# *First Degree Price Discrimination Using Big Data*

*Benjamin Reed Shiller, Economics Department, Brandeis University*

**Working Paper Series**

2013 | 58

# First Degree Price Discrimination Using Big Data

Benjamin Reed Shiller
Department of Economics, Brandeis University

August 20, 2013

**Abstract**

Second and $3^{rd}$ degree price discrimination (PD) receive far more attention than $1^{st}$ degree PD, i.e. person-specific pricing, because the latter requires previously unobtainable information on individuals' willingness to pay. I show modern web behavior data reasonably predict Netflix subscription, far outperforming data available in the past. 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 methods for handling massive datasets. Simulations show using demographics alone to tailor prices raises profits by 0.14%. Including web browsing data increases profits by much more, 1.4%, increasingly the appeal of tailored pricing, and resulting in some consumers paying twice as much as others do for the exact same product.

# 1 Introduction

First degree price discrimination (PD), dating back to at least Pigou [1920] in the literature, theoretically allows the firm to extract full surplus. Yet, the empirical literature instead focuses on other forms of price discrimination, which have been found to allow far less surplus extraction, about a third in studied contexts.[1,2] These other papers implicitly assume that first degree PD is infeasible - firms do not have information on willingness to pay at the individual level.[3] Moreover, orthodox instruction uses this argument to motivate $2^{nd}$ degree and $3^{rd}$ degree PD. While sound historically, this argument may no longer hold. Large datasets on individual behavior, popularly referred to as "big data," are now readily available, and contain information potentially useful for person-specific pricing.[4] For example, web browsing data may indicate psychographic profiles or direct interest in a related product, or reflect latent demographics such as sexual orientation, social phobia, and marital happiness - all information that can be used to form a hedonic estimate of willingness to pay.[5] In this paper, I investigate the extent of incremental information contained in web-browsing data, the profitability of first degree PD, and the resulting distribution of prices different consumers are offered when purchasing the exact same item.

Netflix provides an auspicious context. First, since purchases occur online, Netflix could offer tailored prices, as a couple of other online sellers have tried [Mikians et al. [2012]]. Second, Netflix could effectively price discriminate - its products were differentiated from competitors' products implying pricing power, and arbitrage appears costly enough given the fact that Netflix has long employed another form of price discrimination, $2^{nd}$ degree PD. Last, focusing on Netflix overcomes estimation problems faced by researchers, since Netflix subscription can be imputed in a dataset that also includes browsing behavior variables to tailor pricing.

This paper overcomes obstacles in incorporating large numbers of explanatory variables. Potential problems include insufficient degrees of freedom, overfitting, tractability and convergence issues, and computer memory limitations. Missing data can also be problematic - with many variables there may be few observations with non-missing values for all variables. Model averaging overcomes or mitigates these problems, but without further steps yields biased results in binary and ordered choice models. To address this problem, I show an Ordered choice Model Averaging (OMA) method, a very simple solution to this problem which can be used in standard

[1]Recent empirical papers focus on $3^{rd}$ degree PD [Graddy [1995], Graddy and Hall [2011], Langer [2011]], $2^{nd}$ degree PD [Crawford and Shum [2007], McManus [2008]], intertemporal pricing [Nair [2007]] and bundling [Chu et al. [2011], Shiller and Waldfogel [2011]]

[2]Shiller and Waldfogel [2011] find that bundling, nonlinear pricing, and $3^{rd}$ degree PD cannot extract as profits more than about third of surplus in the market for digital music.

[3]It is worth noting that tailoring prices to consumers is not currently per se illegal, as is evident from widespread use of $3^{rd}$ degree PD. Tailoring prices to downstream firms is however prohibited under the Robinson-Patman Act.

[4]Madrigal [2012], Mayer [2011] note that web-browsing behavior is collected by hundreds of firms.

[5]Psychographic profile categorizes individuals based on attitudes, activities, values, and behavior.

statical programs.[6]

This method is used to determine how useful different sets of variables are in estimating the probability 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 91%.

Next, an empirical model is used to translate the increased precision from web-browsing data into key outcome variables. Specifically, a model derived from canonical quality discrimination theory models is used to estimate demand for Netflix in the observed environment, in which Netflix employed $2^{nd}$ degree PD, but not $1^{st}$ degree PD. The model is then used to simulate pricing and profits in the hypothetical counterfactual occurring if Netflix had implemented $1^{st}$ degree PD.

I find that web browsing behavior substantially raises the amount by which person-specific pricing raises variable profits relative to $2^{nd}$ degree PD - 1.39% if using all data to tailor prices, but only 0.14% using demograhpics alone. Web-browsing daa make $1^{st}$ degree PD more appealing to firms and likely to be implemented, thus impacting consumers. Substantial equity concerns may arise - I find some consumers may be charged twice as much as others are for the same product.

The closest literature, a strand 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.[7] However, they assumed that consumers were myopic. Anecdotal evidence following Amazon's pricing experiment in the early 2000s suggests otherwise [Streitfeld [2000]]. Acquisti and Varian [2005], Fudenberg and Villas-Boas [2005] show theoretically that $1^{st}$ degree PD 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."

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 higher prices to individuals that use the internet during the day on Tuesdays and Thursdays, and visit Wikipedia.org, patterns consumers may not recognize. Moreover, with many variables, there may not be any easy heuristics consumers

[6]The method works on both binary and ordered choice models

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

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 data, providing the incentive for consumers to reveal them.

The remainder of the paper is organized as follows. Section 2 describes the context and industry background. Next, Section 3 describes the data. Section 4 then shows how well various sets of data explain propensity to purchase. Lastly, Sections 5 and 6 present a model and estimate optimal person-specific 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 [net [2006]]. This implies that about 16.7% of internet connected households consumed Netflix during 2006.[8]

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.[9] Moreover, Netflix's customer acquisition algorithm was well known, further differentiating their services.

Netflix's subscriptions plans can be broken into two categories. Unlimited plans allow consumers to receive an unlimited number of DVDs by mail each month, but restrict the number of DVDs in a consumer's possession at one time. Limited plans set both a maximum number of DVDs the consumer can possess at one time, and the maximum number sent in one pay 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.[10] None of the plans allowed video streaming, since Netflix did not launch that service until 2007 [net [2006]].

[8]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.

[9]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 [net [2006]].

[10]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.

Key statistics for later analyses are Netflix's marginal cost 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.[11]

# 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 pages viewed on that website, the referring website, and details on any transactions. 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 inputed 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 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 points of the estimated share of U.S. internet connected housedholds subscribing, presented in Section 2. 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 other dubious categories, leaving $4,789$ popular websites to calculate additional variables.[12] The total number of visits to all websites and to each single website were computed

[11] 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 net [2006] for further details.

[12] yoyo.org provides a user-supplied list of some websites of dubious nature. Merging this list with the ComScore

for each computer user. Similar variables were computed for transactions.

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, percent of online browsing by time of day and by day of week, and broadband indicator. Detailed web-behavior contain variables indicating number visits to a particular website.

## 4 Prediction in Status Quo

This section predicts 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 across these sets of explanatory variables, to inform on the the relative benefit of including web browsing behavior.

First, a Probit model is used to investigate which standard demographic variables are significant predictors of a 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.

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 were no longer significant once basic web behavior variables were added, suggesting they were 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, number visits to each particular website are added one at-a-

data reveal that such websites tend to have very high ($\geq 0.9$) or very low ($\leq 0.1$) rates of visits that were referred, 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.

[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.

time to the significant demographic and basic web behavior explanatory variables. Overall, 29% of websites were significant at the 5% level, and 18% at the 1% level, far more than expected by chance alone. This suggests that for most the effect is causal, rather than a type I error. The twenty websites which best explain Netflix subscription are shown in Table 2. All twenty were positive predictors. Inspection reveals they are comprised of websites which are likely used by movie lovers (IMDB, Rotten Tomatoes), internet savvy users (Wikipedia), those preferring mail ordering (Amazon, Gamefly), and discount shoppers (Bizrate, Price Grabber).

I next investigate the joint prediction of all website variables combined, rather than just considering one at-a-time. Model averaging, averaging many smaller models together to yield a final estimate, is used to overcome several problems common when using data with many explanatory variables.

Problems common when including many explanatory variables, which are addressed by model averaging, are as follows. First, there may not be enough degrees of freedom, preventing estimation. This is especially problematic when interaction effects or higher order terms are included. A second problem is overfitting, leading to biased estimates. Even with many observations, overfitting can occur if errors are not independent. Third, large models may be prone to convergence problems or exceeding computer memory limits. Missing data may also be problematic. Most observations may have a missing value for at least one variable, leaving few observations with all nonmissing values to estimate the model.

Model averaging proceeds as follows. First, the set of explanatory variables $X$ is divided in two. Label these sets $Z_1$ and $Z_2$. $Z_1$ is the set of variables that are deemed by the econometrician to have a high likelihood of importance, in this case, demographics and basic web behavior variables. The variables in $Z_1$ that are significant when $Z_2$ are excluded will be included in all models. Call this set $Z_1'$. The set $Z_2$ includes variables indicating number visits to each website separately, referred to as the detailed web behavior variables. 50,000 subsets $s$ of five variables in $Z_2$ are drawn. The model is re-estimated adding each subset $s$ to $Z_1'$, dropping any variables in $s$ which are not significant.[14]

Each set of variables $(Z_1', s)$ yields its own estimates of the expected value of the latent variable for each individual and the threshold in the Probit model, as well as a value of the maximized likelihood. Taking a weighted average of these values across models enables predictions based on the information contained in all models together.

Without further steps, model averaging yields biased results in binary choice models. It over-predicts the probability of subscription for those least likely to subscribe, and under-predicts

[14]To increase computational efficiency, rather than including all demographic and basic web behavior variables, I rather summarize their contribution by including a single variable $\hat{y}_i$ equal to the value of $X\beta$ when only they are included. Replacing the individual variables with this single variable speeds computation, but restricts the model's ability to estimate marginal effects of a single variable in $Z_1$ conditional on the value of a variable in $Z_2$.

the probability for those most likely to subscribe. The reason is as follows. First, recall that in this binary choice model a consumer buys only if the true value of an underlying latent variable exceeds some threshold. The probability that a consumer subscribes thus depends on how far the estimate of the latent variable is from the threshold, and the accuracy of the estimate of the latent variable, i.e. its error's standard deviation. More extreme probabilities are reached only if the scaling of the latent variable is increased or the standard deviation of the error term is reduced. But neither can occur. The standard deviation of the error term is typically normalized to a value of one, since it is not separately identified in such models. Mechanically, the latent variable's scaling does not increase either - averaging many values does not change the expected value. If anything, averaging reduces the range of latent variables as extreme values yielded in one model by chance are mitigated by averaging.

A simple analogy helps explain the problem that is occurring. Suppose several independent medical tests for a disease all come back positive. If each test alone implies the probability an individual has the disease is 0.85, then multiple independent positive tests should together imply a probability over 0.85. In an Probit model framework, this 0.85 probability can be represented by a value of the latent variable equal to approximately 1, a threshold of 0, and standard deviation of the latent variable equal to 1. Averaging over multiple tests will yield the exact same values of the threshold and latent variable, and hence the same predicted probability as any single test, 0.85.

This bias can easily be corrected with one additional step after averaging the ensemble of models, a method that can be labeled as Ordered choice Model Averaging (OMA). As a final step, another ordered choice model is run with a single explanatory variable which equals the weighted average of the expected value of the latent variable for the individual across models. If the parameter $\kappa$ on this variable is greater than 1, then the scaling of the latent variable estimates increase. This broadens the range of estimated latent variables across individuals, relative to the standard deviation of the error, allowing for more extreme probabilities, and incorporating the information from many models.

