

---

# Targeted Advertising in the Information Age

---

Author: Stephen Douglas BRUESTLE

Hometown: Pennington, NEW JERSEY

Previous degrees:

Economics (M.A.), UNIVERSITY OF VIRGINIA  
(January 2009)

Economic History (B.A.) and Mathematics (B.A.),  
LAFAYETTE COLLEGE (May 2007)

A Dissertation (*or Thesis*) presented to the Graduate Faculty of the  
University of Virginia in Candidacy for the Degree of Doctor of  
Philosophy.

Department of ECONOMICS  
UNIVERSITY OF VIRGINIA

August 2013

Advisors:

Simon ANDERSON

Federico CILIBERTO

Nathan LARSON

4th Reader:

Natasha Zhang FOUTZ

Copyright ©2013 by Stephen D. Bruestle.  
All rights reserved.

## Abstract

In Chapter 1, “As Webpages Get Narrower, Do Ads Get Nicher? An Online Field Experiment in Google Contextual Ads,” Firms target their advertisements to the consumer segments delivered by webpages. I develop an auction model where firms target segments of heterogeneous consumers. From this, I derive an empirical framework, which I use to test whether more niche or more general ads win the auction for more narrowly-focused webpages. To do this, I create many differentiated webpages in an experimental fashion and observe the Google text ads that are placed on them. Then, I compare general webpages, such as a ‘Ford’ webpage, with more narrowly-focused webpages, such as a ‘Ford Truck’ webpage, by using a measures of ad niche-ness. I use a Hierarchical Latent Dirichlet Allocation algorithm from the machine-learning literature to create a robust measure of ad niche-ness. My results show a U-shaped relationship between webpage narrowness and ad niche-ness: Ads for less niche products tend to appear on moderately narrow-focused webpages.

In Chapter 2, “Imperfect Targeting of Advertising and Privacy Regulations,” I investigate how privacy regulations affect consumer welfare through advertising. Tougher privacy regulations reduce the accuracy of information collected on consumers. This discourages targeted advertising. When firms target advertise, privacy regulations ambiguously affect welfare. Less accurate information decreases welfare by inducing a smaller, less-targeted selection of products. Yet less accurate information increases welfare by inducing fewer annoying ads, even without any pricing effects. In extensions, I find that tougher privacy regulations increase the product selection benefit and the ad annoyance cost through reducing the ad price; but greater marketing costs have the opposite effect; and ad avoidance ads has no effect.

In Chapter 3, “Showing Ads To The Wrong Consumers: Strategic Ad Platform

Inefficiency In Online Targeted Advertising,” I find that an ad platform has an incentive to induce lower product prices. Consumers pay a search cost when clicking on an ad. To induce consumer clicking, an ad platform adopts a targeting strategy that induces the merchant to lower its price. This involves showing the ad to some consumers who it rationally expects not to buy the product and not showing the same ad to other consumers who it would rationally expect to buy the product.

JEL Codes: M37, D83, L19, C93, C81

KEYWORDS: targeted advertising, online advertising platforms, advertisement pricing, online field experiment, hierarchical latent dirichlet allocation, privacy regulations

## Dedication

To all those who encourage and inspired me:

To my advisors; for their endless patience and encouragement.

To my wife; for her honesty and loving support.

To my expected child; with whom I plan on sharing my love of education.

To my parents; for their continued belief in me and my abilities.

To my brother; for his misguided insistence that engineering is better than economics.

To my sister; for her generosity, loudness, and love of mathematics.

To my grandmother; for her endless creativity and love of life.

To my grandfather; for his support and insistence that price is the same as quality.

To my grandma; for her endless perseverance, wisdom, and cooking.

To my grandpa; for the twinkle in his eyes and his sweet tooth.

To you; for reading my thesis and your forgiveness of any mistakes.

# Table of Contents

	Page
<b>I As Webpages Get Narrower, Do Ads Get Nichier? An On-line Field Experiment in Google Contextual Ads</b>	<b>1</b>
1 Introduction	1
2 Hotelling Auction for Targeted Advertising	6
3 Data & Experimental Design	12
4 Topic Modeling	18
5 Results	29
6 Conclusion	43
References	43
Figures and Tables	49
Appendix A Many Firms / Endogenous Prices	67
<b>II Imperfect Targeting of Advertising and Privacy Regulations</b>	<b>70</b>
1 Introduction	70
2 Literature Review	75
3 Game	77

4	Equilibrium	78
5	Impact of Signal Accuracy on Welfare	85
6	Marketing Costs	89
7	Endogenous Advertising Prices	95
8	Ad Retention	100
9	Common Signals	111
10	Conclusion	117
	Appendix A Comparative Statics on the Marketing Cost Threshold	122
	Appendix B Proof of Proposition 4	123
III	Showing Ads to the Wrong Consumers: Strategic Ad Platform Inefficiency in Online Targeted Advertising	124
1	Introduction	124
2	Literature Review	128
3	The Model	131
4	Each Consumer's Clicking Decision	134
5	The Merchant's Pricing Decision	135
6	The Ad Platform's Targeting Decision	136
7	Examination of Equilibrium Payoffs	141

<b>8</b>	<b>Take-it-or-leave-it Offer Advertising</b>	<b>143</b>
<b>9</b>	<b>Concluding Remarks</b>	<b>147</b>
<b>Appendix A</b>	<b>Mathematical Proofs and Lemmas</b>	<b>152</b>
<b>Appendix B</b>	<b>Functional Form Conditions</b>	<b>164</b>



## Chapter I

---

# As Webpages Get Narrower, Do Ads Get Nichier? An Online Field Experiment in Google Contextual Ads

### 1 Introduction

Whenever we update our statuses on Facebook, search for videos of cats on YouTube, or email our loved ones on GMail, we see advertisements targeted toward our habits. Our cyber-world is teeming with *targeted advertisements* where different ads are shown to different consumers based on their tastes, locations, or demographics. While much technology and innovation has improved advertisers' abilities to select to whom they send their ads, some of the webpages on which they advertise continue to draw in large, heterogeneous segments of consumers. While there are webpages on narrow topics, such as a webpage on Texas Hold'em Poker strategies, there are many webpages on broad topics (such as webpages dedicated toward card games in general or main index pages linking to more narrowly-focused webpages). Advertisers bid for ad space on both these more generally-focused and more narrowly-focused webpages.

The ads that win the online auction for a more narrowly-focused webpage may not be for niche products. To keep my language simple, I will refer to ads for more niche products as *more niche ads*. For example, what if no niche products

exist that could cater to those who would visit a more narrowly-focused webpage? Would a close product win the auction, or would a more general one succeed? For example, while there are many products targeted toward Nascar drivers, like seatbelt harnesses and racing helmets, and there are many products targeted toward opera singers, like libretos and voice lessons, there may be few, if any, products targeted toward opera singing Nascar drivers. A more narrowly-focused webpage dedicated to Nascar drivers singing the most famous arias therefore may not get more niche ads than a webpage dedicated to Nascar drivers or opera singers. Instead, we could see on the webpage ads targeted toward opera singers, ads targeted toward Nascar drivers, or ads targeted toward a broader audience, such as ads for credit cards or insurance.

In this paper, I test whether ads for products targeted toward smaller market segments focused on marketing niche products or ads for products targeted toward larger market segments focused on marketing general products would win the auction for more narrowly-focused webpages. For example, would a webpage featuring 'Ford Truck Tires' show more ads for Ford Truck snow tires (a more niche product than Ford Trucks) or more ads for Ford cars (a less niche product than Ford Trucks) than the parent 'Ford' webpage. The purpose of this paper is not examine Google's strategy, because I assume they are selling ad space in a simple auction, but to explore the targeting behavior of firms as a function of the preference of each consumer, which is a preference revealed by his visit to a particular webpage.

Despite the recent theoretical economic literature on targeted advertising (see for example: Bergemann and Bonatti, 2011; Iyer et al., 2005; Johnson, 2013), there has not been empirical analyse of targeted advertising. The one expectation is Chandra (2009). He uses the number of competing newspapers as a proxy for how targeted or niche advertising is in a city, which shows that more competing newspapers lowers circulation prices and raises advertising prices. This paper extends the discussion by examining the kinds of ads delivered to a more narrowly-focused

audience.

I start this paper by developing a Hotelling model where firms bid on advertising toward segments (webpages) of heterogeneous consumers in a second-price auction. I use this model to investigate whether an ad for a product targeted toward a smaller market segment selling a niche product or an ad for a product targeted toward a larger market segment selling a general product would win the auction for more narrowly-focused webpages. Using an example, I demonstrate the possibility that the niche-ness of ads served on a webpage can vary non-monotonically with the narrowness of the webpage content. Because of the ambiguity of the theory, I test this empirically.

To test this empirical question, I run a novel field experiment on Google text-based, webpage advertisements. I create webpages that each contain: (1) unique webpage content and (2) a space for one Google text advertisement. For each webpage, Google observes the content and auctions advertising on that webpage based the perceived topic of the webpage.

My experiment is somewhat similar to two online field experiments run by Randall A. Lewis and David H. Reiley at Yahoo! Research (Lewis and Reiley, 2011, 2012). In these experiments, they identify a set of consumers in both Yahoo! and an online retailer's database. Then, they randomly assign these consumers different amounts of advertising. They test how changing online advertising behavior (how many ads a consumer sees) affects consumer behavior (purchasing decisions). My research likewise focuses on the complex subject of online advertising. In this experiment, I test how changing online consumer behavior (revealed to firms by the choice of a webpage) affects advertiser behavior (which firm wins the bid for advertising).

One advantage of creating my own webpages instead of using previously existing webpages or other advertising media, like real magazines or TV stations, is I can control their content. The ad slot value of a preexisting webpage presumably

depends on a slew of variables, such as number of ads on a webpage or traffic, many of which are either unobservable to me or hard to measure and compare. By creating my own webpages, I can be sure that ad slot values depend on my content keywords and nothing else.

I chose to use content from websites that were currently being auctioned on Flippa. Flippa ([flippa.com](http://flippa.com)) is the largest online marketplace dedicated to buying and selling websites. For instance, nearly 26,000 websites were sold on Flippa in 2011, for a value of almost \$31 million. The advantage of using websites currently auctioned on Flippa is that the seller provides data on the highest bid (i.e. the price for the website), age of the website, ad revenue, website traffic, website ranking, and many more variables. I did not use previously-auctioned websites despite that data being available, because the buyer could have changed the website's characteristics after the auction.

