First Degree Price Discrimination Using Big Data

Benjamin R. Shiller (Brandeis University)

Presented at Boston University

Mar 17, 2015
Outline for section 1

1. Introduction
2. Data
3. Predicting Subscription
4. Model Description
5. Model Results
6. Robustness Checks
7. Fit in Literature
8. Conclusion
Quantity discounts and group-specific pricing common in practice (and empirical work)

- Rely on self-sorting or easily verified attribute
- May only extract 1/3 surplus (SW, 2011)

1st degree (person-specific) PD extracts more. But,

- Requires knowledge of WTP
- Historically, rare in practice

However …
Imagine manager can hire one private detective per person

Might reveal information useful for predicting reservation values:

- Direct interest in related products
- Amount of free time
- Unobserved demographics (e.g. sexual orientation)
- Location by time of day
- Other correlated behaviors

“‘Big Data’” captures much of this information!

**Empirical Questions:**

(1) Can behavior reveal WTP, enabling personalized pricing?
(2) Is it a break from the past: i.e. better than demographics (long useable)?
Motivating Research Strategy

Possible Research Strategies (and Problems)

- Use aggregate data
  - Problem: Personalized pricing intentionally inconspicuous. Not clear which/how many firms use it
  - Problem: Market in transition

- Before/after comparison at single firm
  - Problem: Decision to use personalized pricing endogenous

- Estimate individual-level demand in market W/O personalized pricing. Simulate counterfactual with personalized pricing
  - Advantage: Provide method for 1st PD w. massive data
    - Done poorly, personalized pricing ↓ profit
  - Difficulty: Requires purchase decisions and web-behavior in same dataset
Advantages of Netflix

- Overcomes major data challenge: Netflix subscription can be imputed in web-browsing data

- Netflix, which sells online, could tailor price to individual

- Can price discriminate
  - Already uses 2nd degree PD
Outline of Research Strategy

First Analysis - Non-Structural

- Estimate ability of diff. variable sets to predict subscription
  - Standard demographics
  - Basic web-browsing behavior
  - Visits to each of 5,000 websites
    - Reflects unobserved traits and interests in related products
    - Requires machine-learning methods with one tweak (OMA method) to address over-fitting/high dimensionality
  - Verify fit in holdout sample

Second Analysis - Structural

- Estimate individual-level demand
- Then simulate outcomes under personalized pricing
- To my knowledge first paper to combine machine learning with structural economic modeling
Preview of Main Results

First analysis

- Unconditional probability consume Netflix - 16%
- Demographics based predicted probabilities range from 6% to 30%
- Web-browsing based probabilities range from \( \approx 0 \) to 99.8%
  - Variables with most predictive power are intuitive

Second analysis

- Profit increase from 1\(^{st}\) degree PD (above 2\(^{nd}\) degree)
  - Using demographics: 0.8%
  - Using all variables: 12.2%
- Some consumers pay twice what others do for same product
Outline for section 2

1. Introduction
2. Data
3. Predicting Subscription
4. Model Description
5. Model Results
6. Robustness Checks
7. Fit in Literature
8. Conclusion
Raw Data

**Raw data source:** ComScore 2006 machine-user level microdata

For each of over 60,000 users:

- Top-level domain/timestamp/duration
- Online transaction details
- # pages within domain visited
- Referring website
- Demographics
Aggregation to Cross Section and Cleaning

- Keep popular sites sans popups, malware, pornography, etc.

- Aggregate data to cross-section
  - \# times user visits each website
  - When user tends to use internet
  - Internet use intensity

- Impute subscription - subscriber if averages ≥ 2 pages per visit
  - Implies 15.75% subscribe, within 1% of auxiliary estimate of fraction households subscribing
Final Cross-Sectional Dataset

**Three categories of variables**

- **Standard demographics:** race/ethnicity, \#(children), household income, oldest householders age, household size, local population density, census region

- **Basic web behavior:** total website visits, \# unique transactions, % browsing by time of day and day of week, \#(broadband)

- **Detailed web behavior:** total visits to each website separately (4,789 variables in total)
Outline for section 3

1. Introduction
2. Data
3. Predicting Subscription
4. Model Description
5. Model Results
6. Robustness Checks
7. Fit in Literature
8. Conclusion
Overview of Section

In this section:

- Compare how well various sets of variables predict subscription
  - Standard demographics
  - Basic web behavior
  - Detailed web behavior

- Explain machine learning technique used
## Demographics and Basic Behavior

### Probit - Dependent Variable is I(Subscribe)

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

† Explanatory variables normalized, insignificant variables not shown
Websites - One at a Time