The averaged values of the latent variable and threshold in the Probit model, used in the OMA method, depend on the weights placed on each model when computing the average. These weights are estimated as follows. Following the intuition from Occam's razor and earlier literature [Raftery et al. [1997]], the least likely set of models were excluded. Specifically, the top 2% most likely models were kept, and then averaged in relation to a function of the increase in the log likelihood obtained by their inclusion in the holdout sample.[15,16] Specifically, I parameter-

[15]For linear regression models, the expected likelihood of each model is typically found by integrating over the values of each parameter in the model using an analytic expression. For choice models, no such analytic expression exists. To speed estimation, I used the maximized likelihood instead.

[16]This implicitly assumes an uninformative prior - ie. ex-ante all models are assumed equally likely. Pre-existing information on model likelihood can easily be incorporated by changing this assumption.

ized the weights used in averaging as $\frac{e^{\omega*(LL_{z_1',s}-LL_{z_1'})}}{\sum e^{\omega*(LL_{z_1',s}-LL_{z_1'})}}$, where $LL_{z_1',s}$ and $LL_{z_1'}$ denote the log likelihood of model when the set $(z_1', s)$ and $(z_1')$ are include as possible explanatory variables, and $\omega$ is a parameter that determines the relative weights of the most likely to less likely models. The value of $\omega$ was chosen maximize the probability of the data in the holdout sample. Its value, 0.017, implies that most of the weight is on a small number of models. Nearly half the weight falls on one model, about 70% on the top two, 95% on the top 15, and 99% on the top 50.

Figure 1 shows the predictions from model averaging in the holdout sample. Specifically, individuals in the holdout sample are ordered according to their estimated value of the latent variable in the ordered choice model, then grouped. The average predicted probability and observed probabilities are then calculated for each group. Notice that these predicted probabilities, shown in solid blue line, do in fact seem to follow the actual probabilities of subscription.

The main takeaway from this section is summarized in Figure 2. It plots the predicted probability each individual subscribes based on various sets of explanatory variables together on one graph. Note the Y-axis range is larger than in Figure 1, which averaged predicted probability within groups, obscuring extreme probabilities. Including web behavior variables does in fact seem to substantially help prediction. Predicted probabilities of subscription ranged from $5.9 * 10^{-11}$ percent to 91% 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% chance.

Figure 3 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. Demographics predict this probability ranges between 6.5% and 12%. Predictions for this same group based on the full set of variables yield probabilities as high as 75%. 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. Hence, demographics used alone grossly misclassifies some individuals as low or high probability subscribers.

### 4.1 Discussion of Causality

The above results use correlations between explanatory variables and Netflix subscription status to infer individuals' tendencies to value Netflix highly. It is assumed that the web behavior is revealing underlying traits which correlate with valuations for Netflix. E.g. having a strong affinity for movies both causes a consumer to on average like viewing celebrity gossip websites and makes them more likely to consume Netflix, hence their correlation. However, an obvious concern is that this correlation exists for another reason.

An alternative story is that Netflix advertises only on particular websites, and visits to such websites cause consumers to buy Netflix, rather than revealing an underlying individual trait that correlates with valuing Netflix highly. However, some of the best-explaining websites shown in Table 2, Amazon and Wikipedia, do not advertise and hence are not consistent with this concern. Moreover, as explained earlier, the best-explaining websites seem intuitively to appeal to individuals with underlying traits which also would likely cause them to value Netflix highly. Even if this alternative explanation holds, it may not pose a problem. Since most consumers were aware of Netflix at that time, it would appear that advertising is persuasive rather than informative, i.e. seeing an advertisement raises a consumer's underlying value for the product. If advertising is persuasive, then the results are still accurate - prices can be viewed as optimal in the presence of the chosen level of advertising.

# 5 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 each tier, but may differ in how much they value the quality of higher tiers. Firms in the models set prices of the tiers of service, but not qualities.[17]

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 endogeneity concerns, but require more restrictive 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.

## 5.1 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 $\alpha$ is

[17] In the canonical $2^{nd}$ 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. consumer cannot rent half a DVD.

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:

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

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, only the incentive compatibility constraints for neighboring products bind, and consumer $i$ chooses product $j$ if and only if the following inequality condition is satisfied.[18]

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

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 $\lambda_j$, 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, $\sigma$, 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, $\alpha$ cannot be separately identified from the scaling of quality levels, $\lambda_j$, so $\alpha$ 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

[18] It is assumed that both the quality and price of the outside good $q_0$ are zero.

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

The term $\mu_j = -\beta_0 + \lambda_j P_{\Delta j}$ has been introduced to highlight the fact that $\beta_0$ 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:

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

where $F()$ is the CDF of $\epsilon$.

## 5.2 Model Intuition Graphically

Figure 4 helps provide intuition for the model's mechanics. On the X-axis is an individual's valuation for quality, $y_i = X_i\beta + \beta_0 + \epsilon_i$, e.g. affinity for movies. For presentation purposes, the X-axis has been rescaled by subtracting $X_i\beta + \beta_0$, so the scale corresponds to the value of $\epsilon_i$, the uncertainty in the individual's value for quality. The PDF of $\epsilon_i$ for this individual is shown by the curve in the figure.

If the shock $\epsilon_i$ 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 $\epsilon_i$ must exceed is given by $\theta_{i,1}$ from equation 7, shown by a vertical line in Figure 4. 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 an individual $i$ chooses a given tier $j$ equals the area of the PDF of $\epsilon_i$ 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 websites, a variable in set $X$, predicts a tendency to consume Netflix, indicating consumers with many such visits have higher values for Netflix on average. Then the corresponding component of $\beta$ in the equation for $\theta_{i,j}$ would have a positive value. Since $X\beta$ enters negatively in equation 7, its impact on $\theta$ is negative. Hence, in Figure 4, a unit increase in the value of this X shifts all three values of $\theta_{i,j}$ left by the corresponding value of $\beta$, capturing the higher probability that the consumer subscribes to Netflix.[19]

[19] The equations for $\theta_{i,j}$ are estimated subject to the normalized value of the standard deviation of $\epsilon$. Note, however, that the scaling of $\epsilon$ is irrelevant in determining outcomes. If one were to instead assume, say, a higher

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 $\epsilon$, in order to be willing to choose tier $j$ over tier $j-1$. A price increase in $j$ also causes $P_{\Delta j+1} = P_{j+1} - P_j$ to decrease, causing $\theta_{i,j+1}$ to shift left. Hence, when the price of tier $j$ increases, some consumers switch to either the higher or lower adjacent tier. Note however, 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 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 an individual consumes each tier. The expected revenues from the individual in Figure 4 equals $P_1 * Area\ A + P_2 * Area\ B + P_3 * Area\ C$, where the areas depend on prices and the values of $X_i$. Total expected revenues are then found by summing expected revenues across individuals.

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

## 5.3 Estimation

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

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  10%). It is assumed that consumers of these plans would subscribe to one of the unlimited plans, had these limited plans been unavailable.[20] Second, while I can impute whether or not a given individual subscribed to Netflix, for most subscribers it is not known directly which tier they subscribed to. The partially-conceiled tier choice requires slight modications to the likelihood function, causing it to less well resemble the

level of the standard deviation of $\epsilon$, $\sigma$, then the model and data would yield estimates of all other parameters exactly $\sigma$ times higher as well. Due to this countervailing change, $\frac{\partial s_{i,j}}{\partial X_j}$ would be left unchanged.

[20] 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.

likelihood function in standard ordered probit models.

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 a function of parameters $\mu$ and $\beta$ defined in equation 7. The notation $i(j = -1)$ denotes the set of individuals observed not subscribing to Netflix, $i(j = 0)$ denotes the set of individuals subscribing to Netflix, but whose subscription tier is unknown, and $i(j = k)$ denote the sets of individuals observed purchasing tier $k \in (1, 2, 3)$.

# 6 Counterfactual Simulations

This section simulates counterfactual environments in which Netflix implements first degree price discrimination. Specifically, optimal variable profits and the dispersion of prices offered to different individuals are calculated separately using demographics alone and then all variables to explain a consumers willingness to pay.

## 6.1 Calculating Variable Profits

For a given price schedule, the firm's variable profit from individual $i$ equals:

$$\Pi = \sum_{i} \sum_{j=1}^{3} \left(P_{i,j} - c_j\right) \left(F(\theta_{i,j+1}) - F(\theta_{i,j})\right) \quad (10)$$

where $c_j$ is the marginal cost of providing tier $j$ service. The marginal costs and their values were described in section 2. Recall that $\theta_{i,j}$ are a function of price.[21]

[21] 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. To ensure prices meet this requirement, 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:

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

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

Optimal prices with and without tailoring can be found via grid-search. Increments of 5 cents were used.[22] Unreported tests found reducing the increment size further yields similar profit estimates.

### 6.2 Assignment of Unidentified Parameter

In order to simulate scenarios with counterfactual pricing, one must specify consumers' responsiveness to price, since it not identified in data lacking price variation. Equation 7 shows that $\lambda$ determines the rate by which $\theta_{i,j}$ changes with prices, and hence the slope of demand.[23] Rearranging equation 7 to solve for $\lambda_j$, yields:

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

Note in equation 11 that $\lambda_j$, the price parameter, is monotonically increasing in the value of the parameter $\beta_0$, which is not recovered from estimation. Hence higher $\beta_0$ imply strictly higher price sensitivities. This suggests that $\beta_0$ can be determined after estimation using supply side conditions, similar to Gentzkow [2007].

Specifically, I assume Netflix has some pricing power, and estimate the value of $\beta_0$ which implies that observed prices are the prices which maximize Netflix's static profits. Formally, I search over $\beta_0$ to find the value of $\beta_0$ which minimizes the summed square of differences between observed prices for the tiers and simulated profit maximizing prices. The resulting value, 0.622, yields a set of simulated prices that are close to observed prices, $[\$10.30, \$15.00, \$17.70]$ vs. $[\$9.99, \$14.99, \$17.99]$.[24] Since the prices of the three tiers in simulations all depend on a single parameter $\beta_0$, it was not possible to find a value of $\beta_0$ matching all three prices exactly.

[22] 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.

[23] Note the omission of the price sensitivity parameter $\alpha$. $\alpha$ is not separately identified from $\lambda_j$ 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 $\alpha$ to $\lambda_j$ is the coefficient on tier $j$ price.

[24] The estimated value of $\beta_0$ depends on the estimated distribution of the latent variable and thresholds, which can differ slightly depending on the variables used in estimation of the model. However, the value of $\beta_0$ did not depend much on whether demographics or all variables were used as explanatory variables. The two estimates of $\beta_0$ were quite similar: 0.622 vs, 0.568.

### 6.3 Counterfactual Results

Variable profits, prices, sales, and other outcome variables are simulated under both status quo pricing, i.e. $2^{nd}$ degree PD, as well as under $1^{st}$ 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 variable profits from individually tailored pricing.[25] Using all variables to tailor prices, one can yield variable profits 1.39% higher than variable profits obtained using non-tailored $2^{nd}$ degree PD. Using demographics alone to tailor prices raises profits by much less, yielding variable profits only 0.14% higher than variable profits attainable under $2^{nd}$ degree PD. Since adding web browsing data substantially increases the variable profit gain from first degree price discriminating, it increases the likelihood that firms will implement 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 5 shows histograms of prices for the 1 DVD at-a-time tier. The figure includes overlaid histograms, one for person-specific prices using all variables to tailor prices, and another using only demographics. Clearly, a much wider range of prices occurs when all variables are used to individually tailor prices.