I then ran an automated program to observe advertisements on my webpages. My data collection program:<sup>1</sup> (1) randomly brings up one of my webpages in the firefox browser; (2) it grabs the webpage content, the text ad, the date, and the time off the webpage; (3) it randomly jumps to another of my webpages in Firefox. Then it iterating back on (2) to grab the information off the new webpage.

Intuitively, a Lincoln Blackwood, which is Ford's 2002-03 all-black luxury pick-up truck which sold only 3,356 units, is a niche product. In contrast, a Toyota Corolla, which is the perennial best selling car in the world, is not. Therefore, an ad for Lincoln Blackwoods would be more niche than an ad for Toyota Corollas. And a webpage on Lincoln Blackwoods would be more narrowly-focused than a webpage on Toyota Corollas. The methodological problem is finding some scientific and systematic way of measuring how niche an ad is and how narrow a webpage is.

---

<sup>1</sup> I thank John, Thomas, and Anna Bruestle for their programming help on writing my data collection program. The program collects data from the firefox browser, not the source code, because Google programmers made it difficult, if not impossible, to get the ad directly from the source code. It was written in Javascript embedded into the webpages in a way such that Google would not be able to see the program.

This is not trivial, because the measure needs to be exogenous. Therefore, I cannot use number of units sold, price of the product, or price of the ad as proxy variables for niche-ness and narrowness, because these are all endogenous measures.

While an ideal experiment might allow me to pick both the possible ads and the possible webpages, in this experiment, I cannot choose the possible ads. I instead observe real text ads created and bid on by real advertisers. Therefore, I cannot slightly vary the text in ads, to observe different levels of targeting. I cannot add the word ‘red’ or the word ‘truck’ to the text of an ad and see how it changes which webpages it gets posted on. Instead I observe the occurrences of different ads, some which contain the word ‘red’, some which contain the word ‘truck’, and some which do not contain either word. Therefore I create measures of advertisements to be able to identify which ads are similar and which ads are different.

I use a statistical method from the machine-learning literature to uncover the latent relationship in clusters of words to derive a measure for the niche-ness of an ad. This statistical method comes from a stream of machine-learning research known as *topic modeling* (see for example: Blei et al., 2003a; Griffiths and Steyvers, 2004; Minka and Lafferty, 2002; Teh et al., 2006b). Topic modeling algorithms are probabilistic algorithms for uncovering the underlying structure of a set of documents using hierarchical Bayesian analysis of the original texts. They are most often used to categorize documents based on observed patterns of words in the texts. This paper is the first economics application of a topic modeling algorithm.

In the past decade, most of the development of topic modeling has been from adaptations and applications of Blei et al.’s (2003a) Latent Dirichlet Allocation (LDA). This includes analyses of scientific abstracts (Griffiths and Steyvers, 2004) and newspaper archives (Wei and Croft, 2006). LDA is not only the most widely accepted topic model, but it is also the most powerful. A large part of the machine-

learning literature has focused on creating faster and more efficient algorithms for estimating the latent relationship between words and documents using the LDA model, including mean field variational inference (Blei et al., 2003a), collapsed variational inference (Teh et al., 2006a), expectation propagation (Minka and Lafferty, 2002)), and Gibbs sampling (Steyvers and Griffiths, 2006). In this paper, I focus on applying a topic model algorithm to answer an economics question, not developing a new estimation technique nor develop a new topic model.

The big advantages of LDA are: It allows for documents to be generated from multiple topics, it allows the topics of documents to be identified without having to create a new exogenous variable for each document, and it can be identified quickly. Many of the algorithms based on LDA only take a few hours to run on a set of tens of thousands of documents. For a good overview of the important topic models and the importance of LDA, see Blei and Lafferty (2009).

In this paper, I focus on a particular adaptation of LDA from Blei et al. (2003b) called Hierarchical Latent Dirichlet Allocation (HLDA) because it allows me to test the niche-ness and narrowness of documents (my ads and my webpages). HLDA imposes the additional assumption on the basic LDA model that the categories are hierarchical in nature. There is additionally one parent category, with a set of the most general words. Further, there is a set of children categories, with a set of more niche words. And there is a set of grandchildren categories, which are subcategories of the children categories and are composed of the most niche words.<sup>2</sup> I use HLDA by using it to estimate the level of the category of each word in each ad; the lower the level, the more niche the word in the ad. Using this measure, I find strong evidence for a non-monotonic relationship: The less niche ads tend to appear more on moderately narrow webpages.

---

<sup>2</sup> There can be as many levels to the tree as you choose, but I found that only three levels were useful for my dataset.

## 2 Hotelling Auction for Targeted Advertising

In this section, I develop a Hotelling model where firms bid on advertising toward segments (webpages) of heterogeneous consumers in a second-price auction. I use this model to investigate whether an ad for a product targeted toward a smaller market segment (niche product) or an ad for a product targeted toward a larger market segment (general products) would win the auction for a more narrowly-focused webpage. An advertiser for a niche product might bid more per customer for advertising on a narrow webpage that is targeted toward its small segment of potential buyers; or the advertiser might bid less for advertising on a narrow webpage that is targeted toward consumers that would not buy its product. Therefore, it is not obvious what kind of firm will tend to win the right to advertise on a narrower webpage.

I show an example where a general product wins the auction for advertising on a generally-focused webpage, a more niche product wins the auction for advertising on a more narrowly-focused webpage, and a general product wins the auction for advertising on the most narrowly-focused webpage.<sup>3</sup> I could have, however, just as easily have created a counter example where a niche product wins the auction for advertising on a generally-focused webpage, a general product wins the auction for advertising on a more narrowly-focused webpage, and a niche product wins the auction for advertising on the most narrowly-focused webpage. This shows that the relationship between the niche-ness of the advertised product and the narrowness of the webpage is not necessarily monotonic. Therefore, I test it as an empirical question.

In 2.1, I first develop a base Hotelling model where firms bid on advertising seeking to attract individual heterogeneous consumers in a second-price auction. Then, in 2.2, I solve for the firm's bidding strategy where each firm bids its value

---

<sup>3</sup> In A, I extend this to a market equilibrium by introducing endogenous product prices and many firms.

for advertising.<sup>4</sup> Next, in 2.3, I extend my base Hotelling model so that firms bid on segments (webpages) of consumers, instead of on individual consumers. I use this analysis to lay the theoretical foundation for my empirical analysis.

## 2.1 Basic Model

There are two firms  $j = 0, 1$  that each produce a horizontally differentiated good at a constant marginal cost, which is normalized to zero. Each firm  $j$  receives an i.i.d. random, horizontal product characteristic or location of  $x_j \sim U[-1, 1]$ , which is only known by firm  $j$ . For now, each firm  $j$  has an exogenous product price of  $p_j$ .<sup>5</sup> Each firm  $j$  simultaneously chooses its advertisement bidding function  $b_j : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  to maximize its expectation of its profit  $\Pi_j \equiv p_j \int_{\mathbb{R}} \omega_j(x) dx - \int_{\mathbb{R}} b_{1-j}(x) 1\{b_j(x) > b_{1-j}(x)\} dx$ , where  $\omega_j(x) = 1$  if consumer  $x$  buys product  $j$  and  $\omega_j(x) = 0$  if consumer  $x$  does not buy product  $x$ . Consumer  $x$  will only be able to choose to buy product  $j$  if  $b_j(x) > b_{1-j}$ .

There is a continuum of consumers with taste characteristics or locations distributed in a uniform density normalized to one along the real number line. Each consumer  $x \in \mathbb{R}$  chooses either to buy: (a) the product revealed to him through advertising or (b) some outside option to maximize his utility  $u(x)$ . Consumer  $x$  is only shown the product  $j = 0, 1$  where  $b_j(x) > b_{1-j}(x)$ . If consumer  $x$  buys good  $j$ , then he gets a utility of  $u_j(x) \equiv R - t|x_j - x| - p_j$ , and if consumer  $x$  takes the outside option, then he gets a utility of 0. Therefore,  $\omega_j(x) = 1\{u_j > 0\}1\{b_j(x) > b_{1-j}(x)\}$ . Each firm  $j$ 's profit from consumer  $x$  is given by equation (1).

$$\Pi_j(x) = [p_j * 1\{u_j > 0\} - b_{1-j}(x)]1\{b_j(x) > b_{1-j}(x)\} \quad (1)$$

---

<sup>4</sup> I ignore the degenerate equilibria, such as when one firm bids zero and the other firm bids high.

<sup>5</sup> In A, I extend this to a market equilibrium by introducing endogenous product prices and many firms.



## 2.2 Firm Bidding Strategy for Individual Consumers

Figure 1 shows firm 0's pricing decision. Firm 0 can potentially sell to consumers located between  $x_0 - \frac{R}{t}$  and  $x_0 + \frac{R}{t}$  from its location of  $x_0$ . Consumer's valuations for firm 0's product is the dotted line  $R - t|x_0 - x|$ . A consumer located at  $x_0$  would be willing to pay up to  $R$  for the product, while consumers located at  $x_0 - \frac{R}{t}$  and  $x_0 + \frac{R}{t}$  would be willing to pay up to 0 for the product.

[Figure 1 about here.<sup>6</sup>]

If firm 0 sets a price of  $p_0$ , then consumers between  $x_0 - \frac{R-p_0}{t}$  and  $x_0 + \frac{R-p_0}{t}$  would be willing to buy its product if they see its ad. Consequentially, firm 0's value for consumers and by the same reasoning firm 1's value for consumers is given by equation (2). Because firms bid for advertising in a second-price auction, one bidding equilibrium is that all firms bid their value for winning each auction. In this paper, I will only consider this case.<sup>7</sup> Therefore, firms bid according to equation (2).

$$b_j(x) = \begin{cases} 0 & \text{if } x < x_j - \frac{R-p_j}{t} \\ p_j & \text{if } x \in [x_j - \frac{R-p_j}{t}, x_j + \frac{R-p_j}{t}] \\ 0 & \text{if } x > x_j + \frac{R-p_j}{t} \end{cases} \quad (2)$$

Figure 2 shows an example of an auction for targeted advertisements between firm 0 and firm 1. Firm 0 bids  $p_0$  for consumers between  $x_0 - \frac{R-p_0}{t}$  and  $x_0 + \frac{R-p_0}{t}$  and 0 otherwise. Additionally firm 1 bids  $p_1$  for consumers between  $x_1 - \frac{R-p_1}{t}$  and  $x_1 + \frac{R-p_1}{t}$  and 0 otherwise.

[Figure 2 about here.]

---

<sup>6</sup>See page 49 for figures.

<sup>7</sup> I ignore degenerate equilibria where one firm bids zero and the other bids more than either would be willing to pay to advertise to a consumer.