Analysis

- Each website separately is added to demographic and basic web behavior variables
- Evaluated based on p-value

Results:

- 29% significant at 5% level
- 18% significant at 1% level

Far more than expected by chance alone
Best Explaining Websites

<table>
<thead>
<tr>
<th>Rank</th>
<th>Website</th>
<th>Rank</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>amazon</td>
<td>14</td>
<td>pricegrabber</td>
</tr>
<tr>
<td>2</td>
<td>bizrate</td>
<td>15</td>
<td>wikipedia.org</td>
</tr>
<tr>
<td>3</td>
<td>imdb</td>
<td>16</td>
<td>smarter</td>
</tr>
<tr>
<td>4</td>
<td>shopping</td>
<td>17</td>
<td>hoovers*</td>
</tr>
<tr>
<td>5</td>
<td>dealtime</td>
<td>18</td>
<td>alibris</td>
</tr>
<tr>
<td>6</td>
<td>citysearch</td>
<td>19</td>
<td>epinions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>prnewswire.com</td>
</tr>
</tbody>
</table>

Websites intuitively visited by:

- Movie lovers
- Discount shoppers
- Frequent mail-orderers
- Rare products
- Internet savvy

*Hoovers appears to be a similar company to ComScore
Joint Prediction and Overfitting Problem

Two conceptual problems from overfitting

- Too many variables included (some significant by chance)
  - Best model not chosen - less complex model yields better predictions in holdout sample

Try to pick best model. Not choosing best model underestimates potential gain from 1st degree PD

- Best-fitting model in-sample underestimates standard deviation of error term
  - Overestimates ability of model to precisely predict WTP
    - Biased upwards estimate of profits under 1st degree PD
    - Biases upwards range of predicted prob. at observed prices

Re-estimate std(error) in whatever model is chosen using holdout sample. Then model reflects true std(error)
Choosing Model

- Forward step-wise reg.
- After 17\textsuperscript{th} web variable, out of sample fit ↓
  - Some ”next best” contain useful info
    - 48\% of 50 ”next best” improve out of sample fit when added as 18\textsuperscript{th} variable
- Model averaging (w. diff var in 18\textsuperscript{th} position) addresses model uncertainty, improves predictions
Econometric Review - Probit

- Binary outcome $Y_i = 1$ (vs. 0) iff latent variable $Y_i^* = \hat{Y}_i^* + \epsilon$ exceeds threshold $\mu$

- $\text{Prob}(Y_i = 1) = \text{Prob}(Y_i^* > \mu) = 1 - \phi(\hat{Y}_i^* - \hat{\mu})$

- To get model averaging (MA) estimate of $\text{Prob}(Y_i = 1)$, average $(\hat{Y}_i^* - \hat{\mu})$ across models, yielding $(\bar{Y}_i^* - \bar{\mu})$

**Problem:** Mechanically, this doesn’t account for the increased precision that should come from including multiple models (more variables) together via model averaging

- Details next
• Averaging models incorporates more info in prediction, hence std(error) should ↓ yielding more extreme probabilities

• But std(error) fixed = 1 in Probit
  • Probability determined by relative magnitudes of \((\bar{Y}_i^* - \bar{\mu})\) and \(std(\epsilon)\)
  • So instead of ↓ \(std(\epsilon)\), rather ↑ scaling of \((\bar{Y}_i^* - \bar{\mu})\)
    • Accomplishes same thing: more extreme probabilities
    • But doesn’t automatically happen from just averaging models
Ordered choice Model Average (OMA) Method

**Q:** How to increase scaling by *right* amount?

**A:** (1) Estimate series of models, and average \((\hat{Y}_i^* - \hat{\mu})\) for some number of best models to get \((\bar{Y}_i^* - \bar{\mu})\).

(2) In holdout sample, run new Probit with \((\bar{Y}_i^* - \bar{\mu})\) as sole explanatory variable. Its coefficient gives *right* amount to scale up

**Note:** Rescaling magnitudes so has *right* size (relative to error) in holdout sample removes any bias, including from:

- Not capturing increased precision from model averaging
- Using overfit model (2nd conceptual problem stated earlier)
Fit of Full Model

Predicted Probability
Actual Probability of Group

Pr. Subscribe to Netflix vs. Percentile of Latent Variable
Relative Predictive Abilities

- No Variables
- Standard Demographics
- Basic Behavior and Demog.
- All Variables

Percentile of Latent Variable vs. Pr. Subscribe to Netflix
Explained by Geography? - No
Predictions of Full Model Given Demographics’ Prediction

- Lowest 10% probabilities according to demographics range from 6.5% to 12%
- Miss much information captured in web-browsing data
Two categories of plans Netflix offered