Table 5 provides further details on the impact of tailored pricing on the distribution of offered prices for all three subscription tiers. When all variables are used, the consumer estimated to have the highest value for Netflix would face prices about 40% higher when prices are tailored compared to when they are not. Similarly, the $99.9^{th}$ percentile individual would face prices about 20% higher, the $99^{th}$ percentile about 13% higher, and the 95% percentile about 5% higher. The median consumer pays slightly less when prices are tailored, and the lowest offered prices are about 30% less than untailored prices. These results together imply that the highest price offered would be roughly double the lowest price offered.

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 $2^{nd}$ 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 gain from tailored pricing is about two-thirds lower when the upper bound is imposed. As before, the variable profit gain from tailored pricing is much higher when prices are based

[25]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.

on all variables rather than only on demographics (0.44% vs. 0.05%).

A pertinent question is whether $1^{st}$ degree PD substantially improves the fraction of surplus 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.[26]

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 better and bigger datasets. Other data might, for example, include behaviors not captured in browsing history which can be used to predict consumers' valuations. Examples of other data are: location by time of day, collected on smartphones via GPS, and contextual variables derived from user-generated text on twitter, emails, and text-messages.

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 $\epsilon$ in the predicted value for quality $y_i$.[27] Lower standard deviations imply better estimates of $y_i$, but shrinking the error terms $\epsilon$ 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 $\epsilon$.[28]

The results are shown in Figures 6 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 long as changes are small. Eventually, at around a tenfold increase in precision, price percentiles seem to level off. After that, they do not increase much as precision increases further. At that point, the 99.9% percentile prices are about four times the 10% percentile prices.

[26]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.

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

[28]A wider price grid was used in these simulations. To speed computation, the increments between grid points were increased as well, to $1.

# 7 Conclusion

This paper finds that the increase in profits made feasible by $1^{st}$ degree PD is much higher when web browsing behavior, rather than just demographics, is used to predict individuals' valuations. This suggests that $1^{st}$ degree PD might evolve from merely theoretical to practical and widely employed. This will impact consumers, as I find that the estimated range of prices offered to different individuals for use of the same product is quite large.

Widespread $1^{st}$ degree PD may have large efficiency effects, albeit of ambiguous direction. 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, $1^{st}$ degree PD may unilaterally raise profits, but employed jointly it may increase competition, reducing profits and hence innovation incentives.

Consumer behavior may also be affected. Consumers may waste effort masking themselves as low valuation consumers. Or, in an unlikely extreme scenario, consumers could reduce labor effort, knowing that earning higher wages would result in being charged higher prices. In a related application, Feldstein [1995] finds that $1^{st}$ degree PD as applied to college tuition distorts savings rates.

Lastly, $1^{st}$ degree PD raises equity concerns - is it fair for consumers to pay different prices for the same product? The public seems to think not. Kahneman et al. [1986] find $1^{st}$ degree PD was viewed as unfair by 91% of respondents.

# References


Netflix annual report. 2006.

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

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

Ronald H Coase. Durability and monopoly. *Journal of Law and Economics*, 15(1):143–49, 1972.

Gregory S Crawford and Matthew Shum. Monopoly quality degradation and regulation in cable television. *Journal of Law and Economics*, 50(1):181–219, 2007.

Martin Feldstein. College scholarship rules and private saving. *American Economic Review*, 85 (3):552–66, 1995.

Drew Fudenberg and J Miguel Villas-Boas. Behavior-based price discrimination and customer recognition. 2005.

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

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

Kathryn Graddy and George Hall. A dynamic model of price discrimination and inventory management at the fulton fish market. *Journal of Economic Behavior & Organization*, 80(1): 6–19, 2011.

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*, pages 728–741, 1986.

Ashley Langer. Demographic preferences and price discrimination in new vehicle sales. *manuscript, University of Michigan*, 2011.

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. *http://cyberlaw.stanford.edu/blog/2011/07/tracking-trackers-early-results*, 2011.

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

Jakub Mikians, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. Detecting price and search discrimination on the internet. 2012.

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.

Harikesh Nair. Intertemporal price discrimination with forward-looking consumers: Application to the us market for console video-games. *Quantitative Marketing and Economics*, 5(3):239–292, 2007.

AC Pigou. The economics of welfare. 1920.

Adrian Raftery, David Madigan, and Jennifer Hoeting. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191, 1997.

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.

Ben 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.

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

# A Heterogeneous Price Sensitivity Model

When 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 shown below.[29]

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

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 $\alpha_i$ are person-specific parameters that reflect individual $i$'s valuation for quality and price sensitivity. This utility specification is similar to Mussa and Rosen's (1978), but allows for differences across consumers in price sensitivity $\alpha_i$.

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

[29] The conditional indirect utility refers to the indirect utility conditional on choosing a specific option.

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

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 3 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 $\lambda_j$, respectively.

It is assumed that the quality of the outside good $q_0$ equals zero. With this assumption, equation 4 for $j = 1$ includes the individual rationality constraint, i.e. conditions necessary for the individual to prefer some tier of service, as opposed to the outside good.

Next, the variables $y_i$ and $\alpha_i$ 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 $\gamma$ and $\beta$ reflect differences across consumers explainable with the data. Note that only a single error term has been introduced. 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, $\sigma$, 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, $\gamma_0$ cannot be separately identified from the scaling of quality levels, $\lambda_j$, so $\gamma_0$ 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

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

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

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

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

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

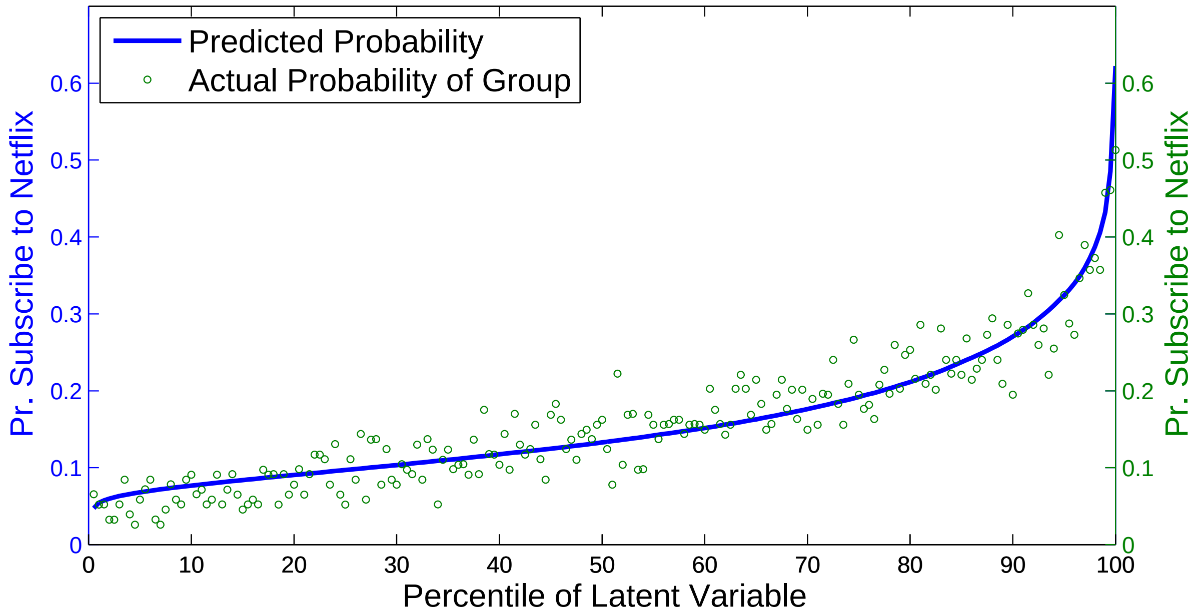


Figure 1: Model Fit - Predicted Probilities in Holdout Sample When All Variables Used

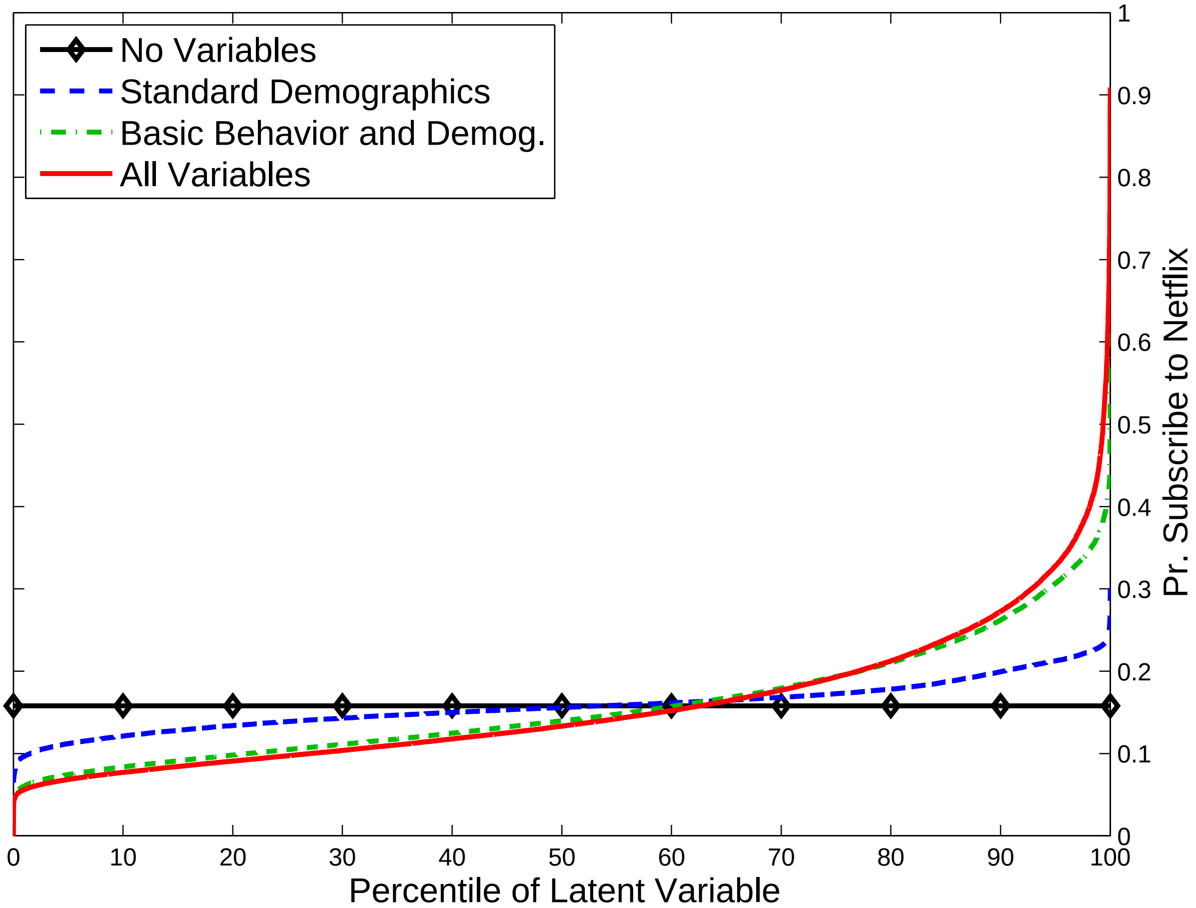


Figure 2: Range of Predicted Probabilities, Using Various Sets of Explanatory Variables