Therefore, for consumers between  $x_0 - \frac{R-p_0}{t}$  and  $x_1 - \frac{R-p_1}{t}$ : firm 0 wins the auction and pays firm 1's bid of 0. Firm 0 makes  $p_0$  from each of these consumers. For consumers between  $x_1 - \frac{R-p_1}{t}$  and  $x_0 + \frac{R-p_0}{t}$ : firm 0 wins the auction and pays firm 1's bid of  $p_1$ . Firm 0 makes  $p_0 - p_1$  from each of these consumers. For consumers between  $x_0 + \frac{R-p_0}{t}$  and  $x_1 + \frac{R-p_1}{t}$ : firm 1 wins the auction and pays firm 0's bid of  $p_0$ . Firm 1 makes  $p_1$  from each of these consumers. It follows that firm 0's profit is  $\Pi_0 = p_0[(x_1 - \frac{R-p_1}{t}) - (x_0 - \frac{R-p_0}{t})] + (p_0 - p_1)[(x_0 + \frac{R-p_0}{t}) - (x_1 - \frac{R-p_1}{t})]$ , firm 1's profit is  $\Pi_1 = p_1[(x_1 + \frac{R-p_1}{t}) - (x_0 + \frac{R-p_0}{t})]$ , and the advertising revenue is  $\Pi_A = p_1[(x_0 + \frac{R-p_0}{t}) - (x_1 - \frac{R-p_1}{t})]$ .

Figure 3 demonstrates the difference between niche and general products. Expensive luxury cars like a 2011 Lexus LS 460, where 9,568 units were sold in the US with a manufacturer's suggested retail price of \$73,000, are niche products. In contrast, cheaper mass-produced cars like a 2011 Toyota Corolla, where 240,259 units were sold in the US with a msrp of \$16,230, are general products.<sup>8</sup> Niche products often are more expensive, because of less competition, better ability to price discriminate through selling many niche products instead of one general product, and higher marginal cost of production from not taking as much advantage from an economy of scale.

[Figure 3 about here.]

Yet in a full model, niche-ness is a function of consumer preferences, and product price is a result of niche-ness and other market characteristics; in the real world not all niche products are expensive.

## 2.3 Webpages

In this section, I extend the model presented in subsection 2.1 to include webpages. I do this by splitting the set of tastes or locations  $x$  into 'webpage' intervals. Each interval represents all the consumers visiting one webpage, which is assumed to

---

<sup>8</sup> Figures from Toyota press releases available at <http://pressroom.toyota.com>.

be an exogenous process. Each firm  $j$ 's advertising bidding function  $b_j$  is restricted to be constant for each interval. Therefore, firms can only target consumers based on which webpage they visit.

Figure 4 shows the case of one advertiser. The consumers not visiting the webpage, or equivalently not visiting the webpage interval, cannot buy the product, because they would never be informed about the product through the advertising on the webpage. The advertiser only values consumers who see the ad and would buy the product. He values advertising to those consumers  $p_0$ . He bids a per consumer bid  $b_0$  equivalent to his average profit per consumer who visits the webpage.

[Figure 4 about here.]

Although this is a pay-per-impression (PPI) auction, where the bidding is based on the number of consumers who see the ad, there are a number of other equivalent auctions. This auction is equivalent to each firm  $j$  making one bid  $B_j$  for all advertising on the webpage, where the highest bidder pays the second highest bid. It is also equivalent to a simple pay-per-click (PPC) or pay-per-action (PPA) auction, where firms pay per the number of consumers who buys its product or clicks on its ad.<sup>9</sup> Here, if a mass  $q_j$  of the consumers on the website would buy firm  $j$ 's product if it saw firm  $j$ 's ad, and if firm  $j$  bids  $\tilde{b}_j$ , then the winning bidder would be the firm with the highest  $\tilde{b}_j q_j$  and would pay  $\tilde{b}_{1-j} q_{1-j} / q_j$  per consumer who buys its product or clicks on the ad.<sup>10</sup>

Figure 5, demonstrates how a more narrowly-focused webpage would deliver a smaller segment of consumers to advertisers. A generally-focused webpage, like 'Ford', would deliver a larger, more heterogeneous interval of consumers to advertisers. A more narrowly-focused webpage, like 'Ford Trucks', would deliver

<sup>9</sup> Assuming that these are equivalent and we are not playing a search game.

<sup>10</sup> In a more complicated model of an auction for online advertising this result breaks down (Agarwal et al., 2009, see for example: ).

a smaller, less heterogeneous interval of consumers to advertisers. An even more narrowly-focused webpage, like 'Ford Truck Parts', would deliver an even smaller, even less heterogeneous interval of consumers to advertisers. Note that webpage narrow-ness is not related to product niche-ness. They can be different segments along the same number line.

[Figure 5 about here.]

Consider the case of the three webpages depicted in Figure 6. Webpage (a) is the most generally-focused webpage, webpage (b) is more narrowly-focused than webpage (a), and webpage (c) is the most narrowly-focused. Firm 0 is more niche than firm 1 because it has a smaller interval of potential consumers.

[Figure 6 about here.]

For each firm  $j \in \{0, 1\}$ , firm  $j$ 's bid for advertising to all consumers in webpage interval (a) is  $B_j + B_s$ . Therefore, if  $B_j > B_{1-j}$ , then firm  $j$  will win the auction for advertising to all the consumers in the webpage interval. Consequentially, firm 0 wins the auctions for advertising on webpages (a) and (c), and firm 1 wins the auctions for advertising on webpage (b). General webpages, like webpage (a), deliver a large variety of consumers. Firm 0 wins the auction for webpage (a), because it values a large variety of consumers. As a webpage becomes more niche, like going from webpage (a) to webpage (b), the webpage will attract a more homogeneous set of consumers. A more niche firm will then be able to make more profit from each of these consumers through its higher prices. Firm 1 wins the auction for webpage (b), because it values each of its niche group of consumers more than firm 0. As a webpage becomes even more niche, the chance of it matching with a niche product decreases. Firm 0 wins the auction for webpage (c) because its broadness allowed it to value consumers that firm 1 did not.

### 3 Data & Experimental Design

In this section, I describe the online field experiment that I ran to test whether an ad for a product targeted toward a smaller market segment selling a niche product or an ad for a product targeted toward a larger market segment selling a general product would win the auction for different types of webpages that were broadly-focused and narrowly-focused. I create webpages that each contain: (1) a unique content and (2) a space for one Google text advertisement. For each webpage, Google observes the title and auctions advertising on that webpage based on the perceived topic of the webpage. I collect my dataset by having my Firefox browser randomly cycle through my webpages, and record the ads on each webpage.

#### 3.1 Creation of Webpages

Each of my webpages contains content I chose and one ad determined by the advertisers in an auction. For example, in Figure 7, I posted the content ‘Bentley Convertible.’<sup>11</sup> Google auctioned the ad space in an online auction. The winning ad (i.e. the only ad I observed) was “New 200 Convertible: Build & Price Your New Chrysler® 200 Convertible. Find a Dealer Now!”

[Figure 7 about here.]

I signed up for having online contextual Google advertising for my webpages using Google’s online advertising program: ‘Google Ad Sense’.<sup>12</sup> This allowed me to create frames on my webpages where Google can place targeted ads.<sup>13</sup> By

---

<sup>11</sup> I posted the content of my webpages in two places: (1) in the body of the webpage and (2) in the title of the webpage (the string of text between the Firefox symbol and “Mozil...”). For the purposes of this paper, I always set the content in those two places as equal strings of text, and I refer to that string as the content of the webpage. Future marketing research could set them as unequal strings of text to see the relative weight Google places on the title versus the body of the webpage.

<sup>12</sup> I read the advertising contract that I made with Google to make sure I was not violating it.

<sup>13</sup> The process of observing what content is on a webpage is not instantaneous. Google has to first observe the webpage, through an initial visit to the webpage. It then takes time, usually 15

choosing the size of the frame of the ad, I chose the number of ads that I could see on my webpage. I chose ads of 125 pixels by 125 pixels, which is only big enough for one text ad of no more than 100 characters (including spaces). I also restricted Google to only allow text ads on my webpages.<sup>14</sup>

I posted my webpages under a domain name (or URL) that I acquired through GoDaddy in 2010. I could not completely randomize my choice of my domain name because Google often chooses to target based on the domain name and even a random name might contain a phrase that advertisers target based on. Therefore, I chose a six character domain name that did not receive any search results on Google.<sup>15</sup> I have left the domain name out of the paper, so that I may continue to use the domain name for future experiments.<sup>16</sup>

I set up my webpages so that I would receive as little outside traffic as possible. For instance, I created a false main page that does not link to any of the webpages that I use in my experiment. This way, any surfers, Google employees, or web crawlers would stumble on my false main page when then enter in my domain and not the webpages I use in my experiment.<sup>17</sup> Throughout my experiment, Google never linked to any of my webpages, and I received no clicks on any of my ads.<sup>18</sup>

---

minutes to a few hours, for Google to change the advertising from a general advertisement to a targeted advertisement. People in the computer science field refer to this process as *indexing* the webpage. I allowed Google a full day before gathering any of my data regarding the ads; Google had much more time than it usually takes to index each of the webpages.

<sup>14</sup> I could have allowed picture ads or video ads. I chose not to so that my sample would be only text ads.

<sup>15</sup> In addition, I chose the domain name such that any contiguous string of three or more characters in the name did not receive any search results on Google.

<sup>16</sup> It is available on request.

<sup>17</sup> This false main page was also necessary for my application to have advertising. A Google employee manually looks at the main page of anyone applying to the 'Google Ad Sense' program.

<sup>18</sup> In the three years, I have only received four clicks on any of my ads. Given my hundreds of thousands of observations, I don't think this had any effect on how ads were targeted on my webpages.

### 3.2 Choice of Webpage Content

One of the big advantages of creating my own webpages, is I can control what content is on them. For example, if I wanted to compare the advertising on a 'sports car' webpage with the advertising on a 'red sports car' webpage, I could create two webpages: one with the content 'sports car' and another with the content 'red sports car' instead of having to find two similar webpages or magazines that only vary in the word 'red'. I could even test the ad results on topics not already covered by previously existing webpages. For example, there may not exist a 'red sports car' magazine, but I can create a 'red sports car' webpage.