- Unlimited # DVDs sent each month, but can only possess at one time:
  - 1 DVD ($9.99)
  - 2 DVDs ($14.99)
  - 3 DVDs ($17.99)
  - 4 DVDs ($23.99) (almost no one chooses this plan)

- Limited # sent each month (very unpopular in data)
  - Assume they would choose an unlimited plan had limited plans not existed
Graphical Model Intuition

- Area A
- Area B
- Area C

Probability Density

Individual i’s Valuation For Quality

$\theta_{i,1,t}$

$\theta_{i,2,t}$

$\theta_{i,3,t}$
Utility Function

Standard utility function for $2^{nd}$ degree PD demand:

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

- $y_i$ - measure of individual i’s valuation for product type
- $q_j$ - quality of tier $j$
- $\alpha$ - price sensitivity
- $I_i$ - income (drops out later)
- $P_j$ - price of tier $j$
j preferred to k iff:

\[ y_i \geq \alpha \frac{P_j - P_k}{q_j - q_k} \]

j chosen if preferred to neighboring options:

\[ \alpha \frac{P_j - P_{j-1}}{q_j - q_{j-1}} \leq y_i < \alpha \frac{P_{j+1} - P_j}{q_{j+1} - q_j} \]

Replace \( y_i \) with regression expression, and normalizations:

\[ \alpha \lambda_j P_{\Delta j} \leq \beta_0 + X_i \beta + \sigma \epsilon_i < \alpha \lambda_{j+1} P_{\Delta j+1} \]

Rearrange isolating \( \epsilon \):

\[ \theta_{i,j} \leq \epsilon_i < \theta_{i,j+1}, \text{ where} \]

\[ \theta_{i,j} = -\beta_0 + \lambda_j P_{\Delta j} - X_i \beta = \mu_j - X_i \beta \]

Probability choosing j (used in maximum likelihood):

\[ s_{i,j} = F (\theta_{i,j+1}) - F (\theta_{i,j}) \]
\( j \) preferred to \( k \) iff:

\[
y_i \geq \alpha \frac{P_j - P_k}{q_j - q_k}
\]

\( j \) chosen if preferred to neighboring options:

\[
\alpha \frac{P_j - P_{j-1}}{q_j - q_{j-1}} \leq y_i < \alpha \frac{P_{j+1} - P_j}{q_{j+1} - q_j}
\]

Replace \( y_i \) with regression expression, and normalizations:

\[
\alpha \lambda_j P_{\Delta j} \leq \beta_0 + X_i \beta + \sigma \epsilon_i < \alpha \lambda_{j+1} P_{\Delta j+1}
\]

Rearrange isolating \( \epsilon \):

\[
\theta_{i,j} \leq \epsilon_i < \theta_{i,j+1}, \text{ where } \theta_{i,j} = -\beta_0 + \lambda_j P_{\Delta j} - X_i \beta = \mu_j - X_i \beta
\]

Probability choosing \( j \) (used in maximum likelihood):

\[
s_{i,j} = F(\theta_{i,j+1}) - F(\theta_{i,j})
\]
$j$ preferred to $k$ iff:

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

$j$ chosen if preferred to neighboring options:

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

Replace $y_i$ with regression expression, and normalizations:

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

Rearrange isolating $\epsilon$:

$$\theta_{i,j} \leq \epsilon_i < \theta_{i,j+1}, \text{ where}$$

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

Probability choosing $j$ (used in maximum likelihood):

$$s_{i,j} = F(\theta_{i,j+1}) - F(\theta_{i,j})$$
j preferred to k iff:

\[ y_i \geq \alpha \frac{P_j - P_k}{q_j - q_k} \]

j chosen if preferred to neighboring options:

\[ \alpha \frac{P_j - P_{j-1}}{q_j - q_{j-1}} \leq y_i < \alpha \frac{P_{j+1} - P_j}{q_{j+1} - q_j} \]

Replace \( y_i \) with regression expression, and normalizations:

\[ \alpha \lambda_j P_{\Delta j} \leq \beta_0 + X_i \beta + \sigma \epsilon_i < \alpha \lambda_{j+1} P_{\Delta j + 1} \]

Rearrange isolating \( \epsilon \):

\[ \theta_{i,j} \leq \epsilon_i < \theta_{i,j+1}, \text{ where} \]

\[ \theta_{i,j} = -\beta_0 + \lambda_j P_{\Delta j} - X_i \beta = \mu_j - X_i \beta \]

Probability choosing j (used in maximum likelihood):

\[ s_{i,j} = F(\theta_{i,j+1}) - F(\theta_{i,j}) \]
Taking Stock

- Model averaging demonstrated to improve predictive fit
  - E.g. Netflix Prize Challenge
  - Helpful for trying many thousands of potential explanatory variables.