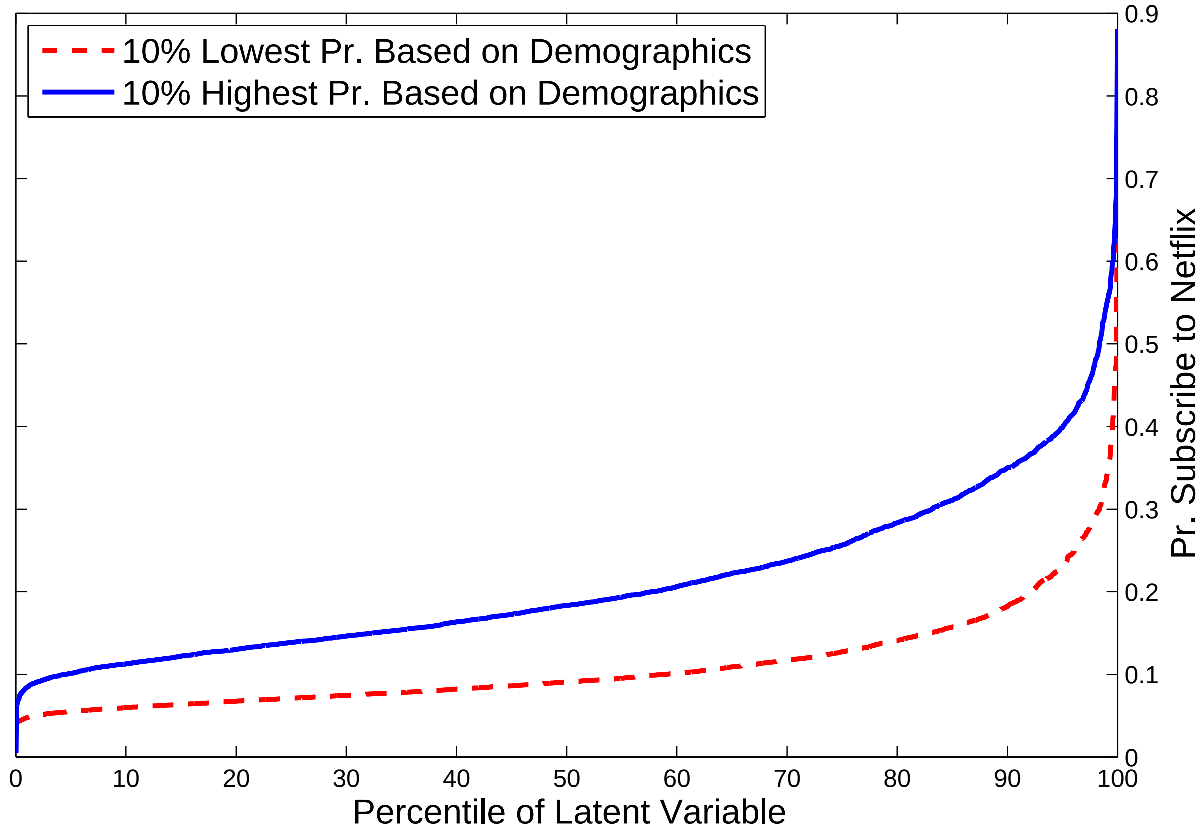


Figure 3: Range of Predicted Probabilities For Subsets of Individuals

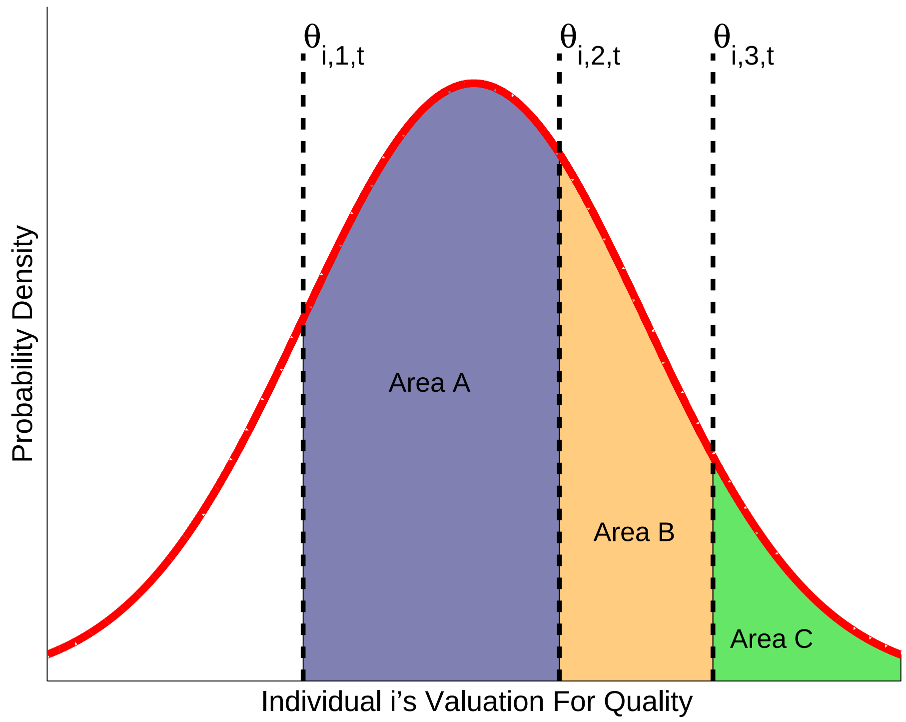


Figure 4: Graphical Depiction of Model

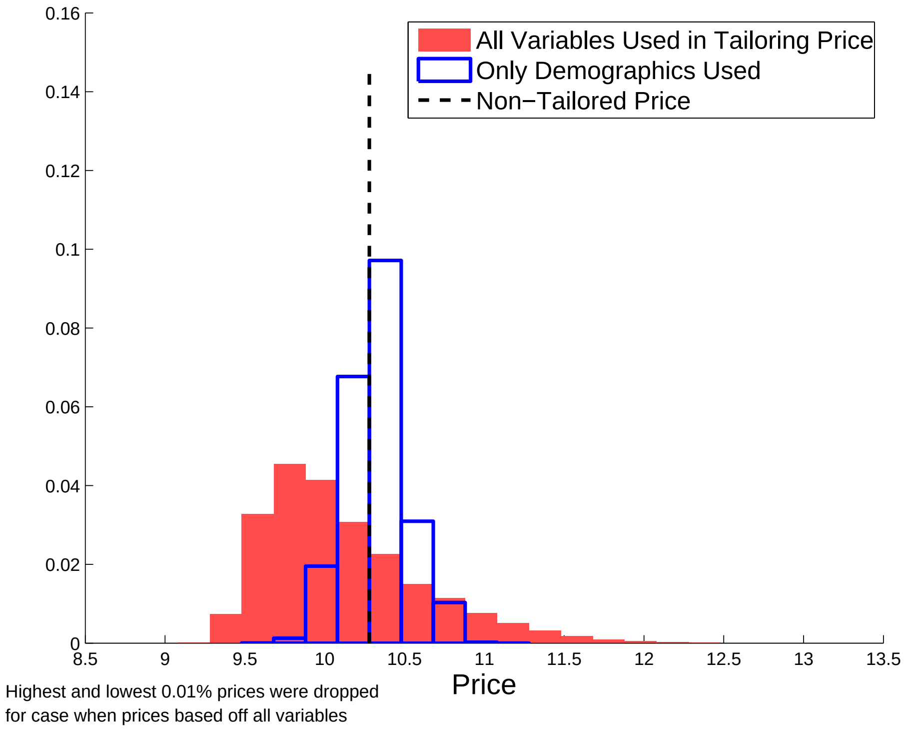


Figure 5: Histogram of Individually Tailored Prices - 1 DVD At-A-Time Plan

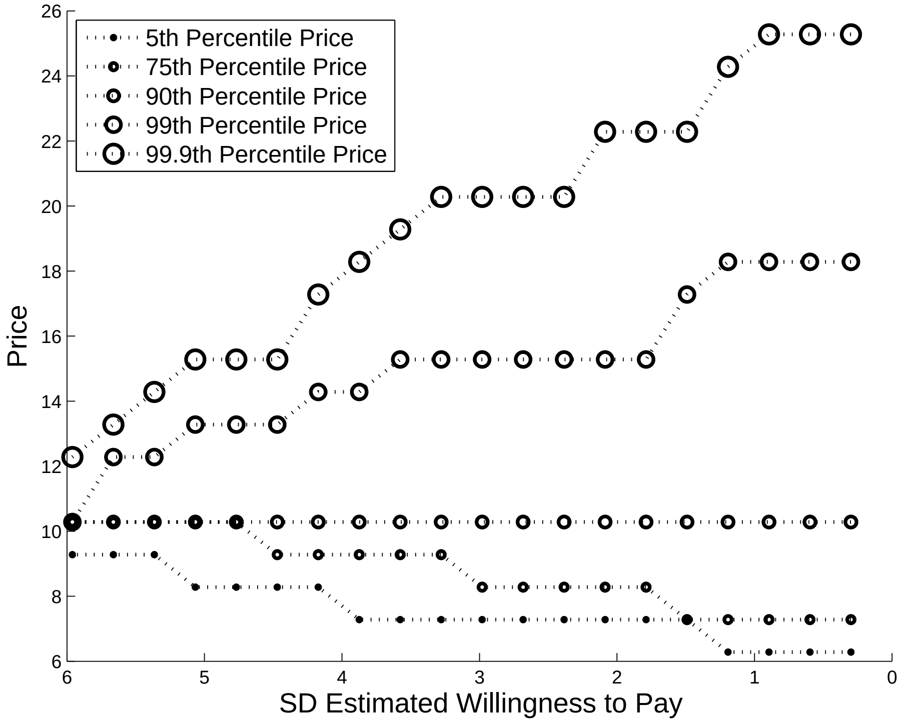


Figure 6: Graphical Depiction of Dispersion of Prices - 1 DVD At-A-Time Plan

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

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

Table 3: Simulated Changes in Various Outcomes Resulting From $1^{st}$ Degree PD

| | Percent Change When Price Based on: | |
|---|---|---|
| | Demographics | All Variables |
| Variable Profits | 0.14% | 1.39% |
| Sales (DVDs At-A-Time) | 0.85% | 1.84% |
| Subscribers | 0.17% | 1.03% |
| Aggregate Consumer Surplus | 8.13% | 1.54% |

Table 4: Simulated Changes in Various Outcomes Resulting From Tailored Discounts Off Optimized $2^{nd}$ Degree PD Prices

| | Percent Change When Discounts Based On: | |
|---|---|---|
| | Demographics | All Variables |
| Variable Profits | 0.05% | 0.44% |
| Sales (DVDs At-A-Time) | 2.56% | 6.82% |
| Subscribers | 2.41% | 6.64% |
| Aggregate Consumer Surplus | 12.95% | 4.16% |

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

| Price | 1 DVD At-A-Time | | 2 DVDs At-A-Time | | 3 DVDs At-A-Time | |
|---|---|---|---|---|---|---|
| Percentile | Demog. | All Variables | Demog. | All Variables | Demog. | All Variables |
| 0 | −6.8% | −29.7% | −6.3% | −28.3% | −6.2% | −27.4% |
| 0.1 | −5.8% | −9.7% | −5.7% | −9.3% | −5.6% | −9.0% |
| 1 | −4.4% | −8.8% | −4.0% | −8.3% | −3.9% | −8.2% |
| 10 | −2.4% | −6.8% | −2.3% | −6.7% | −2.3% | −6.5% |
| 25 | −1.5% | −5.3% | −1.3% | −5.3% | −1.4% | −5.4% |
| 50 | −0.5% | −2.9% | −0.3% | −3.0% | −0.6% | −3.1% |
| 75 | 0.5% | 0.5% | 0.7% | 0.3% | 0.6% | 0.3% |
| 90 | 2.4% | 4.9% | 2.3% | 4.7% | 2.0% | 4.2% |
| 99 | 3.9% | 13.6% | 3.6% | 12.7% | 3.4% | 12.2% |
| 99 | 5.3% | 21.9% | 5.0% | 20.7% | 4.8% | 19.8% |
| 100 | 7.7% | 41.8% | 7.3% | 39.7% | 7.0% | 38.1% |

# First Degree Price Discrimination Using Big Data

Benjamin Reed Shiller
Department of Economics, Brandeis University

September 5, 2013

**Abstract**

Second and $3^{rd}$ degree price discrimination (PD) receive far more attention than $1^{st}$ degree PD, i.e. person-specific pricing, because the latter requires previously unobtainable information on individuals' willingness to pay. I show modern web behavior data reasonably predict Netflix subscription, far outperforming data available in the past. 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 methods for handling massive datasets. Simulations show using demographics alone to tailor prices raises profits by 0.14%. Including web browsing data increases profits by much more, 1.4%, increasingly the appeal of tailored pricing, and resulting in some consumers paying twice as much as others do for the exact same product.