Artificially constructing webpages in this manner has the disadvantage of not being similar to previously-existing content on the Internet. Because of this disadvantage, I chose to use the text content<sup>19</sup> from previously existing webpages on the Internet as the text content of my webpages.<sup>20</sup>

I chose to use content from websites that were currently being auctioned on Flippa.<sup>21</sup> Flippa (flippa.com) is the largest online marketplace dedicated to buying and selling websites. For instance, nearly 26,000 websites were sold on Flippa in 2011, for a value of almost \$31 million.<sup>22</sup> The advantage of using websites currently auctioned on Flippa is that the seller provides data on the highest bid (i.e. the price for the website), age of the website, ad revenue, website traffic, website ranking, and many more variables. I did not use previously-auctioned websites despite that data being available, because the buyer could have changed the website's characteristics after the auction.

It is reasonable to assume that any information provided about a website on Flippa is accurate because (1) most of these statistics are verified by third party

---

<sup>19</sup>I removed the HTML code and only used the text content.

<sup>20</sup> In order to avoid copyright issues, I randomized the order of the words. Randomizing the order of the words has no affect on the algorithms I use in this paper. It may affect the output of the algorithms used by Google. I leave this question to future marketing research.

<sup>21</sup> These websites were being auctioned as of May 23<sup>rd</sup> 2013.

<sup>22</sup> For more information on Flippa see <https://flippa.com/about>.

sources such as Google and (2) sellers sign a legal contract that their information is accurate. Despite this, there is plenty of missing data, because sellers did not have to share any information on a website. To account for these issues and avoid an issue of selection bias, I only used the final winning bid for the website, which is available for every website. I used it as a proxy for the advertising revenue of the website.

I chose to use the sixty-nine websites that met all of these criteria: (1) the website was being currently auctioned<sup>23</sup> on Flippa on May 23<sup>rd</sup> 2013,<sup>24</sup> (2) the website was identified as an automobile website;<sup>25</sup> (3) the URL was not hidden by the seller;<sup>26</sup> and (4) the website had text content;<sup>27</sup>.

I programed a webcrawler to gather the text off of each of the webpages. For each website, the webcrawler started by downloading the main page where the URL was provided by the Flippa data. Then it gathered each webpage linked to from the main page within the website domain.<sup>28</sup> It next gathered each webpage linked to those pages and iterated, with a maximum of five webpages per website. This was a self-imposed limit because some websites had hundreds of webpages, and some had only a few webpages.<sup>29</sup>

---

<sup>23</sup>through a first-price sealed bid auction with the possibility of a reserve price and the possibility of a buy-it-now price.

<sup>24</sup> 65 of the 1,334 websites that were currently for sale were not being auctioned. Instead, the seller set it as a 'private sale', which means that buyers would make offers and the seller would choose which offer to accept. I did not use these websites because the final price would not be published on Flippa.

<sup>25</sup> Either by self reported categories or a search for the word 'automobile'. This way the websites were all in one industry, the automobile industry.

<sup>26</sup> In Flippa, sellers have the option of hiding their URL and only revealing statistics on the site they are selling. The idea behind this is that some sellers may not want their users or some other party knowing they are selling their website. It would have been impossible for me to use content of their sites because I did not know which site was theirs. Of the 1,334 websites that were currently being auctioned on May 23<sup>rd</sup> 2013, only 14 had their URL hidden.

<sup>27</sup> Some sellers choose to sell domain names without websites that presumably are premium url names.

<sup>28</sup> A website's domain is all the stuff before the '.com', '.net', etc.

<sup>29</sup> I did this to prevent myself from creating too many webpages and biasing my observations too much toward those with many webpages.



For each webpage: I next took the text content off of the webpage;<sup>30</sup> I randomized the order of the words;<sup>31</sup> and I used the text as the content of a new webpage as described in section 3.1.

### 3.3 Data Collection

I then ran an automated program to observe advertisements on my webpages. My data collection program:<sup>32</sup> (1) randomly brings up one of my webpages in the firefox browser; (2) it grabs the webpage content, the text ad, the date, and the time off the webpage;<sup>33</sup> (3) it randomly jumps to another of my webpages in Firefox. Then it iterating back on (2) to grab the information off the new webpage.

Although one iteration can take a fraction of a second, I purposely slowed it down to 20 iterations per minute to keep the traffic on my webpages low. Although technically this experiment does not violate the advertising contract I made with Google,<sup>34</sup> I did not want to gain special attention from Google. Google tracked how much traffic each of my webpages gets per day, including my automated program. If I had received too much traffic, then I would likely have gained the attention of a Google employee who might choose to shut down advertising on my webpages. By keeping my traffic low and avoiding clicking on my ads, I am confident that my experiment escaped notice or at least Google did not choose to do anything that would impact my study.

---

<sup>30</sup>I removed any HTML code through sophisticated use of regular expressions.

<sup>31</sup> I did this to avoid any possible copyright infringement on any webpage.

<sup>32</sup> I thank John, Thomas, and Anna Bruestle for their programming help on writing my data collection program. The program collects data from the firefox browser, not the source code, because Google programmers made it difficult, if not impossible, to get the ad directly from the source code. It was written in Javascript embedded into the webpages in a way such that Google would not be able to see the program.

<sup>33</sup> The program gathers data by printing the output to a text file, because Google programmers have made it impossible to gather data through the source code of my webpages.

<sup>34</sup> I did not click on any of the ads on my own webpages.

### 3.4 Data Description

From May 27<sup>th</sup> to September 21<sup>st</sup> 2013, I collected 893,614 observations like the sample observations I show in Table 8.

[Table 8 about here.]

For every observation, I collected: the time and date, webpage content, ad URL, and ad textual content. Different observations of the same webpage sometime gave me different ads, as shown by my two observations of my “configuration diesels ... ” webpage in my sample data. Therefore, I did not observe a single ad for each webpage, instead, I observed a single ad for each observation. The same ad is often observed on different webpages, as shown by my two observations of an ad for the Trenton used car sale in my sample data.

Each ad contains an average of 13.939 words, for a total of 12,456,083 words. When I removed repeated ads, each ad contains an average of 13.685 words per ad, for a total of 218,555 words. Each of my 138 unique webpage contains 202.203 words for a total of 27,904 words. I often observed the same word repeated in different ads and webpages; I observed only 20,014 unique words.

## 4 Topic Modeling

Intuitively, a Lincoln Blackwood, which is Ford’s 2002-03 all-black luxury pick-up truck which sold only 3,356 units, is a niche product. In contrast, a Toyota Corolla, which is the perennial best selling car in the world, is not. Therefore, an ad for Lincoln Blackwoods would be more niche than an ad for Toyota Corollas. And a webpage on Lincoln Blackwoods would be more narrowly-focused than a webpage on Toyota Corollas. The methodological problem is finding some scientific and systematic way of measuring how niche an ad is and how narrow a webpage is. This is not trivial, because the measure needs to be exogenous. Therefore, I cannot

use number of units sold, price of the product, or price of the ad as proxy variables for niche-ness and narrowness, because these are all endogenous measures.

While an ideal experiment might allow me to pick both the possible ads and the possible webpages, in this experiment, I could not choose the possible ads; instead, I observed real text ads created and bid on by real advertisers. Therefore, I could not slightly vary the text in ads to observe different levels of targeting. I could not add the word 'red' or the word 'truck' to the text of an ad and see how it changed the ad's performance. Instead, I observed the occurrences of different ads. Some of these contain the word 'red', some of these contain the word 'truck', and some of these did not contain either word.

While I could choose to vary my webpages by adding the word 'red' or the word 'truck', there is no way of knowing how that changes the narrowness of a webpage, without some way of measuring the narrowness or the niche-ness of a word derived from how words are used in real ads and webpages. I consequentially chose to use the words from real webpages (randomizing to avoid copyright issues). Because of this, I was also restricted by the words on the webpages I observed.

Furthermore, I could not simply choose a set of keywords in an ad to count with the idea that an ad is more specific if it has more keywords. These measures face, the same problem that was faced in previous studies in economics using keywords in newspapers to identify political bias (see for example: Agirgas, 2011; Gentzkow and Shapiro, 2006; Larcinese et al., 2011). I would be choosing and justifying the set of keywords to count in a non-empirical way. I would be making a judgement call. While I may have been able to justify a set of keywords heuristically, I could not ignore the possibility that there was another set of heuristically justifiable keywords that I had not considered.

In previous versions of this experiment, I tried various heuristic measurements of ad niche-ness and webpage narrowness including: (1) the number of Google search results of a word where fewer results meant a niche word, (2) the number of

nouns in a ad or webpage where more nouns meant the ad or webpage was more specific, and (3) the number of automobile manufacturer and make names where more of these keywords mentioned meant the ad or webpage was more specific. All these test produced weak and inconclusive results.

I construct a measure for ad niche-ness using a *topic modeling* algorithm. I estimate the niche-ness of each word in each ad through examining latent patterns between clusters of words in my ads. Topic modeling algorithms are probabilistic algorithms for under-covering the underling structure of a set of documents using hierarchical Bayesian analysis of the original texts. They are most often used to categorize documents based on observed patterns of words in the texts.

In the past decade, most of the development of topic modeling has been from adaptations and applications of Blei et al.'s (2003a) Latent Dirichlet Allocation (LDA). This includes analysis of scientific abstracts (Griffiths and Steyvers, 2004) and newspaper archives to be generated from multiple topic; in this study, it allowed the topics of documents to be estimated without having to create a new exogenous variable for each document. It estimates the latent random drawn of topics and identifies the underlying latent parameters of its model quickly. Many of the algorithms based on LDA only take a few hours to run on a set of tens of thousands of documents. For a good overview of the important topic models and the importance of LDA, see Blei and Lafferty (2009).

In 4.1, I will first show you the basic topics estimated by the LDA model. Then, I will present the LDA model (4.2), show that the underlying parameters of the model are identified (4.3) by variation in the data, and discuss the algorithms used to estimate the LDA model (4.4). I present the results from running an LDA algorithm on my model first because I seek to illustrate what the model does before we get into the details of how it works.

In 4.5, I present Blei et al.'s (2003b) hierarchical extension to the LDA model, which I use to construct a measure of niche-ness/narrowness. I find strong evidence

for a non-monotonic relationship: The moderately narrow webpages have the less niche ads.

#### 4.1 Latent Dirichlet Allocation Estimation of My Data

Table 9 shows five categories identified from the 15,970 unique ads and 138 unique webpages using Blei et al.'s (2003a) mean field variational inference algorithm for the LDA model.<sup>35</sup> Each column shows the estimated top thirty words for each category.<sup>36</sup>

Based on these estimations of word clusters, the econometrician chooses names for each topic. These names should be interpreted as how the econometrician interprets the data.<sup>37</sup> For example, Topic 1 seems to be composed of many words under the 'Computing & Servers' topic (ex: 'cloud', 'hosting', 'ftp', etc.). Additionally, Topic 2 seems to be composed of many words under the 'Tires & Cars' topic (ex: 'tire', 'michelin', 'wheels', etc.).