- But not immediately compatible with the structural framework

- The **Ordered-Choice Model Average (OMA)** method (introduced earlier) bridges this gap, making compatible.

- Will allow evaluation of ability of newly available browsing data to predict WTP.
Returning to Graphical Model Intuition

- Suppose knew $\frac{\partial \theta_{i,j,t}}{\partial P_{k,t}}$

- Can determine optimal set of prices for individual $i$
  \[ \arg\max_{P_{i,1}, P_{i,2}, P_{i,3}} \sum_{j=1}^{3} (P_{i,j} - c_j) \left( F(\theta_{i,j+1}) - F(\theta_{i,j}) \right) \]

- Similarly find optimal prices when all individuals charged same
Estimating Price Sensitivity

- Since no price variation, price sensitivity not identified in estimation
- But can use supply-side conditions to estimate it
  - Price sensitivity monotonic function of one model parameter
  - Once this parameter specified, all parameters known
  - Can find its value using supply side conditions
    - I.e. find value making observed prices optimal
Outline for section 5

1. Introduction
2. Data
3. Predicting Subscription
4. Model Description
5. Model Results
6. Robustness Checks
7. Fit in Literature
8. Conclusion
## Simulated Changes Resulting From $1^{st}$ Degree PD

<table>
<thead>
<tr>
<th></th>
<th>Percent Change When Price Based on:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Demographics</td>
</tr>
<tr>
<td><strong>Total Profits</strong></td>
<td>0.79%</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
</tr>
<tr>
<td><strong>Sales (DVDs At-a-Time)</strong></td>
<td>0.85%</td>
</tr>
<tr>
<td></td>
<td>(1.09)</td>
</tr>
<tr>
<td><strong>Subscribers</strong></td>
<td>0.17%</td>
</tr>
<tr>
<td></td>
<td>(0.62)</td>
</tr>
<tr>
<td><strong>Aggregate Consumer Surplus</strong></td>
<td>$-0.18%$</td>
</tr>
<tr>
<td></td>
<td>(1.12)</td>
</tr>
</tbody>
</table>

Bootstrapped standard errors in parentheses
### Simulated Changes Resulting From Tailored Discounts Off Optimized 2\textsuperscript{nd} Degree PD Prices

<table>
<thead>
<tr>
<th></th>
<th>Percent Change When Discounts Based On:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Demographics</td>
<td>All Variables</td>
</tr>
<tr>
<td>Total Profits</td>
<td>0.28%</td>
<td>3.19%</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.51)</td>
</tr>
<tr>
<td>Sales (DVDs At-a-Time)</td>
<td>2.56%</td>
<td>8.11%</td>
</tr>
<tr>
<td></td>
<td>(0.71)</td>
<td>(0.89)</td>
</tr>
<tr>
<td>Subscribers</td>
<td>2.53%</td>
<td>7.61%</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.55)</td>
</tr>
<tr>
<td>Aggregate Consumer Surplus</td>
<td>3.38%</td>
<td>8.17%</td>
</tr>
<tr>
<td></td>
<td>(0.57)</td>
<td>(0.64)</td>
</tr>
</tbody>
</table>

Bootstrapped standard errors in parentheses
Histograms of Tailored Prices

- Red: All Variables Used in Tailoring Price
- Blue: Only Demographics Used
- Dashed: Non-Tailored Price

Highest and lowest 0.01% prices were dropped for cases when prices based off all variables.
- Highest price 61% above non-tailored optimal price
- Lowest price 22% below non-tailored optimal price
- Highest price roughly twice lowest price
- Median prices lower than non-tailored price

<table>
<thead>
<tr>
<th>Price Percentile</th>
<th>1 DVD At-A-Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Demog.</td>
</tr>
<tr>
<td>Lowest</td>
<td>-6.8%</td>
</tr>
<tr>
<td></td>
<td>(0.6)</td>
</tr>
<tr>
<td>1</td>
<td>-4.4%</td>
</tr>
<tr>
<td></td>
<td>(0.4)</td>
</tr>
<tr>
<td>25</td>
<td>-1.5%</td>
</tr>
<tr>
<td></td>
<td>(0.5)</td>
</tr>
<tr>
<td>50</td>
<td>-0.5%</td>
</tr>
<tr>
<td></td>
<td>(0.5)</td>
</tr>
<tr>
<td>75</td>
<td>0.5%</td>
</tr>
<tr>
<td></td>
<td>(0.5)</td>
</tr>
<tr>
<td>90</td>
<td>2.4%</td>
</tr>
<tr>
<td></td>
<td>(0.3)</td>
</tr>
<tr>
<td>99</td>
<td>3.9%</td>
</tr>
<tr>
<td></td>
<td>(0.6)</td>
</tr>
<tr>
<td>99.9</td>
<td>5.3%</td>
</tr>
<tr>
<td></td>
<td>(0.9)</td>
</tr>
<tr>
<td>Highest</td>
<td>7.7%</td>
</tr>
<tr>
<td></td>
<td>(1.7)</td>
</tr>
</tbody>
</table>
Firms may predict WTP better with geolocation patterns (via smartphones) or contextual variables (via twitter, email, texts)
Outline for section 6