# Contents

# 1 Introduction

First degree price discrimination (PD), dating back to at least Pigou [1920] in the literature, theoretically allows the firm to extract full surplus. Yet, the empirical literature instead focuses on other forms of price discrimination, which have been found to allow far less surplus extraction, about a third in studied contexts.[1,2] These other papers implicitly assume that first degree PD is infeasible - firms do not have information on willingness to pay at the individual level.[3] Moreover, orthodox instruction uses this argument to motivate $2^{nd}$ degree and $3^{rd}$ degree PD. While sound historically, this argument may no longer hold. Large datasets on individual behavior, popularly referred to as "big data," are now readily available, and contain information potentially useful for person-specific pricing.[4] For example, web browsing data may indicate psychographic profiles or direct interest in a related product, or reflect latent demographics such as sexual orientation, social phobia, and marital happiness - all information that can be used to form a hedonic estimate of willingness to pay.[5] In this paper, I investigate the extent of incremental information contained in web-browsing data, the profitability of first degree PD, and the resulting distribution of prices different consumers are offered when purchasing the exact same item.

Netflix provides an auspicious context. First, since purchases occur online, Netflix could offer tailored prices, as a couple of other online sellers have tried [Mikians et al. [2012]]. Second, Netflix could effectively price discriminate - its products were differentiated from competitors' products implying pricing power, and arbitrage appears costly enough given the fact that Netflix has long employed another form of price discrimination, $2^{nd}$ degree PD. Last, focusing on Netflix overcomes estimation problems faced by researchers, since Netflix subscription can be imputed in a dataset that also includes browsing behavior variables to tailor pricing.

This paper overcomes obstacles in incorporating large numbers of explanatory variables. Potential problems include insufficient degrees of freedom, overfitting, tractability and convergence issues, and computer memory limitations. Missing data can also be problematic - with many variables there may be few observations with non-missing values for all variables. Model averaging overcomes or mitigates these problems, but without further steps yields biased results in binary and ordered choice models. To address this problem, I show an Ordered choice Model Averaging (OMA) method, a very simple solution to this problem which can be used in standard

[1] Recent empirical papers focus on $3^{rd}$ degree PD [Graddy [1995], Graddy and Hall [2011], Langer [2011]], $2^{nd}$ degree PD [Crawford and Shum [2007], McManus [2008]], intertemporal pricing [Nair [2007]] and bundling [Chu et al. [2011], Shiller and Waldfogel [2011]]

[2] Shiller and Waldfogel [2011] find that bundling, nonlinear pricing, and $3^{rd}$ degree PD cannot extract as profits more than about third of surplus in the market for digital music.

[3] It is worth noting that tailoring prices to consumers is not currently per se illegal, as is evident from widespread use of $3^{rd}$ degree PD. Tailoring prices to downstream firms is however prohibited under the Robinson-Patman Act.

[4] Madrigal [2012], Mayer [2011] note that web-browsing behavior is collected by hundreds of firms.

[5] Psychographic profile categorizes individuals based on attitudes, activities, values, and behavior.

statical programs.[6]

This method is used to determine how useful different sets of variables are in estimating the probability 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 91%.

Next, an empirical model is used to translate the increased precision from web-browsing data into key outcome variables. Specifically, a model derived from canonical quality discrimination theory models is used to estimate demand for Netflix in the observed environment, in which Netflix employed $2^{nd}$ degree PD, but not $1^{st}$ degree PD. The model is then used to simulate pricing and profits in the hypothetical counterfactual occurring if Netflix had implemented $1^{st}$ degree PD.

I find that web browsing behavior substantially raises the amount by which person-specific pricing raises variable profits relative to $2^{nd}$ degree PD - 1.39% if using all data to tailor prices, but only 0.14% using demograhpics alone. Web-browsing data make $1^{st}$ degree PD more appealing to firms and likely to be implemented, thus impacting consumers. Substantial equity concerns may arise - I find some consumers may be charged twice as much as others are for the same product.

The closest literature, a strand 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.[7] However, they assumed that consumers were myopic. Anecdotal evidence following Amazon's pricing experiment in the early 2000s suggests otherwise [Streitfeld [2000]]. Acquisti and Varian [2005], Fudenberg and Villas-Boas [2005] show theoretically that $1^{st}$ degree PD 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."

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 higher prices to individuals that use the internet during the day on Tuesdays and Thursdays, and visit Wikipedia.org, patterns consumers may not recognize. Moreover, with many variables, there may not be any easy heuristics consumers

[6]The method works on both binary and ordered choice models

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

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 data, providing the incentive for consumers to reveal them.

The remainder of the paper is organized as follows. Section 2 describes the context and industry background. Next, Section 3 describes the data. Section 4 then shows how well various sets of data explain propensity to purchase. Lastly, Sections 5 and 6 present a model and estimate optimal person-specific 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 [net [2006]]. This implies that about 16.7% of internet connected households consumed Netflix during 2006.[8]

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.[9] Moreover, Netflix's customer acquisition algorithm was well known, further differentiating their services.

Netflix's subscriptions plans can be broken into two categories. Unlimited plans allow consumers to receive an unlimited number of DVDs by mail each month, but restrict the number of DVDs in a consumer's possession at one time. Limited plans set both a maximum number of DVDs the consumer can possess at one time, and the maximum number sent in one pay 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.[10] None of the plans allowed video streaming, since Netflix did not launch that service until 2007 [net [2006]].

[8] 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.

[9] 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 [net [2006]].

[10] 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.

Key statistics for later analyses are Netflix's marginal cost 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.[11]

# 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 pages viewed on that website, the referring website, and details on any transactions. 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 inputed 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 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 points of the estimated share of U.S. internet connected housedholds subscribing, presented in Section 2. 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 other dubious categories, leaving $4,789$ popular websites to calculate additional variables.[12] The total number of visits to all websites and to each single website were computed

[11] 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 net [2006] for further details.

[12] yoyo.org provides a user-supplied list of some websites of dubious nature. Merging this list with the ComScore

for each computer user. Similar variables were computed for transactions.

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, percent of online browsing by time of day and by day of week, and broadband indicator. Detailed web-behavior contain variables indicating number visits to a particular website.

## 4 Prediction in Status Quo

This section predicts 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 across these sets of explanatory variables, to inform on the the relative benefit of including web browsing behavior.

First, a Probit model is used to investigate which standard demographic variables are significant predictors of a 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.

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 were no longer significant once basic web behavior variables were added, suggesting they were 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, number visits to each particular website are added one at-a-

data reveal that such websites tend to have very high ($\geq 0.9$) or very low ($\leq 0.1$) rates of visits that were referred, 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.

[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.

time to the significant demographic and basic web behavior explanatory variables. Overall, 29% of websites were significant at the 5% level, and 18% at the 1% level, far more than expected by chance alone. This suggests that significance is not predominantly due to type I errors. It is implicitly assumed that consumers' observed web browsing behavior is driven by the same innate characteristics as their Netflix choices.[14] This is supported intuitively by the the types of websites found to be significant. The twenty websites which best explain Netflix subscription are shown in Table 2. All twenty were positive predictors. Inspection reveals they are comprised of websites which are likely used by movie lovers (IMDB, Rotten Tomatoes), internet savvy users (Wikipedia), those preferring mail ordering (Amazon, Gamefly), and discount shoppers (Bizrate, Price Grabber).

I next investigate the joint prediction of all website variables combined, rather than just considering one at-a-time. Model averaging, averaging many smaller models together to yield a final estimate, is used to overcome several problems common when using data with many explanatory variables.

Problems common when including many explanatory variables, which are addressed by model averaging, are as follows. First, there may not be enough degrees of freedom, preventing estimation. This is especially problematic when interaction effects or higher order terms are included. A second problem is overfitting, leading to biased estimates. Even with many observations, overfitting can occur if errors are not independent. Third, large models may be prone to convergence problems or exceeding computer memory limits. Missing data may also be problematic. Most observations may have a missing value for at least one variable, leaving few observations with all nonmissing values to estimate the model.

Model averaging proceeds as follows. First, the set of explanatory variables $X$ is divided in two. Label these sets $Z_1$ and $Z_2$. $Z_1$ is the set of variables that are deemed by the econometrician to have a high likelihood of importance, in this case, demographics and basic web behavior variables. The variables in $Z_1$ that are significant when $Z_2$ are excluded will be included in all models. Call this set $Z_1'$. The set $Z_2$ includes variables indicating number visits to each website separately, referred to as the detailed web behavior variables. 50,000 subsets $s$ of five variables in $Z_2$ are drawn. The model is re-estimated adding each subset $s$ to $Z_1'$, dropping any variables in $s$ which are not significant.[15]

[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.

[15]To increase computational efficiency, rather than including all demographic and basic web behavior variables, I rather summarize their contribution by including a single variable $\hat{y}_i$ equal to the value of $X\beta$ when only they are included. Replacing the individual variables with this single variable speeds computation, but restricts the model's ability to estimate marginal effects of a single variable in $Z_1$ conditional on the value of a variable in $Z_2$.

Each set of variables $(Z_1', s)$ yields its own estimates of the expected value of the latent variable for each individual and the threshold in the Probit model, as well as a value of the maximized likelihood. Taking a weighted average of these values across models enables predictions based on the information contained in all models together.

Without further steps, model averaging yields biased results in binary choice models. It over-predicts the probability of subscription for those least likely to subscribe, and under-predicts the probability for those most likely to subscribe. The reason is as follows. First, recall that in this binary choice model a consumer buys only if the true value of an underlying latent variable exceeds some threshold. The probability that a consumer subscribes thus depends on how far the estimate of the latent variable is from the threshold, and the accuracy of the estimate of the latent variable, i.e. its error's standard deviation. More extreme probabilities are reached only if the scaling of the latent variable is increased or the standard deviation of the error term is reduced. But neither can occur. The standard deviation of the error term is typically normalized to a value of one, since it is not separately identified in such models. Mechanically, the latent variable's scaling does not increase either - averaging many values does not change the expected value. If anything, averaging reduces the range of latent variables as extreme values yielded in one model by chance are mitigated by averaging.

A simple analogy helps explain the problem that is occurring. Suppose several independent medical tests for a disease all come back positive. If each test alone implies the probability an individual has the disease is 0.85, then multiple independent positive tests should together imply a probability over 0.85. In an Probit model framework, this 0.85 probability can be represented by a value of the latent variable equal to approximately 1, a threshold of 0, and standard deviation of the latent variable equal to 1. Averaging over multiple tests will yield the exact same values of the threshold and latent variable, and hence the same predicted probability as any single test, 0.85.

This bias can easily be corrected with one additional step after averaging the ensemble of models, a method that can be labeled as Ordered choice Model Averaging (OMA). As a final step, another ordered choice model is run with a single explanatory variable which equals the weighted average of the expected value of the latent variable for the individual across models. If the parameter $\kappa$ on this variable is greater than 1, then the scaling of the latent variable estimates increase. This broadens the range of estimated latent variables across individuals, relative to the standard deviation of the error, allowing for more extreme probabilities, and incorporating the information from many models.