[Table 9 about here.]

Figure 10 shows the topics for sample documents from my data. In LDA, each word is a latent draw from a single topic, and different words in the same document may be drawn from different topics. LDA is a *mixture model*. It allows for documents to be drawn from multiple topics. For example, the bottom document in figure 10 is estimated to have been drawn from topics 1 through 4.

---

<sup>35</sup> I describe this algorithm in 4.4.

<sup>36</sup> The most common word in one topic could be common in another topic, so generally the most frequent word in a topic is not considered the top word in a topic. Top words are those with the highest term-score (from: Blei and Lafferty, 2009, see equation (3)), which was inspired by the Term Frequency and Inverse Document Frequency (TFIDF) score of vocabulary terms used in Baeza-Yates and Ribeiro-Neto (1999).

$$term\_score(k, v) = \hat{\beta}_{k,v} \log\left(\frac{\hat{\beta}_{k,v}}{(\prod_{j=1}^K \hat{\beta}_{j,v})^{1/K}}\right) \quad (3)$$

Here,  $\beta_{k,v}$  is the probability of observing vocabulary word  $v$  in topic  $k$  and  $\hat{\beta}_{k,v}$  is its estimate.

<sup>37</sup> Sometimes the names are chosen as the top word for a topic.

[Figure 10 about here.]

## 4.2 Latent Dirichlet Allocation Model

In this section, I present the LDA model from Blei et al. (2003a) to introduce the basic topic model. In 4.5, I will present and use an extension of this model that allows for hierarchical topics to estimate the niche-ness of the words in my ads.

In LDA, I assume that all the words in  $D$  documents, which is the set of all unique webpages and unique ads, drawn from a set of  $V$  vocabulary words through the following latent process:

For each document  $d = 1, \dots, D$ , the number  $N_d$  of words in the document  $d$  are drawn from some random distribution. The assumption of what random distribution is not critical to anything that follows because I am modeling the choice of words, not the choice of the number of words. It does not have to be independent; it can be correlated with the other data generating variables. In the case of my ads, Google limits the number of characters an advertiser uses to one hundred. In general, it is treated as an ancillary variable and is treated as exogenously given.

Next, a  $K$ -dimensional random vector  $\vec{\theta}_d$  is drawn from a Dirichlet Distribution with a  $K$ -dimensional parameter  $\vec{\alpha}$ , where  $K$  is an exogenously given number of topics. Each document is drawn from each topic in different proportions.  $\vec{\theta}_d$  is the vector of these proportions for document  $d$ . The  $k^{\text{th}}$  element  $\theta_{d,k}$  of vector  $\vec{\theta}_d$  will be the probability any given word in document  $d$  is drawn from topic  $k$ . The Dirichlet distribution draws random variables  $\vec{\theta}_d$  on the  $(K - 1)$ -simplex ( $\sum_{k=1}^K \theta_{d,k} = 1$ ) and has a probability density function given by equation (4).<sup>38</sup>

$$p(\vec{\theta}_d | \vec{\alpha}) = \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_{d,k}^{(\alpha_k - 1)} \quad (4)$$

---

<sup>38</sup> It is standard practice to restrict all the  $\alpha$ s to be equivalent, which is equivalent to restricting all topics to be about the same size. This makes parameter estimation much quicker. I do not do this here because the HLDA model extends the more general Dirichlet distribution.

Then, for each word  $n = 1, \dots, N_d$  in document  $d$ , a random topic  $z_n \in 1, \dots, K$  is drawn from the multinomial distribution with a parameter of  $\vec{\theta}_d$ . The probability of choosing topic  $k$  is  $\theta_{d,k}$ . The draws of topics are technically not independent because they depend on the document specific parameter  $\vec{\theta}_d$ . They are essentially *conditionally independent and identically distributed*, where the conditioning refers to the document specific parameter  $\vec{\theta}_d$ . Therefore, in LDA, it is assumed that the order of the words in a document does not matter.<sup>39</sup>

Then for each word  $n = 1, \dots, N_d$  in document  $d$ , a random word  $w_n \in 1, \dots, V$  is drawn from the multinomial distribution with a parameter  $\beta_{z_n}$  conditioning on the topic  $z_n$  of word  $n$ . The probability of choosing vocabulary word  $v$  is the parameter  $\beta_{z_n,v}$ .

An example of this process is shown in Figure 11. There are two bags of words. Each bag of words is a different topic ( $K = 2$ ). Topic 1 is ‘Trucks’. A random word from a document about ‘Trucks’ has a 25% chance of being the word ‘pickup’, a 20% chance of being the word ‘truck’, and so forth. Topic 2 is ‘SUVs’. A random word from a document about ‘SUVs’ has a 25% chance of being the word ‘suv’, a 20% chance of being the word ‘toyota’, and so forth.

[Figure 11 about here.]

For each document  $d = 1, 2$ , and 3, the number of words is drawn from some random process. There are  $N_1 = 11$  words in document 1,  $N_2 = 8$  words in document 2, and  $N_3 = 15$  words in document 3. Then, for each document  $d$ , a vector  $\vec{\theta}$  is drawn, representing the probability from drawing from each bag of words. For words in document 1, the probability from drawing words from the first bag (topic 1) is 100%, so  $\theta_1 = 100\%, \theta_2 = 0\%$ . For words in document 2, the probability from drawing words from the first bag (topic 1) is 50% and from the second bag (topic 2) is 50%, so  $\theta_1 = 50\%, \theta_2 = 50\%$ . Likewise,  $\theta_1 = 0\%, \theta_2 = 100\%$ .

---

<sup>39</sup> This assumption is later relaxed in dynamic topic models (DTM) (Blei and Lafferty, 2006, see: ).

Words are then selected for each document  $d$  in the following way: For each word  $n = 1, \dots, N_d$ , a random bag of words (or topic) is drawn. The probability of drawing the bag of words  $k$  (topic  $k$ ) is  $\theta_{d,k}$ . Therefore all the words in document 1 are drawn from the first bag of words. All the words in document 3 are drawn from the second bag of words. Each word in document 2 then has a 50% chance of being drawn from the first bag of words and a 50% chance of being drawn from the second bag of words.

### 4.3 Identification

The advantage of the LDA model over its predecessor Probabilistic Latent Semantic Indexing (pLSI) model (Hofmann, 1999) is that the LDA can easily be identified. In this section, I will explain why we have enough observations to identify the parameters in the LDA model, and I will explain why we observe enough variation in the data to confidently identify LDA. I do this by first discussing identification in the pLSI model, which I use to explain the identification in the LDA model. This same discussion applies to the HLDA model, which is a restricted version of the LDA model.

The pLSI model (Hofmann, 1999) is the first mixture model in topic modeling; it is the first topic sorting model that allows for documents to be drawn from multiple topics. The basic probability of a realization of a document given by the pLSI model is shown in equation (5).

$$p(\vec{w}|d, \beta) = \prod_{n=1}^{N_d} \left( \sum_{z_n} \beta_{z_n, w_n} \theta_{d, z_n} \right) \quad (5)$$

$$\beta_{k,v} = p(w = v | z = k)$$

$$\theta_{d,k} = p(z = k | d)$$

Here, each document  $d$  is a realization of a  $N_d$ -vector of words  $\vec{w}$ . Each word is



selected by first drawing a random topic from the document specific distribution of topics where  $\theta_{d,k}$  is the document specific probability of drawing topic  $k$ . Then a vocabulary word is drawn from the topic specific distribution of words where  $\beta_{k,v}$  is the topic  $k$  specific probability of drawing vocabulary word  $v$ . Therefore the pLSI model has  $K(D + V)$  underlying parameters:  $K$  parameters for each document (each  $\theta_{d,k}$ ) and  $K$  parameters for each word (each  $\beta_{k,v}$ ).

The LDA model (Blei et al., 2003a) simplifies the pLSI model to allow for relatively fewer parameters to identify. The probability of a realization of a document given by the LDA model is shown in equation (6).

$$p(\vec{w}|\vec{\alpha}, \beta) = \int p(\vec{\theta}_d|\vec{\alpha}) \prod_{n=1}^{N_d} \left( \sum_{z_n} \beta_{z_n, w_n} \theta_{d, z_n} \right) d\theta \quad (6)$$

Each probability  $\theta_{d,k}$  of a topic  $k$  in document  $d$  is now a random realization from the same Dirichlet distribution. Therefore the basic LDA model has  $K(1 + V)$  underlying parameters: the  $K$ -parameters in the vector  $\vec{\alpha}$  that control the Dirichlet distribution and  $K$  parameters for each word. Adding another document to the dataset adds to the number of underlying parameters in the pLSI model by  $K$ , but does not add to the number of underlying parameters in the LDA model unless a new word is drawn.

Increasing the number of documents collected increases the number of vocabulary words. An additional document should be expected to add new words to the vocabulary list at a decreasing rate. This poses two identification problems: (1) the maximum likelihood estimates of the basic model would say that this is a zero probability event, and (2) there will always be a set of rare words with few observations.

Blei et al. (2003a) solved both of these problems by applying a form of smoothing

to the  $K \times V$  parameters  $\beta$ .<sup>40</sup> They treat each vector  $\vec{\beta}_k$  (the set of probabilities of seeing each word in each topic  $k$ ) as a random draw from a symmetric Dirichlet distribution with a latent parameter  $\eta$ .<sup>41</sup> This produces a random vector on the  $V - 1$ -simplex. If a new word is added to the vocabulary set, then this becomes a random draw on the  $V$ -simplex. Therefore, every word observed has a positive probability of being drawn in each topic (solving problem 2), and there is a positive probability of drawing a new word when extrapolating to new documents (solving problem 1). Consequentially, the LDA model has  $K + 1$  parameters that need to be identified:  $\vec{\alpha}$  and  $\eta$ . Everything else ( $\theta, \beta, F, \dots$ ) are all random variables that we can estimate given identifying  $\vec{\alpha}$  and  $\eta$ .

If  $\alpha_k$  is relatively smaller than the other  $\alpha$ s, then Topic  $k$  occurs relatively less frequently than the other topics. If  $\alpha_k$  is relatively larger than the other  $\alpha$ s, then Topic  $k$  occurs relatively more frequently than the other topics. Consequentially, the relative size of the  $\alpha$ s is identified by the frequency of occurrences of documents with that topic, or in other words, by how common the cluster of words appears in my data.