1. Introduction
2. Data
3. Predicting Subscription
4. Model Description
5. Model Results
6. Robustness Checks
7. Fit in Literature
8. Conclusion
Two Concerns

1. Thresholds depend on few hundred obs. tier purchases.
   - **Solution:** Re-run binary model - not buy vs. buy any
     - In ordered model, higher tier buyers would buy lower tier
     - Simulate profit ↑ from personalized pricing if only one tier.

2. Price sensitivity $\alpha$ imputed from static profit-max FOC.
   - **Concern** - Netflix maybe underpricing to max long-run profits
   - **Solution** - Test sensitivity, trying $\alpha$ implying profit-maximizing prices twice observed prices
     - If results similar, suggests *percent change* in profits depends primarily on finding differences across consumers, NOT magnitude of the price sensitivity.
## Robustness Checks Results

### Robustness Checks

Percent Increase in Profits From Personalized Pricing, When Based On:

<table>
<thead>
<tr>
<th>Model</th>
<th>Demographics</th>
<th>All Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Model</td>
<td>0.79%</td>
<td>12.18%</td>
</tr>
<tr>
<td>Robustness Checks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Tier Sold</td>
<td>0.81%</td>
<td>12.27%</td>
</tr>
<tr>
<td>Diff. Price Sensitivity</td>
<td>1.18%</td>
<td>12.54%</td>
</tr>
</tbody>
</table>
Outline for section 7

1. Introduction
2. Data
3. Predicting Subscription
4. Model Description
5. Model Results
6. Robustness Checks
7. Fit in Literature
8. Conclusion
Related Economics Literature

**Monopoly theory**

- $\pi(1^{st}) = \text{full surplus} > \max(\pi(2^{nd}), \pi(3^{rd}))$ if heterogeneous valuations
- Bundling extracts full surplus if $\text{MC} = 0$ and differences in valuations across consumers not too persistent [Bakos & Brynjolfsson, 1999]

**Empirical** - Many examples of non-$1^{st}$ degree PD

- 1st Degree. Only aware of one, on university tuition (Waldfogel, 2014)
- 3$^{rd}$ degree PD [Graddy, 1995; Langer, 2011]
- 2$^{nd}$ degree PD [Crawford & Shum, 2007; McManus 2008]
- Intertemporal pricing [Nair, 2007]
- Bundling [Chu, Leslie, & Sorensen, 2011; Shiller & Waldfogel, 2011]
Computer science

- Evidence a few firms 1\textsuperscript{st} degree PDing on web [Mikians et al. 2012; Hannak et al. 2014]
  - Best known places: Staples, Orbitz
  - Since published, I've seen other examples

Marketing

- Rossi et al. (1996) spawned empirical literature on individual-level pricing based on prior purchase history of same product

However

- Fudenberg and Villas-Boas (2005) and Acquisti and Varian (2005)- theoretically such pricing not more profitable if consumers forward-looking/obscure behavior
- But if pricing instead based on web browsing - Too many rules!
  - Bounded rationality/no simple heuristics
  - Not worth changing hundreds of online behaviors
<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
</tbody>
</table>
Conclusion

Implications

- Monopoly profits/increased efficiency
- For duopoly, can raise or lower profits [Corts, 1998; Thisse and Vives, 1988]
- Wasted effort masking as low WTP type
- Privacy/data property rights
- Labor participation
- Fair?

Widspread?

- PD exists even in seemingly competitive market [Graddy, 1995; Shepard, 1991]
Targeted messages:

- Already done a bit with "‘landing page optimization’"

Targeted behavioral economics/marketing

- Behavioral economics show consumers make some "‘mistakes,’" i.e aren’t acting rationally
  - Several have pointed out several common mistakes made, and ways to exploit these mistakes to raise profits
- In future, firms could tailor strategies to each individual’s "‘mistakes’"