The averaged values of the latent variable and threshold in the Probit model, used in the OMA method, depend on the weights placed on each model when computing the average. These weights are estimated as follows. Following the intuition from Occam's razor and earlier literature [Raftery et al. [1997]], the least likely set of models were excluded. Specifically, the top 2% most likely models were kept, and then averaged in relation to a function of the increase in the log likelihood obtained by their inclusion in the holdout sample.[16,17] Specifically, I parameterized the weights used in averaging as $\frac{e^{\omega*(LL_{z_1',s}-LL_{z_1'})}}{\sum e^{\omega*(LL_{z_1',s}-LL_{z_1'})}}$, where $LL_{z_1',s}$ and $LL_{z_1'}$ denote the log likelihood of model when the set $(z_1', s)$ and $(z_1')$ are included as possible explanatory variables, and $\omega$ is a parameter that determines the relative weights of the most likely to less likely models. The value of $\omega$ was chosen maximize the probability of the data in the holdout sample. Its value, 0.017, implies that most of the weight is on a small number of models. Nearly half the weight falls on one model, about 70% on the top two, 95% on the top 15, and 99% on the top 50.

Figure 1 shows the predictions from model averaging in the holdout sample. Specifically, individuals in the holdout sample are ordered according to their estimated value of the latent variable in the ordered choice model, then grouped. The average predicted probability and observed probabilities are then calculated for each group. Notice that these predicted probabilities, shown in solid blue line, do in fact seem to follow the actual probabilities of subscription.

The main takeaway from this section is summarized in Figure 2. It plots the predicted probability each individual subscribes based on various sets of explanatory variables together on one graph. Note the Y-axis range is larger than in Figure 1, which averaged predicted probability within groups, obscuring extreme probabilities. Including web behavior variables does in fact seem to substantially help prediction. Predicted probabilities of subscription ranged from $5.9 * 10^{-11}$ percent to 91% 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% chance.

Figure 3 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. Demographics predict this probability ranges between 6.5% and 12%. Predictions for this same group based on the full set of variables yield probabilities as high as 75%. 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. Hence, demographics used alone grossly misclassifies some individuals as low or high probability subscribers.

[16] For linear regression models, the expected likelihood of each model is typically found by integrating over the values of each parameter in the model using an analytic expression. For choice models, no such analytic expression exists. To speed estimation, I used the maximized likelihood instead.

[17] This implicitly assumes an uninformative prior - ie. ex-ante all models are assumed equally likely. Pre-existing information on model likelihood can easily be incorporated by changing this assumption.

# 5 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 each tier, but may differ in how much they value the quality of higher tiers. Firms in the models set prices of the tiers of service, but not qualities.[18]

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 concerns over the endogeneity of the coefficient on price, 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.

## 5.1 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 $\alpha$ 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:

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

[18] In the canonical $2^{nd}$ 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. consumer cannot rent half a DVD.

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, only the incentive compatibility constraints for neighboring products bind, and consumer $i$ chooses product $j$ if and only if the following inequality condition is satisfied.[19]

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

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 $\lambda_j$, 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, $\sigma$, 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, $\alpha$ cannot be separately identified from the scaling of quality levels, $\lambda_j$, so $\alpha$ 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 $\beta_0$ 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:

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

where $F()$ is the CDF of $\epsilon$.

[19] It is assumed that both the quality and price of the outside good $q_0$ are zero.

### 5.2 Model Intuition Graphically

Figure 4 helps provide intuition for the model's mechanics. On the X-axis is an individual's valuation for quality, $y_i = X_i\beta + \beta_0 + \epsilon_i$, e.g. affinity for movies. For presentation purposes, the X-axis has been rescaled by subtracting $X_i\beta + \beta_0$, so the scale corresponds to the value of $\epsilon_i$, the uncertainty in the individual's value for quality. The PDF of $\epsilon_i$ for this individual is shown by the curve in the figure.

If the shock $\epsilon_i$ 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 $\epsilon_i$ must exceed is given by $\theta_{i,1}$ from equation 7, shown by a vertical line in Figure 4. 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 an individual $i$ chooses a given tier $j$ equals the area of the PDF of $\epsilon_i$ 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 websites, a variable in set $X$, predicts a tendency to consume Netflix, indicating consumers with many such visits have higher values for Netflix on average. Then the corresponding component of $\beta$ in the equation for $\theta_{i,j}$ would have a positive value. Since $X\beta$ enters negatively in equation 7, its impact on $\theta$ is negative. Hence, in Figure 4, a unit increase in the value of this X shifts all three values of $\theta_{i,j}$ left by the corresponding value of $\beta$, capturing the higher probability that the consumer subscribes to Netflix.[20]

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 $\epsilon$, 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, 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 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 an individual consumes each tier. The expected revenues from the individual in Figure 4 equals $P_1 * Area\ A + P_2 * Area$

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

$B + P_3 * Area\ C$, where the areas depend on prices and the values of $X_i$. Total expected revenues are then found by summing expected revenues across individuals.

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

### 5.3 Estimation

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

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 10%). It is assumed that consumers of these plans would subscribe to one of the unlimited plans, had these limited plans been unavailable.[21] Second, while I can impute whether or not a given individual subscribed to Netflix, for most subscribers it is not known directly which tier they 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.[22]

The log likelihood function equals:

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

[21] 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.

[22] The ordered-choice thresholds for the $2^{nd}$ and $3^{rd}$ tiers ($\mu_2$ and $\mu_3$) 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 $\mu_1$, 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.

where the data $D$ contain subscription choice and explanatory variables, and $\theta_{i,j}$ is a function of parameters $\mu$ and $\beta$ defined in equation 7. The notation $i(j = -1)$ denotes the set of individuals observed not subscribing to Netflix, $i(j = 0)$ denotes the set of individuals subscribing to Netflix, but whose subscription tier is unknown, and $i(j = k)$ denote the sets of individuals observed purchasing tier $k \in (1, 2, 3)$.

# 6 Counterfactual Simulations

This section simulates counterfactual environments in which Netflix implements first degree price discrimination. Specifically, optimal variable profits and the dispersion of prices offered to different individuals are calculated separately using demographics alone and then all variables to explain a consumers willingness to pay.

## 6.1 Calculating Variable Profits

For a given price schedule, the firm's variable profit from individual $i$ equals:

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

where $c_j$ is the marginal cost of providing tier $j$ service. The marginal costs and their values were described in section 2. Recall that $\theta_{i,j}$ are a function of price.[23]

Optimal prices with and without tailoring can be found via grid-search. Increments of 5 cents were used.[24] Unreported tests found reducing the increment size further yields similar profit estimates.

[23]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. To ensure prices meet this requirement, 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:

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

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

[24]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.

### 6.2 Assignment of Unidentified Parameter

In order to simulate scenarios with counterfactual pricing, one must specify consumers' responsiveness to price, since it not identified in data lacking price variation. Equation 7 shows that $\lambda$ determines the rate by which $\theta_{i,j}$ changes with prices, and hence the slope of demand.[25] Rearranging equation 7 to solve for $\lambda_j$, yields:

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

Note in equation 11 that $\lambda_j$, the price parameter, is monotonically increasing in the value of the parameter $\beta_0$, which is not recovered from estimation. Hence higher $\beta_0$ imply strictly higher price sensitivities. This suggests that $\beta_0$ can be determined after estimation using supply side conditions, similar to Gentzkow [2007].

Specifically, I assume Netflix has some pricing power, and estimate the value of $\beta_0$ which implies that observed prices are the prices which maximize Netflix's static profits. Formally, I search over $\beta_0$ to find the value of $\beta_0$ which minimizes the summed square of differences between observed prices for the tiers and simulated profit maximizing prices.[26] The resulting value, 0.622, yields a set of simulated prices that are close to observed prices, $[\$10.30, \$15.00, \$17.70]$ vs. $[\$9.99, \$14.99, \$17.99]$.[27] Since the prices of the three tiers in simulations all depend on a single parameter $\beta_0$, it was not possible to find a value of $\beta_0$ matching all three prices exactly.

### 6.3 Counterfactual Results

Variable profits, prices, sales, and other outcome variables are simulated under both status quo pricing, i.e. $2^{nd}$ degree PD, as well as under $1^{st}$ 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 variable profits from individually tailored pricing.[28]

[25] Note the omission of the price sensitivity parameter $\alpha$. $\alpha$ is not separately identified from $\lambda_j$ 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 $\alpha$ to $\lambda_j$ is the coefficient on tier $j$ price.

[26] 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 $\beta_0$ yielding optimal simulated prices more than twice those observed still yielded very similar results when expressed in percent changes.

[27] The estimated value of $\beta_0$ depends on the estimated distribution of the latent variable and thresholds, which can differ slightly depending on the variables used in estimation of the model. However, the value of $\beta_0$ did not depend much on whether demographics or all variables were used as explanatory variables. The two estimates of $\beta_0$ were quite similar: 0.622 vs, 0.568.

[28] Percentages rather than absolute profits were reported because simulated variable profits in the status quo

Using all variables to tailor prices, one can yield variable profits 1.39% higher than variable profits obtained using non-tailored $2^{nd}$ degree PD. Using demographics alone to tailor prices raises profits by much less, yielding variable profits only 0.14% higher than variable profits attainable under $2^{nd}$ degree PD. Since adding web browsing data substantially increases the variable profit gain from first degree price discriminating, it increases the likelihood that firms will implement 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 5 shows histograms of prices for the 1 DVD at-a-time tier. The figure includes overlaid histograms, one for person-specific prices using all variables to tailor prices, and another using only demographics. Clearly, a much wider range of prices occurs when all variables are used to individually tailor prices.

Table 5 provides further details on the impact of tailored pricing on the distribution of offered prices for all three subscription tiers. When all variables are used, the consumer estimated to have the highest value for Netflix would face prices about 40% higher when prices are tailored compared to when they are not. Similarly, the $99.9^{th}$ percentile individual would face prices about 20% higher, the $99^{th}$ percentile about 13% higher, and the 95% percentile about 5% higher. The median consumer pays slightly less when prices are tailored, and the lowest offered prices are about 30% less than untailored prices. These results together imply that the highest price offered would be roughly double the lowest price offered.

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 $2^{nd}$ 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 gain from tailored pricing is about two-thirds lower when the upper bound is imposed. As before, the variable profit gain from tailored pricing is much higher when prices are based on all variables rather than only on demographics (0.44% vs. 0.05%).

A pertinent question is whether $1^{st}$ degree PD substantially improves the fraction of surplus 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.[29]

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.

[29] 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

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 better and bigger datasets. Other data might, for example, include behaviors not captured in browsing history which can be used to predict consumers' valuations. Examples of other data are: location by time of day, collected on smartphones via GPS, and contextual variables derived from user-generated text on twitter, emails, and text-messages.

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 $\epsilon$ in the predicted value for quality $y_i$.[30] Lower standard deviations imply better estimates of $y_i$, but shrinking the error terms $\epsilon$ 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 $\epsilon$.[31]

The results are shown in Figures 6 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 long as changes are small. Eventually, at around a tenfold increase in precision, price percentiles seem to level off. After that, they do not increase much as precision increases further. At that point, the 99.9% percentile prices are about four times the 10% percentile prices.

# 7 Conclusion

This paper finds that the increase in profits made feasible by $1^{st}$ degree PD is much higher when web browsing behavior, rather than just demographics, is used to predict individuals' valuations. This suggests that $1^{st}$ degree PD might evolve from merely theoretical to practical and widely employed. This will impact consumers, as I find that the estimated range of prices offered to different individuals for use of the same product is quite large.

Widespread $1^{st}$ degree PD may have large efficiency effects, albeit of ambiguous direction. Most textbooks espouse its efficiency based on partial equilibrium analysis. However, when

the profits across tiers $j$ for that individual.