Figure 12 shows how  $\bar{\alpha} = \sum_{k=1}^K \alpha_k$  is identified.  $\bar{\alpha}$  controls the mixtures of topics in documents. A large  $\bar{\alpha}$  occurs when there is a large overlap in the clusters of words used for topics. A small  $\bar{\alpha}$  occurs when there is a small overlap in the clusters of words used for topics. A large  $\bar{\alpha}$  means the topics are very similar and the words in each document tends to be drawn from a more even spread of topics. A small  $\bar{\alpha}$  means the topics are very disjointed and documents are usually drawn from only one topic.

[Figure 12 about here.]

$\eta$  is identified by looking at the distribution of words in the topics. If the words in each topic are drawn mostly from an even spread of probabilities across a large

---

<sup>40</sup> Unfortunately, simple Laplace smoothing is no longer justified as a maximum a posteriori method for this type of problem.

<sup>41</sup>  $\vec{\eta} = \langle \eta, \eta, \eta, \eta, \dots \rangle$ .

cluster of words, then  $\eta$  is relatively large. If the words in each topic are drawn mostly from a small cluster of a few words, then  $\eta$  is relatively small.

#### 4.4 Overview of Estimation Techniques

The key to identifying the LDA model is through the likelihood function given by equation (6). Unfortunately, equation (6) is too intractable to estimate reliably using maximum likelihood because of the coupling between  $\theta$  and  $\beta$  (see: Dickey, 1983). Several approximation techniques have been developed. The most commonly accepted are: mean field variational inference (Blei et al., 2003a), Gibbs sampling (Griffiths and Steyvers, 2004), expected propagation (Minka and Lafferty, 2002), and collapsed variational inference (Teh et al., 2006a). In this section, I will describe the two most common techniques to illustrate the general idea of how they work. I refer to the authors of these techniques for the details.

Blei et al. (2003a) developed the first algorithm as a mean field variational inference technique, which is what I used to identify the topics used in Tables 9 and 10. This basic idea of this technique is to: (1) use Jensen's inequality to find an adjustable lower bound on the log of the likelihood given in equation (6). The estimates for the document and word-specific parameters ( $\vec{z}$  and  $\vec{\theta}$ ) are then chosen that produce the tightest possible lower bound. Then (2) use these estimates to find the best estimates for the document generating parameters ( $\beta$ ,  $\vec{\alpha}$  and  $\eta$ ), then iterate back on (1). The error introduced by Jensen's inequality will converge to zero as the parameter estimates converge.

The most commonly accepted algorithm (mostly from its speed) is Griffiths and Steyvers's (2004) Gibbs sampling technique. I will use an adaptation of this technique to identify the HLDA extension to LDA in 4.5. Gibbs sampling is a form of Markov Chain Monte Carlo (MCMC). The basic idea of this technique is that all of the parameters can be estimated from the realizations of all the words topic assignments  $z$ . Therefore, posterior estimates of the set of probabilities that a word

is in each topic can be calculated from the topic assignments of the other words. The algorithm (1) assigns each word an arbitrary topic,<sup>42</sup> (2) calculates the set of probabilities that a word is in each topic from the current topic assignments of the other words, and (3) randomly draws a topic for each word from its distribution. It then iterates back on (2).

## 4.5 Hierarchical Latent Dirichlet Allocation

In this section, I restrict the LDA model to have a hierarchical topic structure (Blei et al., 2003b). I begin by defining the hierarchical structure of topics through the *nested Chinese restaurant process*. Then, I explain how this modifies the LDA model. In the next section, I will use these results to analyze the effect of webpage narrowness on ad niche-ness.

Imagine a process where  $M$  customers enter a Chinese restaurant. The first customer sits at the first table. Each additional customer  $m$  sits at a random table from the probabilities given in equation (7).

$$\begin{aligned} p(\text{occupied table } k | \text{previous customers}) &= \frac{m_k}{\gamma + m - 1} \\ p(\text{start a new table} | \text{previous customers}) &= \frac{\gamma}{\gamma + m - 1} \end{aligned} \tag{7}$$

Here,  $m_k$  is the number of previous customers sitting at table  $k$ , and  $\gamma$  is a parameter. Customers are more likely to sit at a table with more people because people are drawn to others. As more customers enter the restaurant, it is less likely that a customer will sit at a new table.

On each table, there is a flier with instructions to a new and different Chinese restaurant. All the customers at each table read their flier and decide to go to the new restaurant the next day. All the customers at the table start the process over again at the new restaurant the next day. In this way, each customer to draws a random path down a hierarchical tree of chinese restaurants.

This process is known as the *nested Chinese restaurant process*. In this same way,

---

<sup>42</sup> This could be a guess or a random assignment.

each document  $d$  is assumed to have drawn a random path down a hierarchical tree of topics. Once the path down the tree of topics is selected, then the words in the document are generated by an LDA process among the topics in its random path, as shown in figure 13.

[Figure 13 about here.]

This introduces two new parameters into the model:  $\gamma$  (which controls the probability of sitting at a new table) and the number of levels of the hierarchical tree. In general, HLDA algorithms do not estimate these parameters. These parameters control how many topics observed in the data. Optimizing the number of topics, usually produces many small topics (Blei et al., 2003a). To make the results meaningful, I limit the number of topics by setting  $\gamma = .25$  and the number of levels to 3 (levels 0, 1, and 2). My results are robust for different  $\gamma$ s around  $\gamma = .25$ , and  $\gamma = .25$  is consistent with the machine learning literature (see 5.1 for details). I was unable to estimate more than three levels because of the small size of the documents. If the documents were longer, then I could estimate more levels.<sup>43</sup>

Because the number of topics is no longer fixed, the HLDA model treats the  $\alpha$  (the parameters of the Dirichlet algorithm for drawing topic probabilities) as random draws from a GEM Distribution. This introduces two new variables: the mean and the scale of the GEM distribution. The GEM mean is the proportion of general words relative to niche words. The GEM scale controls how strictly documents should follow the general versus specific word proportions. I chose a GEM mean of .2 because it produced relatively few small topics (topics with only a few words) (see 5.1 for details). Additionally, I chose a GEM scale of 100% because I wanted to preserve a good balance of niche and general words. My results are robust for my choice of these parameters.<sup>44</sup>

---

<sup>43</sup> The content on the webpages were long enough to estimate more levels. However, there were too few webpages for this to be meaningful.

<sup>44</sup> I got similar results when I varied these parameters.

This leaves two sets of variables for the algorithm to identify:  $\vec{\eta}$  and  $\alpha$ . As I mentioned in 4.3, Blei et al. (2003a) treats each vector  $\vec{\beta}_k$  (the set of probabilities of seeing each word in each topic  $k$ ) as a random draw from a symmetric Dirichlet distribution with a latent parameter  $\eta$ . In LDA,  $\eta$  is the same for all topics. In HLDA, we can relax this and let  $\eta$  be the same for all topics in the same level. Therefore, I have three parameters to identify  $\eta_0$  (the  $\eta$  for the level 0 topic),  $\eta_1$  (the  $\eta$  for the level 1 topics), and  $\eta_2$  (the  $\eta$  for the level 2 topics), plus an additional parameter for each topic.

In running this experiment, I estimated the levels of the words by taking the mode of the estimated level from iterations 1,000, 2,000, 3,000, ..., and 10,000 of Blei et al. (2003b)'s Gibbs sampling algorithm for HLDA. The language modeling literature does it this way to reduce the error in estimating the levels of each word (see Blei et al. (2003a) for details).

## 5 Results

This section reports the results from my online field experiment. I start by explaining the process that I used to select reasonable input parameters for Blei et al.'s (2003b) Gibbs sampling algorithm for HLDA on my data set of unique webpages and ads (see 5.1). I next show the algorithm converges and the parameters are identified (see 5.2). Further, I analyze the distribution of topic levels of words in my webpages and ads (see 5.3), which I use to measure the niche-ness of words in my ads and narrowness in my webpages. I then examine the top words from the topics, or equivalently the word clusters, estimated by the algorithm. I demonstrate that the algorithm estimates which words are niche and which words are general (see 5.4). I continue by regressing my measure of webpage narrowness on my measure of ad niche-ness and show that there is a non-monotonic relationship (see 5.5). I conclude this section by analyzing the residuals to make sure the estimates of my regressions are reasonable (see 5.6).

## 5.1 Input Parameter Determination

In this section I describe how I chose two of my variables: (1)  $\gamma = .25$ , which controls the probability of a document forming a new topic (see equation (7)), and (2) a GEM mean = .2, which controls the proportion of general words relative to niche words. The other input variables were chosen to be consistent with the literature.

Merely finding the optimum<sup>45</sup> number, size, and shape of the word clusters does not provide as meaningful a result. In Blei et al.'s (2003b) Gibbs sampling HLDA algorithm, there are several parameters the econometrician chooses to control the number, size, and shape of the word clusters that the algorithm identifies (see 4.5). It is likely that optimizing the number of topics, for example, would define each document as its own word cluster.

I find  $\gamma$  and the GEM mean by looking at the structure of the topic trees formed under different values of  $\gamma$  and the GEM mean. I then chose the topic tree that satisfies these three conditions: (1) it minimizes the number of small topics, (2) it minimizes the number of topics, and (3) it maximizes the number of level 1 and level 2 words. These conditions mean that my model fits the data better and therefore can produce more meaningful results.

For example, Table 14 illustrates how changing the  $\gamma$ , which controls the probability of a document forming a new topic (see equation (7)), affects the number, size, and shape of the topics estimated by Blei et al.'s (2003b) Gibbs sampling HLDA algorithm. Each column shows the topics estimated by the algorithm at the 10,000<sup>th</sup> and final iteration of the algorithm for three different  $\gamma$ s: .2, .25, and .3.

[Table 14 about here.]

For example, for  $\gamma = .25$ , there is one level 0 topic, which is the general topic that is common to all 13,467 unique documents. It is estimated 122,177 words from these unique documents were drawn from this topic. There are ten level 1 topics,

---

<sup>45</sup>Through finding the parameters that maximize likelihood.

which are the topics more niche than the level 0 topic. These topics are estimated to have generated more than 41,065 of the words in my documents. And there are forty-four level 2 topics, which are the most niche topics; thirty-four of which generated an estimated twenty words or fewer.

A  $\gamma$  of .25 seems to fit better and have more meaningful results than a  $\gamma$  of .2 and .3, because (1) I get fewer small topics, (2) I get fewer topics, and (3) I get more level 1 and level 2 words.

Table 15 shows the number of topics in each level with a low word count by  $\gamma$  for  $\gamma = .01, .05, .1, .15, .2, .25, .3, .35$ , and .4.

[Table 15 about here.]

