Central Region	-0.041	-0.024
	(0.000)	(0.000)
Census South Region	-0.029	
	(0.000)	
Census West Region	0.049	0.062
	(0.000)	(0.000)
Black Indicator	-0.035	-0.028
	(0.000)	(0.000)
Hispanic Indicator	-0.065	-0.024
	(0.000)	(0.000)
Household Income Range Squared	0.020	
	(0.000)	
Household Size Range	0.023	
	(0.000)	
Population Density (Zipcode)	0.021	
	(0.000)	
Total Website Visits		0.398
		(0.000)
Broadband Indicator		0.050
		(0.000)
Total Website Visits Squared		-0.216
		(0.000)
% of Web Use on Tuesdays		-0.024
		(0.000)
% of Web Use on Thursdays		-0.037
		(0.000)
# Unique Transactions		0.023
		(0.000)
N	30,642	30,642
LL	-13,246.403	-12,797.706

Standard errors, in parentheses, computed via likelihood ratio test. † All variables normalized to have zero mean and standard deviation equal to one.

Table 2: Websites Best Predicting Netflix Subscription (All Are Positive)

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

Table 3: Simulated Changes in Various Outcomes Resulting From First-Degree PD

	Percent Change When Price Based on:	
	Demographics	All Variables
Total Profits	0.79%	12.18%
	(0.11)	(0.68)
Variable Profits	0.14%	2.14%
	(0.02)	(0.12)
Sales (DVDs At-a-Time)	0.85%	2.67%
	(1.09)	(1.23)
Subscribers	0.17%	1.65%
	(0.62)	(0.84)
Aggregate Consumer Surplus	−0.18%	−7.75%
	(1.12)	(1.38)

Bootstrapped standard errors in parentheses. To speed computation simulations in bootstrapping employed a 10 cent increment in price gridsearches.

Table 4: Simulated Changes in Various Outcomes Resulting from Tailored Discounts Off Optimized Second-Degree PD Prices

	Percent Change When Discounts Based On:	
	Demographics	All Variables
Total Profits	0.28%	3.19%
	(0.17)	(0.51)
Variable Profits	0.05%	0.56%
	(0.03)	(0.09)
Sales (DVDs At-a-Time)	2.56%	8.11%
	(0.71)	(0.89)
Subscribers	2.53%	7.61%
	(0.22)	(0.55)
Aggregate Consumer Surplus	3.38%	8.17%
	(0.57)	(0.64)

Bootstrapped standard errors in parentheses. To speed computation simulations in bootstrapping employed a 10 cent increment in price gridsearches.

Table 5: Percent Difference Between Individually-Tailored Prices and Non-Tailored Prices

Price	1 DVD Plan		2 DVDs Plan		3 DVDs Plan	
Percentile	Demog.	All Vars.	Demog.	All Vars.	Demog.	All Vars.
Lowest	−6.8%	−22.4%	−6.3%	−21.3%	−6.2%	−20.6%
	(0.6)	(5.7)	(0.7)	(5.4)	(0.7)	(5.2)
0.1	−5.8%	−9.2%	−5.7%	−9.0%	−5.6%	−8.8%
	(0.6)	(0.5)	(0.6)	(0.5)	(0.6)	(0.5)
1	−4.4%	−8.3%	−4.0%	−8.0%	−3.9%	−7.9%
	(0.4)	(0.3)	(0.5)	(0.3)	(0.5)	(0.5)
10	−2.4%	−7.3%	−2.3%	−7.0%	−2.3%	−6.8%
	(0.5)	(0.4)	(0.4)	(0.5)	(0.5)	(0.6)
25	−1.5%	−5.8%	−1.3%	−5.7%	−1.4%	−5.7%
	(0.5)	(0.5)	(0.5)	(0.5)	(0.6)	(0.5)
50	−0.5%	−3.4%	−0.3%	−3.3%	−0.6%	−3.4%
	(0.5)	(0.5)	(0.5)	(0.6)	(0.5)	(0.6)
75	0.5%	0.0%	0.7%	−0.3%	0.6%	−0.6%
	(0.5)	(0.4)	(0.5)	(0.5)	(0.5)	(0.5)
90	2.4%	4.9%	2.3%	4.3%	2.0%	4.0%
	(0.3)	(0.3)	(0.4)	(0.5)	(0.5)	(0.6)
99	3.9%	18.0%	3.6%	17.0%	3.4%	16.1%
	(0.6)	(0.8)	(0.6)	(0.9)	(0.7)	(0.9)
99.9	5.3%	32.6%	5.0%	30.7%	4.8%	29.4%
	(0.9)	(2.0)	(1.0)	(2.0)	(1.0)	(2.0)
Highest	7.7%	61.3%	7.3%	58.0%	7.0%	55.9%
	(1.7)	(5.6)	(1.7)	(5.5)	(1.7)	(5.4)

Bootstrapped standard errors in parentheses. To speed computation simulations in bootstrapping employed a 10 cent increment in price gridsearches.