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

[31]A wider price grid was used in these simulations. To speed computation, the increments between grid points were increased as well, to \$1.

employed by multiple firms, this result may not hold. In oligopolistic [Spulber [1979]] and differentiated product [Thisse and Vives [1988]] markets, $1^{st}$ degree PD may unilaterally raise profits, but employed jointly it may increase competition, reducing profits and hence innovation incentives.

Consumer behavior may also be affected. Consumers may waste effort masking themselves as low valuation consumers. Or, in an unlikely extreme scenario, consumers could reduce labor effort, knowing that earning higher wages would result in being charged higher prices. In a related application, Feldstein [1995] finds that $1^{st}$ degree PD as applied to college tuition distorts savings rates.

Lastly, $1^{st}$ degree PD raises equity concerns - is it fair for consumers to pay different prices for the same product? The public seems to think not. Kahneman et al. [1986] find $1^{st}$ degree PD was viewed as unfair by 91% of respondents.

## References


Netflix annual report. 2006.

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

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

Ronald H Coase. Durability and monopoly. *Journal of Law and Economics*, 15(1):143–49, 1972.

Gregory S Crawford and Matthew Shum. Monopoly quality degradation and regulation in cable television. *Journal of Law and Economics*, 50(1):181–219, 2007.

Martin Feldstein. College scholarship rules and private saving. *American Economic Review*, 85 (3):552–66, 1995.

Drew Fudenberg and J Miguel Villas-Boas. Behavior-based price discrimination and customer recognition. 2005.

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

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

Kathryn Graddy and George Hall. A dynamic model of price discrimination and inventory management at the fulton fish market. *Journal of Economic Behavior & Organization*, 80(1): 6–19, 2011.

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*, pages 728–741, 1986.

Ashley Langer. Demographic preferences and price discrimination in new vehicle sales. *manuscript, University of Michigan*, 2011.

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. *http://cyberlaw.stanford.edu/blog/2011/07/tracking-trackers-early-results*, 2011.

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

Jakub Mikians, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. Detecting price and search discrimination on the internet. 2012.

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.

Harikesh Nair. Intertemporal price discrimination with forward-looking consumers: Application to the us market for console video-games. *Quantitative Marketing and Economics*, 5(3):239–292, 2007.

AC Pigou. The economics of welfare. 1920.

Adrian Raftery, David Madigan, and Jennifer Hoeting. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191, 1997.

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.

Ben 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.

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

# A Heterogeneous Price Sensitivity Model

When 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 shown below.[32]

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

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 $\alpha_i$ are person-specific parameters that reflect individual $i$'s valuation for quality and price sensitivity. This utility specification is similar to Mussa and Rosen's (1978), but allows for differences across consumers in price sensitivity $\alpha_i$.

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 3 to yield a range of $y_i$ required for individual $i$ to buy

[32] The conditional indirect utility refers to the indirect utility conditional on choosing a specific option.

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 $\lambda_j$, respectively.

It is assumed that the quality of the outside good $q_0$ equals zero. With this assumption, equation 4 for $j = 1$ includes the individual rationality constraint, i.e. conditions necessary for the individual to prefer some tier of service, as opposed to the outside good.

Next, the variables $y_i$ and $\alpha_i$ 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 $\gamma$ and $\beta$ reflect differences across consumers explainable with the data. Note that only a single error term has been introduced. 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, $\sigma$, 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, $\gamma_0$ cannot be separately identified from the scaling of quality levels, $\lambda_j$, so $\gamma_0$ 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

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

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

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

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

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

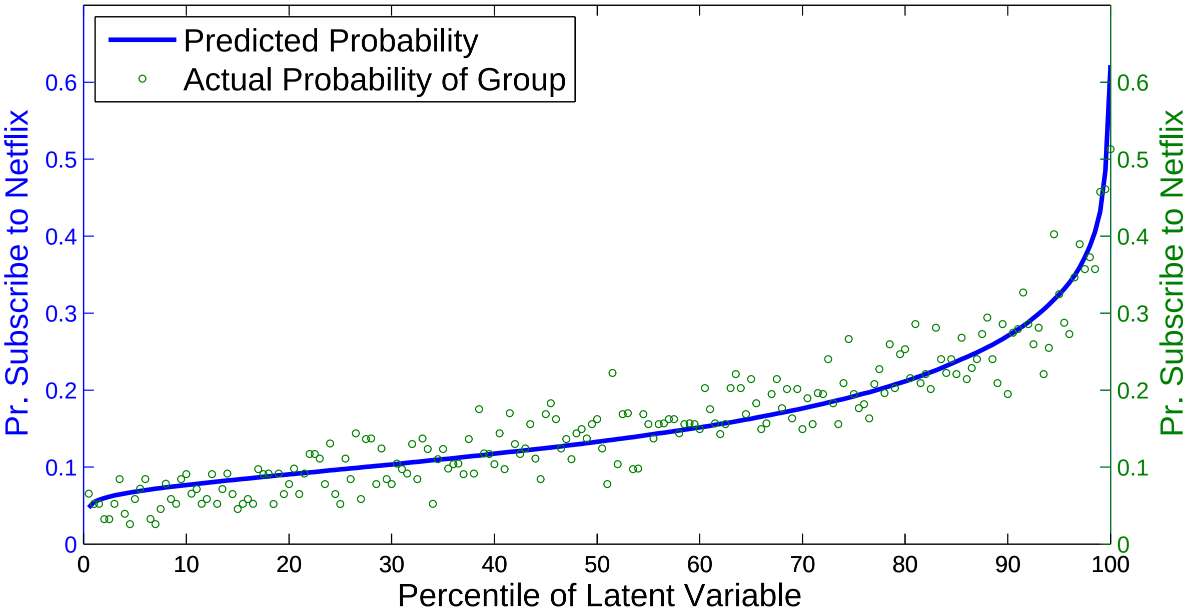


Figure 1: Model Fit - Predicted Probilities in Holdout Sample When All Variables Used

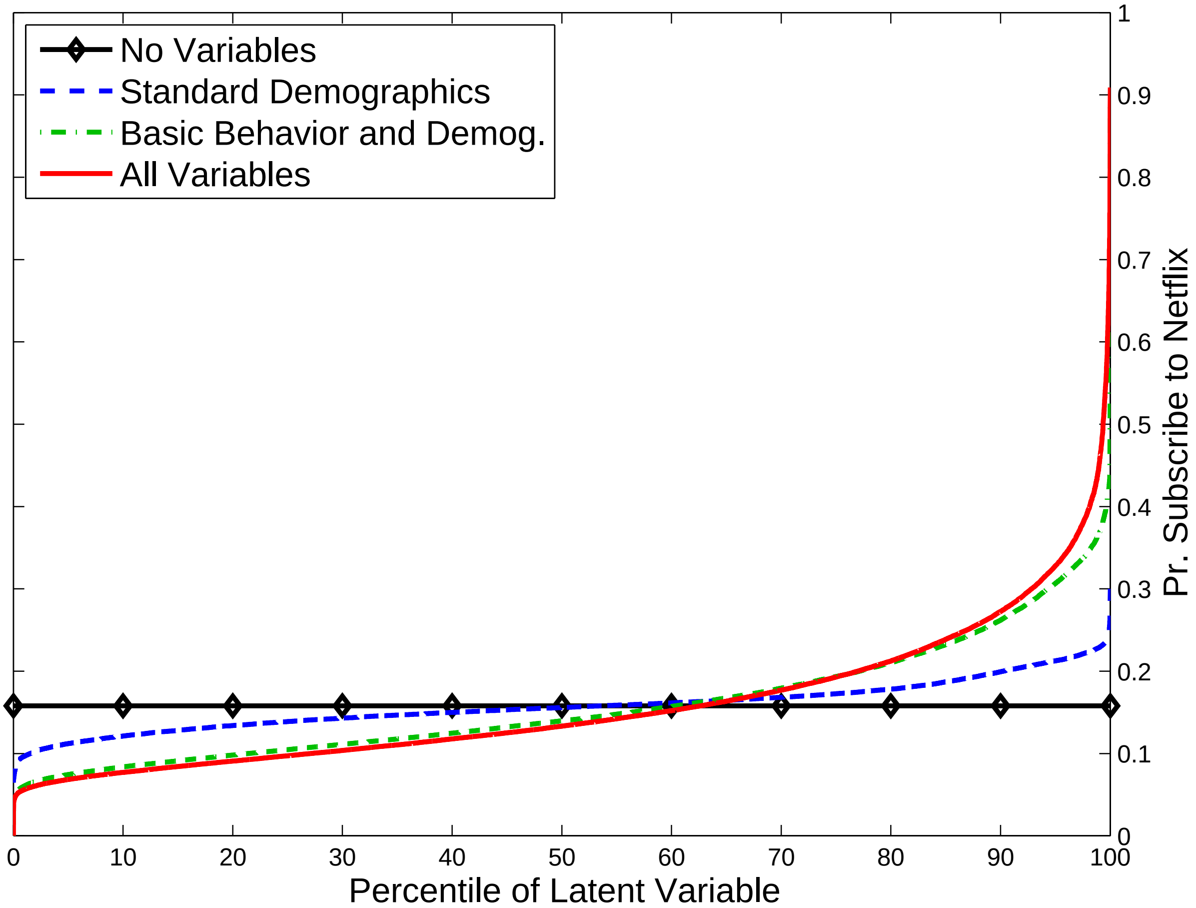


Figure 2: Range of Predicted Probabilities, Using Various Sets of Explanatory Variables