Table 15 illustrates how  $\gamma = .25$  produces fewer small topics:  $\gamma = .25$  produces only one level one topic with fewer than fifty words and it produces fewer small level two topics than  $\gamma = .2$  and  $\gamma = .3$ . Therefore, I feel justified in my choice of a  $\gamma = .25$ .

## 5.2 Convergence of Algorithm

In this section, I test the convergence of the Blei et al.'s (2003b) Gibbs sampling HLDA algorithm on my data. In language modeling, it is conventional to use perplexity instead of likelihood (see: Blei et al., 2003a; Rosen-Zvi et al., 2004). Perplexity is the predicted probability of being able to predict words in new unseen documents. Perplexity is monotonically decreasing in the likelihood and is equivalent to the inverse of the geometric mean of the per-word likelihood. A lower perplexity indicates a higher likelihood, or equivalently better performance.

Figure 16 shows the perplexity for all 10,000 iterations of Blei et al.'s (2003b) Gibbs sampling HLDA algorithm conditional on the given parameters ( $\gamma$  and GEM mean) and the estimated parameters ( $\vec{\eta}$ , and  $\vec{\alpha}$ ) being true. After the first thousand iterations, the perplexity has an average of about 2.176 million with a small standard deviation of 0.00758 million. Given that this is a Markov Chain Monte Carlo



algorithm, and there would naturally be some variation in the estimation of the perplexity, this shows significant evidence that the model was identified.

[Figure 16 about here.]

Figure 17 (b) shows the perplexity conditional on only the  $\vec{\eta}$  parameters being true. As I mentioned in 4.5,  $\eta_l$  is the symmetric Dirichlet parameter that determines the random draw of a probability  $\beta_{k,v}$  of seeing a word  $v$  in a topic  $k$  from level  $l$ . This perplexity tends to be lower (indicating a higher likelihood) than the perplexity in Figure 17, because it is less constrained. After the first thousand iterations, this perplexity has an average of about 2.0444 million with a small standard deviation of 0.00796 million.

[Figure 17 about here.]

I could not do the same for the other identified parameters  $\vec{\alpha}$ , because the number of topics changes from iteration to iteration. Instead, I relied on the total perplexity discussed above and shown in Figure 16 to show that its estimates converge.

### 5.3 Distribution of Results

In this section, I examine the distribution of my HLDA measure. As I mention earlier, the levels of words were calculated by taking the mode of the estimated level from iterations 1,000, 2,000, 3,000, ..., and 10,000 of Blei et al.'s (2003b) Gibbs sampling algorithm for HLDA. I then take an average of these levels for each document as my measure of ad niche-ness and as my measure of webpage narrowness.

Table 18 shows the sample frequency of the level of each word (level 0,1, or 2), first unconditionally and then conditionally on the level of another given word in the same ad or the same webpage.

[Table 18 about here.]

Table 18 first shows the sample frequencies across all 893,614 ads (Table 18a). Next, it shows the sample frequencies across all 13,467 unique ads (Table 18b), and it next shows the sample frequencies across all 138 unique webpages (Table 18c). The sample frequency  $f(i|j)$  of the level of a word being  $i$  conditional on another given word in the same document (ad or webpage) being level  $j$  was calculated by

$$f(i|j) = \frac{\sum_d \left( c_j^d * \frac{c_i^d - 1_{\{i=j\}}}{c_0^d + c_1^d + c_2^d - 1} \right)}{\sum_d c_j^d / d} \quad (8)$$

where  $c_i^j$  is the count of the number of words in document  $d$  that are of level  $i$ .

Across all observations (Table 18a), I estimate that 88.60% of the words in an ad were generated from the general level 0 topic, 10.13% from a level 1 topic, and 1.27% from a level 2 topic. If another given word is level 0, then I estimate that 89.82% of the words in an ad were generated from the general level 0 topic, 9.06% from a level 1 topic, and 1.15% from a level 2 topic. If another given word is level 1, then I estimate 83.23% of the words in an ad were generated from the general level 0 topic, 15.07% from a level 1 topic, and 1.71% from a level 2 topic. If another given word is level 2, then I estimate 83.28% of the words in an ad were generated from the general level 0 topic, 13.50% from a level 1 topic, and 3.26% from a level 2 topic.

Across unique ads (Table 18b), I estimate that 86.05% of the words in an ad were generated from the general level 0 topic, 12.11% from a level 1 topic, and 1.85% from a level 2 topic. If another given word is level 0, then I estimate that 87.89% of the words in an ad were generated from the general level 0 topic, 10.54% from a level 1 topic, and 1.59% from a level 2 topic. If another given word is level 1, then I estimate 78.99% of the words in an ad were generated from the general level 0 topic, 18.41% from a level 1 topic, and 2.61% from a level 2 topic. If another given word is level 2, then I estimate 78.39% of the words in an ad were generated from the general level 0 topic, 17.25% from a level 1 topic, and 4.38% from a level 2 topic.

Across unique webpages (Table 18c), I estimate that 77.56% of the words in a webpage were generated from the general level 0 topic, 20.19% from a level 1 topic, and 2.25% from a level 2 topic. If another given word is level 0, then I estimate that 82.85% of the words in a webpage were generated from the general level 0 topic, 15.77% from a level 1 topic, and 1.38% from a level 2 topic. If another given word is level 1, then I estimate 56.75% of the words in a webpage were generated from the general level 0 topic, 41.31% from a level 1 topic, and 1.94% from a level 2 topic. If another given word is level 2, then I estimate 69.27% of the words in a webpage were generated from the general level 0 topic, 27.10% from a level 1 topic, and 3.62% from a level 2 topic.

If the conditional probabilities were  $f(i|i) = 1$  and  $f(i|\text{not } i) = 0$ , then the average of the estimated levels of each word in a document or a webpage would either be 0, 1, or 2. Yet I observe a lot of mixing in the word level in my documents and webpages, so I observe many different average word levels between 0 and 2. If the conditional probabilities were such that  $f(0|3) = f(3|0) = 0$ , then the average estimated word level and number of words in a document would be sufficient to know how many words of each level are estimated to be in a document. In the case of my data, I am not losing much information about an ad or a webpage by looking at the average estimated word level, because I have (1)  $f(i|i) > f(i|j \neq i)$  and (2)  $f(1|2) > f(1|0)$  in all cases.

Figure 19 shows the cumulative distribution function of my HLDA measure of ad niche-ness for all 893,614 observations (the thick red line), for the 15,970 unique ads (the thinner gray line), and for the 77,507 unique ad-webpage combinations (the dashed blue line). In addition, it shows the cumulative distribution function of my HLDA measure of webpage narrowness for the 138 unique webpages (the dotted green line). The CDF of my HLDA measure for all observations is weakly above the CDF of my HLDA measure for unique ads because the most common ads tend to be less niche. Because of this the CDF of my measure for unique

ad-webpage combinations is sandwiched between the CDFs of my measure for all observations and unique ads (although very close to the CDF of my unique ads). In addition, the CDF of my HLDA measure for all webpages is weakly below the CDFs of my HLDA measure for my ads because webpages tend to be more narrow than ads are niche; webpages have more words, so they have a greater ability to become narrow than ads have at becoming niche.

[Figure 19 about here.]

In general, each of these CDFs looks like a smooth distribution that is cut off at zero.<sup>46</sup> For example, 27.08% of my unique ads and 32.52% of my observations have a measure of niche-ness of zero. This comes from all words in the ad or a webpage being estimated as being generated from the general level zero topic. Because the cut-off point at a HLDA measure of zero matters (more so for the ads than the webpages), I chose to analyze these data using a Tobit regression, which allows for a minimum and maximum observed dependant variable.

## 5.4 HLDA Estimates Niche-ness of Words

The purpose of this section is to determine if the HLDA algorithm gives good estimates for which words are general and which words are niche/narrow. I do this by looking at the top words from the common topics. As I mentioned in footnote 36, the most common word in one topic could be common in another topic, so generally the most frequent word in a topic is not considered the top word in a topic. Top words are those with the highest term-score (from: Blei and Lafferty, 2009, see equation (3)), which was inspired by the Term Frequency and Inverse Document Frequency (TFIDF) score of vocabulary terms used in Baeza-Yates and Ribeiro-Neto (1999). I only examine the most common topics because their large number of observations makes them estimated more accurately.

---

<sup>46</sup> It is not perfectly smooth, because the number of words in an ad was limited by the 100 character max.

Figure 20 shows the top words from Topics Identified by Blei et al.'s (2003b) Gibbs sampling algorithm for HLDA. Each blue box represents a different topic identified by HLDA. It contains the words with the highest term-score for that topic. The algorithm identified fifty-five topics; Figure 20 shows the eight of the most common topics.

[Figure 20 about here.]

In level 0, there is one topic, which are those that are the common words. The top eight words from this topic are 'free', '&', 'the', 'your', 'to', 'for', 'a', and 'and'. These words are what I would expect to be common among automobile ads and webpages. They are also the words that I would describe as the most general, because they do little to differentiate automobile ads and webpages.

In level 1, there are ten topics, which are those that are the less common words than in level 0. The three topics shown are the only topics estimated to cover 2,862 or more words in my data (unique ads and webpages). These words would differentiate the ad more than those in level 0, so I would describe them as making the ad more niche. The words in these topics are more differentiated than those in level 0. For example, there are words like 'freight', 'scan', 'manuals', 'wheels', 'pm', 'may', and 'detroit'.

In level 2, there are forty-four topics, which are those that are the least common words. Figure 20 shows the four most common of these topics, which are the only topics that satisfy both: (1) have a parent topic of the three most common level 1 topics and (2) cover 170 or more words in my data. These words are the least common; they are more niche/narrow than level 0 and level 1 words. For example, there are words like 'influential', 'flowers', 'one-view', 'protocols', 'prohibited', 'frosty', 'jimmy', and 'e-bike'.

## 5.5 Webpage Narrowness vs. Ad Niche-ness

In this section, I compare webpage narrowness to ad niche-ness. I construct a measure of ad niche-ness and webpage narrowness by taking the average of the topic levels of the words that I estimated using HLDA.

Table 21 (b) shows my regressions of webpage narrowness, which is the average level of the topics of the words in an webpage, on ad niche-ness, which is the average level of the topics of the words in an ad.

[Table 21 about here.]