A Heterogeneous Price Sensitivity Model

When the data include time-varying prices, one can estimate a more flexible version of the model, which allows for heterogeneous price sensitivities. Details are below.

The conditional indirect utility that consumer i receives from choosing nondurable product j in period t is:

$$u_{i,j} = y_i q_j + \alpha_i \left(I_i - P_{j,t} \right) \tag{12}$$

where q_j is the quality of product j, $P_{j,t}$ is its price in period t, and I_i is the income of individual i. The products are indexed in increasing order of quality. I.e. if $j > k$, then $q_j > q_k$. The parameters y_i and α_i are person-specific parameters that reflect individual i's valuation for quality and price sensitivity, respectively. This utility specification is similar to Mussa and Rosen (1978), but allows for differences across consumers in price sensitivity α_i.

For consumer i to weakly prefer product j to product k, the following incentive compatibility constraint must hold:

$$y_i q_j + \alpha_i \left(I_i - P_{j,t} \right) \geq y_i q_k + \alpha_i \left(I_i - P_{k,t} \right) \tag{13}$$

If q_j is greater than q_k, this reduces to:

$$y_i \geq \alpha_i \frac{P_{j,t} - P_{k,t}}{q_j - q_k} \tag{14}$$

If $\frac{P_{j,t} - P_{k,t}}{q_j - q_k}$ is strictly increasing in j, then no quality tier is a strictly dominated choice for all possible values of y_i. In that case, only the incentive compatibility constraints for neighboring products bind, and we can use equation 14 to yield a range of y_i required for individual i to buy each tier j. Specifically, a consumer i chooses product j if and only if the following inequality condition is satisfied:

$$\alpha_i P_{\Delta j,t} \lambda_j \leq y_i < \alpha_i P_{\Delta (j+1),t} \lambda_{j+1} \tag{15}$$

where $P_{j,t} - P_{j-1,t}$ and $(q_j - q_{j-1})^{-1}$ have been replaced by the notation $P_{\Delta j,t}$ and λ_j, respectively.

Next, the variables y_i and α_i in the above inequality condition are replaced with linear regression expressions, $\beta_0 + X_i \beta + \sigma \epsilon_{i,t}$ and $\gamma_0 + X_i \gamma$, respectively. The parameter vectors γ and β reflect differences across consumers explainable with the data. The above inequality with these changes is:

$$(\gamma_0 + X_i\gamma)P_{\Delta j,t}\lambda_j \leq \beta_0 + X_i\beta + \sigma\epsilon_{i,t} < (\gamma_0 + X_i\gamma)P_{\Delta(j+1),t}\lambda_{j+1} \tag{16}$$

A couple of normalizations are required. First, σ, the standard deviation of the error term, is not separately identified from the scaling of the remaining parameters in the model. As is standard in ordered-choice models, it is normalized to 1. Second, γ_0 cannot be separately identified from the scaling of quality levels, λ_j, so γ_0 is also arbitrarily normalized to 1. Incorporating these changes, and rearranging yields:

$$\theta_{i,j,t} \leq \epsilon_{i,t} < \theta_{i,j+1,t} \tag{17}$$

where

$$\theta_{i,j,t} = -\beta_0 - X_i\beta + P_{\Delta j,t}\lambda_j + X_i P_{\Delta j,t}\phi_j \tag{18}$$

and where the parameter vector $\phi_j = \lambda_j\gamma$.

Finally, the probability that product j is consumed by individual i equals:

$$s_{i,j,t} = F\left(\theta_{i,j+1,t}\right) - F\left(\theta_{i,j,t}\right) \tag{19}$$

where $F()$ is the CDF of ϵ. The probabilities $s_{i,j,t}$ can subsequently be used in maximum likelihood estimation.