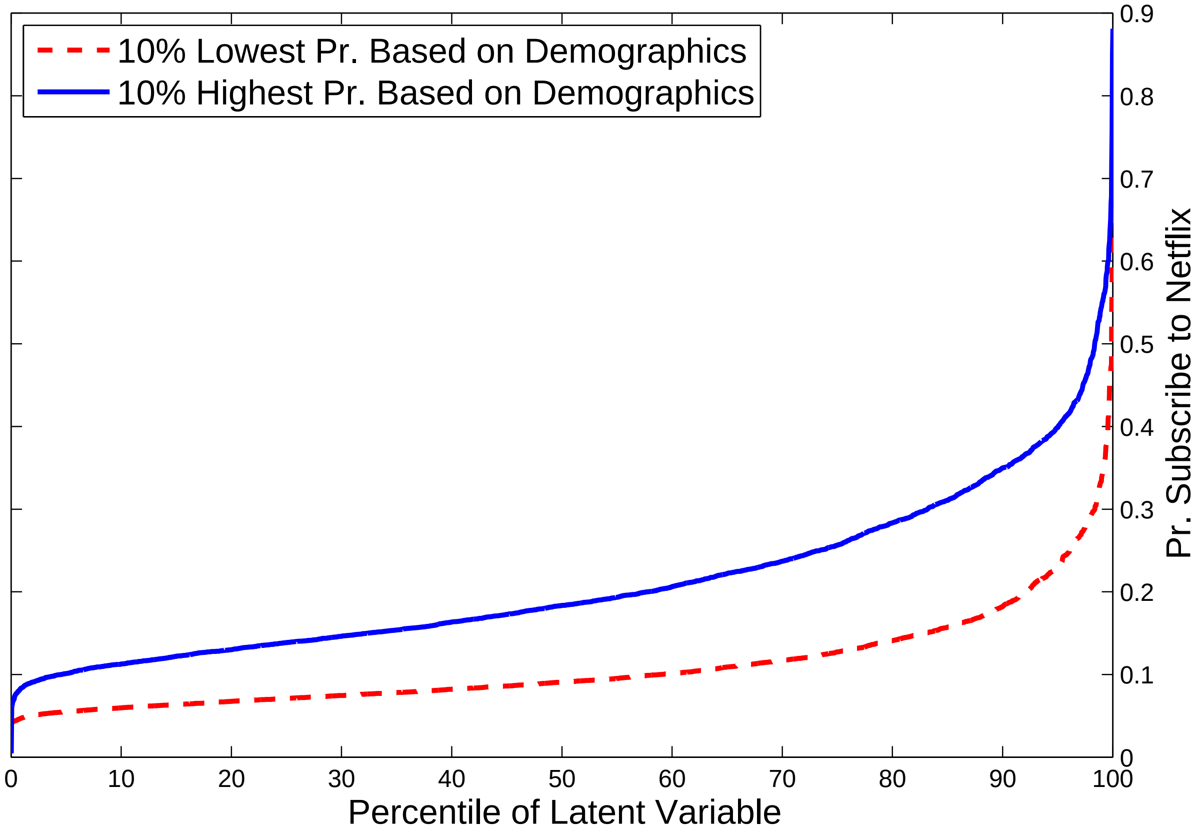


Figure 3: Range of Predicted Probabilities For Subsets of Individuals

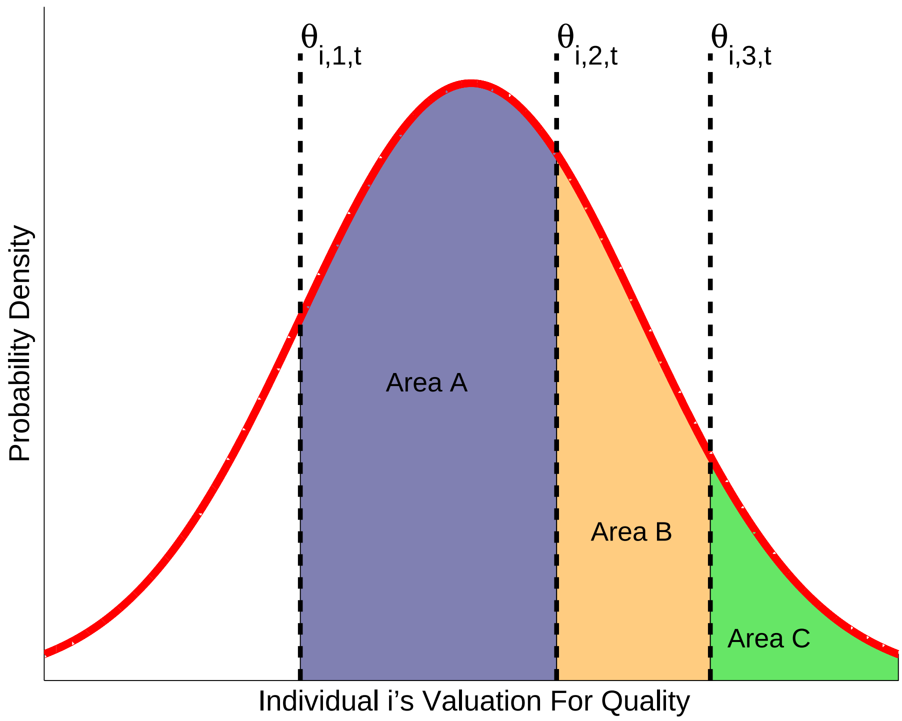


Figure 4: Graphical Depiction of Model

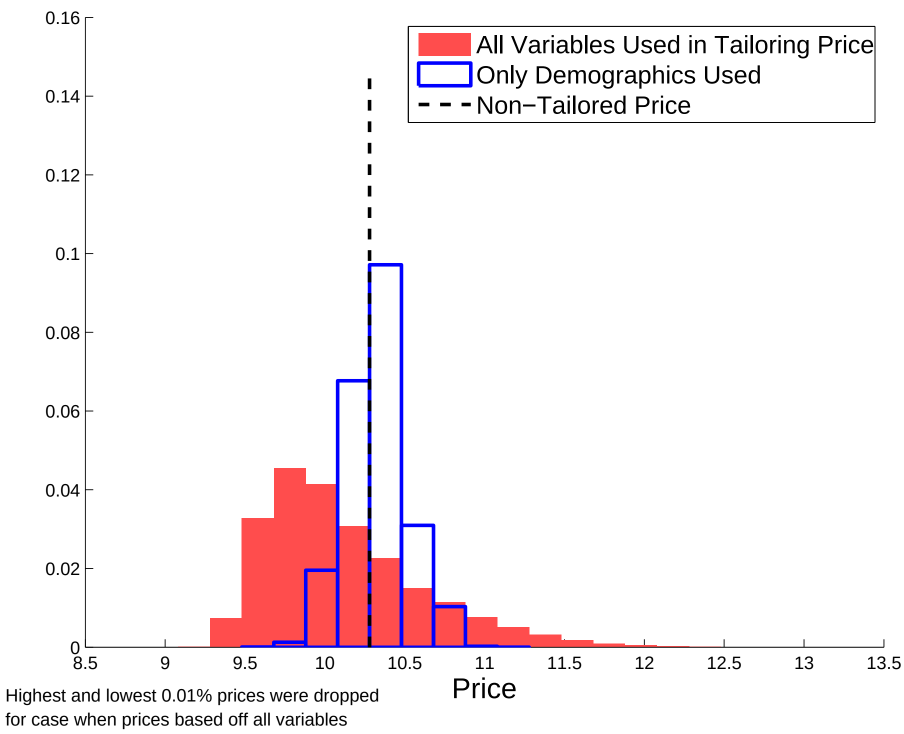


Figure 5: Histogram of Individually Tailored Prices - 1 DVD At-A-Time Plan

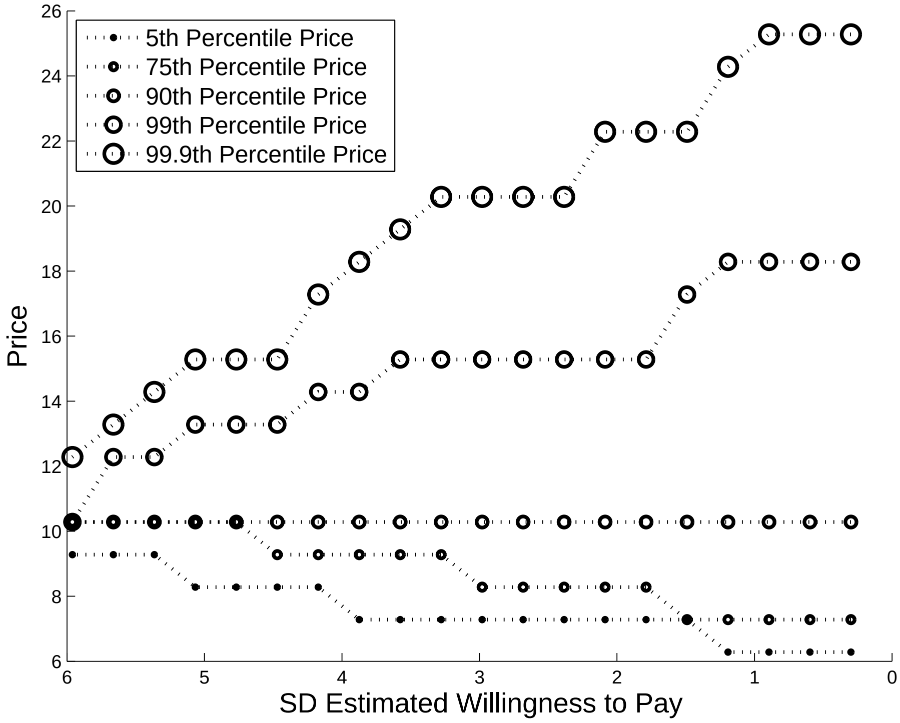


Figure 6: Graphical Depiction of Dispersion of Prices - 1 DVD At-A-Time Plan

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

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

Table 3: Simulated Changes in Various Outcomes Resulting From $1^{st}$ Degree PD

| | Percent Change When Price Based on: | |
|---|---|---|
| | Demographics | All Variables |
| Variable Profits | 0.14% | 1.39% |
| | (0.02) | (0.18) |
| Sales (DVDs At-A-Time) | 0.85% | 1.84% |
| | (1.03) | (0.95) |
| Subscribers | 0.17% | 1.03% |
| | (0.29) | (0.32) |
| Aggregate Consumer Surplus | 1.54% | 8.13% |
| | (0.76) | (0.91) |

Bootstrapped standard errors in parentheses. To speed computation, in model averaging only the 100 most likely models in the orginal sample were considered, and simulations employed a 10 cent increment.

Table 4: Simulated Changes in Various Outcomes Resulting From Tailored Discounts Off Optimized $2^{nd}$ Degree PD Prices

| | Percent Change When Discounts Based On: | |
|---|---|---|
| | Demographics | All Variables |
| Variable Profits | 0.05% | 0.44% |
| | (0.02) | (0.06) |
| Sales (DVDs At-A-Time) | 2.56% | 6.82% |
| | (0.76) | (0.75) |
| Subscribers | 2.41% | 6.64% |
| | (0.11) | (0.30) |
| Aggregate Consumer Surplus | 4.16% | 12.95% |
| | (0.58) | (0.85) |

Bootstrapped standard errors in parentheses. To speed computation, in model averaging only the 100 most likely models in the orginal sample were considered, and simulations employed a 10 cent increment.

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

| Price | 1 DVD At-A-Time | | 2 DVDs At-A-Time | | 3 DVDs At-A-Time | |
|---|---|---|---|---|---|---|
| Percentile | Demog. | All Variables | Demog. | All Variables | Demog. | All Variables |
| Lowest | −6.8% | −29.7% | −6.3% | −28.3% | −6.2% | −27.4% |
| | (0.8) | (15.9) | (0.7) | (16.1) | (0.7) | (16.1) |
| 0.1 | −5.8% | −9.7% | −5.7% | −9.3% | −5.6% | −9.0% |
| | (0.7) | (0.6) | (0.6) | (0.5) | (0.7) | (0.5) |
| 1 | −4.4% | −8.8% | −4.0% | −8.3% | −3.9% | −8.2% |
| | (0.7) | (0.5) | (0.5) | (0.5) | (0.5) | (0.4) |
| 10 | −2.4% | −6.8% | −2.3% | −6.7% | −2.3% | −6.5% |
| | (0.6) | (0.4) | (0.4) | (0.3) | (0.4) | (0.4) |
| 25 | −1.5% | −5.3% | −1.3% | −5.3% | −1.4% | −5.4% |
| | (0.5) | (0.5) | (0.4) | (0.5) | (0.5) | (0.5) |
| 50 | −0.5% | −2.9% | −0.3% | −3.0% | −0.6% | −3.1% |
| | (0.5) | (0.4) | (0.4) | (0.4) | (0.4) | (0.4) |
| 75 | 0.5% | 0.5% | 0.7% | 0.3% | 0.6% | 0.3% |
| | (0.5) | (0.6) | (0.4) | (0.5) | (0.4) | (0.5) |
| 90 | 2.4% | 4.9% | 2.3% | 4.7% | 2.0% | 4.2% |
| | (0.6) | (0.5) | (0.5) | (0.4) | (0.5) | (0.5) |
| 99 | 3.9% | 13.6% | 3.6% | 12.7% | 3.4% | 12.2% |
| | (0.7) | (1.0) | (0.6) | (0.8) | (0.6) | (0.8) |
| 99 | 5.3% | 21.9% | 5.0% | 20.7% | 4.8% | 19.8% |
| | (1.0) | (3.7) | (1.0) | (3.5) | (1.0) | (3.4) |
| Highest | 7.7% | 41.8% | 7.3% | 39.7% | 7.0% | 38.1% |
| | (1.9) | (10.7) | (1.8) | (10.1) | (1.8) | (9.8) |

Bootstrapped standard errors in parentheses. To speed computation, in model averaging only the 100 most likely models in the orginal sample were considered, and simulations employed a 10 cent increment.