In Table 21 (b), I first ran the simple Tobit regression

$$y = \beta_0 + \beta_1 x + \epsilon \quad (\text{R1 and R1a})$$

$$\text{where } y \equiv \frac{c_1^a + 2 * c_2^a}{c_0^a + c_1^a + c_2^a} \in \{0, 2\}$$

$$x \equiv \frac{c_1^w + 2 * c_2^w}{c_0^w + c_1^w + c_2^w}$$

where the dependant variable  $x$  is the number of words in the webpage title,  $\beta_0$  and  $\beta_1$  are parameters to be estimated,  $\epsilon$  is the error term where the conditional mean of  $\epsilon$  given  $x$  is zero, and  $c_l^d$  is the count of the number of non-repeated words in the document( $a$  for ad and  $w$  for webpage) that are estimated to be level  $l = 0, 1, 2$ . This would make my dependant variable  $y$  into the average of the estimated word level in the ad (my measure of ad niche-ness) and my independent variable  $x$  into the average of the estimated word level in the webpage (my measure of webpage narrowness). Note that  $y$  is bounded between 0 and 2 (thus the Tobit).<sup>47</sup>

Running this regression (R1) on all observations, I find that webpage narrowness increases my measure  $y$  of ad niche-ness by 0.012. This result is significant at the

---

<sup>47</sup>  $x$  is also bounded, but this does not affect the regression.

.1% level. Given that each ad contains an average of 13.939 words, if all the words in a webpage were to increase by one level of my measure, then one word in about six ads on the webpage would increase by one level. Changing one word in an ad can profoundly change the meaning of the ad. Consider changing the ad “Get your car price now!” (all top level 0 words) to the ad “Get your studebaker price now!” The second ad would be much more niche.

Because I repeatedly observe the same ad on the same webpage at different times, my results may be biased toward more frequent, less niche ads. Therefore, I also ran the regression (R1a) on unique observed ad-webpage combinations and removed any repeated observations of the same ad-webpage combination. I found that increasing the number of words in the webpage title decreases my measure  $y$  of ad niche-ness by 0.007. This result is significant at the 5% level. This would mean that if all the words in my webpage were to increase by one level of my measure, then one word in about ten ads on the webpage would decrease by one level. This significant difference in direction indicates that the relationship between webpage narrowness and ad niche-ness is not linear.

A significant amount of variation in my observations may depend on the daypart, as has been shown in other studies. For instance, Lowy (2003) showed significant differences in Internet audiences in gender, age, income level, size of audience, work / home use, and type of Internet use between five different dayparts: the early morning (Monday-Friday, 6am - 8am), the daytime (Monday-Friday, 8am - 5pm), the evening (Monday-Friday, 5pm - 11pm), the late night (Monday-Friday, 11pm - 6am), and the weekend (Saturday-Sunday). Therefore, I needed to check whether variations in my observations depended on the daypart as well. To test for this, I ran the regression

$$y = \beta_0 + \beta_1 x + \sum_{i=1}^4 \beta_{\text{daypart}_i} t_{\text{daypart}_i} + \epsilon \quad (\text{R2})$$

where  $t_{\text{daypart}_1}$  is one if the observation was during the early morning,  $t_{\text{daypart}_2}$  is one if the observation was during the evening,  $t_{\text{daypart}_3}$  is one if the observation was during the late night, and  $t_{\text{daypart}_4}$  is one if the observation was during the weekend. If all  $t_{\text{daypart}_1} = t_{\text{daypart}_2} = t_{\text{daypart}_3} = t_{\text{daypart}_4} = 0$ , then the observation was during the daytime. Therefore, a dummy variable for daytime was left out of the regression to avoid multicollinearity. In addition  $\beta_{\text{daypart}_1}, \dots, \beta_{\text{daypart}_4}$  are now additional parameters to be estimated, and the error term  $\epsilon$  has now a conditional mean of zero given  $x, \beta_{\text{daypart}_1}, \beta_{\text{daypart}_2}, \beta_{\text{daypart}_3}$ , and  $\beta_{\text{daypart}_4}$ .

Running this regression (R2) on all observations, I find that observing my webpage during the early morning (compared to during the daytime) increases my measure  $y$  of ad niche-ness by 0.004. For all other times, ad niche-ness decreased: during the evening ad niche-ness decreased by 0.012, during the late night ad niche-ness decreased by 0.004, and during the weekend ad niche-ness decreased by 0.011. These results show significant differences in the ad niche-ness depending on daypart, with the exceptions of early morning not being significantly different than the daytime and evening not being significantly different than the weekend.

I also find that adding these daypart dummy variables does not change my estimates for the coefficient  $\beta_1$ , the relationship between ad niche-ness and webpage narrowness. I designed my experiment so that my data collecting program randomly observed my webpages across time. Therefore, daypart and webpage content are independent of each other.

I cannot run this regression on unique observed ad-webpage combinations because I observe the same ad-webpage combination at different times. Therefore I ran the regression

$$y = \beta_0 + \beta_1 x + \sum_{i=1}^4 \beta_{\text{daypart}_i} \bar{t}_{\text{daypart}_i} + \epsilon \quad (\text{R2a})$$

where  $\bar{t}_{\text{daypart}_i}$  is the average  $t_{\text{daypart}_i}$  for that particular ad-webpage combination.



This should be seen as how frequently I see an ad-webpage combination at different times.

Running this regression (R2a) on unique observed ad-webpage combinations, I find that observing my webpage during the early morning (compared to the daytime) increases my measure  $y$  of ad niche-ness by 0.014; during the evening ad niche-ness increases by 0.013; during the late night ad niche-ness increases by 0.001; and during the weekend ad niche-ness decreases by 0.009. These results show marginal differences in the ad niche-ness depending on daypart; namely, the evening and the weekend are significantly different from daytime at the 5% level.

Because my theory and my mixed linear results suggest that the relationship between webpage narrowness and ad niche-ness may not be monotonic, I also ran the following quadratic regressions

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon \quad (\text{R3 and R3a})$$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \sum_{i=1}^4 \beta_{\text{daypart}_i} t_{\text{daypart}_i} + \epsilon \quad (\text{R4})$$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \sum_{i=1}^4 \beta_{\text{daypart}_i} \bar{t}_{\text{daypart}_i} + \epsilon \quad (\text{R4a})$$

Running regressions (R3) and (R4) on all observations, I find a negative coefficient of  $-0.119$  on the linear term and a positive coefficient of  $0.103$  on the quadratic term. This creates a U-shaped relationship between webpage narrowness and ad niche-ness: When webpage narrowness is below  $.578$ , increasing webpage narrowness decreases ad niche-ness, and when webpage narrowness is above  $.578$ , increasing webpage narrowness increases ad niche-ness. In other words: The least niche ads tend to appear more on the moderately-narrow webpages. These results were all significant at the  $.1\%$  level for both regressions. In addition, this non-monotonicity did not change my estimates of the coefficients on the effects of

dayparts in (R4) from the results I found in (R2).

Running regressions (R3a) and (R4a) on all unique observed ad-webpage combinations, I find a similar relationship; I find a negative coefficient of about  $-0.042$  on the linear term and a positive coefficient of  $0.027$  on the quadratic term. This creates a U-shaped relationship between webpage narrowness and ad niche-ness: When webpage narrowness is below  $.778$ , increasing webpage narrowness decreases ad niche-ness, and when webpage narrowness is above  $.778$ , increasing webpage narrowness increases ad niche-ness. In other words: The least niche ads tend to appear more on the moderately-narrow webpages. These results were also all significant at the  $.1\%$  level for both regressions. In addition, this non-monotonicity did not change my estimates of the coefficients on the effects of dayparts in (R4a) from the results I found in (R2a).

## 5.6 Regression Estimates and Estimates

In this section, I examine my results from each regression run in Table 21 (b) and described in section 5.5. The purpose of this section is to ensure my regression does not estimate values outside the possible bounds of my measure of ad niche-ness. Because my measure is the average of the levels of words in an ad and because levels are between 0 and 2, I find that my measure of ad niche-ness is between 0 and 2. Therefore, in this section, I check whether my regressions give me estimates between 0 and 2.

Table 22 shows sample statistics for the regression estimates and residuals from the regressions in Table 21 (b) and described in section 5.5.

[Table 22 about here.]

In Table 22, the estimate of a regression is  $\hat{y}$  where

$$\hat{y} \equiv \max\{\min\{\hat{y}^*, 2\}, 0\} \quad (9)$$

$$\hat{y}^* \equiv \begin{cases} \hat{\beta}_0 + \hat{\beta}_1 x & \text{for (R1 and R1a)} \\ \hat{\beta}_0 + \hat{\beta}_1 x + \sum_{i=1}^4 \hat{\beta}_{\text{daypart}_i} t_{\text{daypart}_i} & \text{for (R2)} \\ \hat{\beta}_0 + \hat{\beta}_1 x + \sum_{i=1}^4 \hat{\beta}_{\text{daypart}_i} \bar{t}_{\text{daypart}_i} & \text{for (R2a)} \\ \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 & \text{for (R3 and R3a)} \\ \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \sum_{i=1}^4 \hat{\beta}_{\text{daypart}_i} t_{\text{daypart}_i} & \text{for (R4)} \\ \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \sum_{i=1}^4 \hat{\beta}_{\text{daypart}_i} \bar{t}_{\text{daypart}_i} & \text{for (R4a)} \end{cases} \quad (10)$$

Here each  $\hat{\beta}$  is the regression estimate of the true parameter  $\beta$ . And the residual is defined as the difference  $y - \hat{y}$ .

For regressions (R1) - (R4) or equivalently for regressions using all observations, I find my estimate have the same mean of 0.126 as my measure of ad niche-ness and my residuals have a mean of zero, which is what you would expect from an Tobit regression. The standard deviation of my measure  $y$  of ad niche-ness is 0.136, which is about the same as the standard deviation of the residuals  $y - \hat{y}$ . This comes from the fact that although my estimates are significant, the language in webpages and ads still has a lot of variation that my regressions do not explain. Each of the estimates  $\hat{y}$  have a small standard deviation, and these estimates are all well within the bounds 0 and 2. They are all bounded below from about .11 and from above by about .3; none of the estimates are out of bounds of the possible measures of ad niche-ness.

I find similar results for regressions (R1a) - (R4a) or equivalently for regressions using unique observations of ad-webpage combinations. I find my estimate has the same mean of 0.124 as my measure of ad niche-ness and my residuals have a mean of about zero, which is what is expected from a Tobit regression. The standard deviation of my measure  $y$  of ad niche-ness is 0.146, which is about the

same as the standard deviation of the residuals  $y - \hat{y}$ . Each of the estimates  $\hat{y}$  have a small standard deviation, and these estimates are also all well within the bounds 0 and 2; none of these estimates are out of bounds of the possible measures of ad niche-ness.

## 6 Conclusion

Using a Gibbs sampling algorithm on the HLDA model, I found strong evidence for a non-monotonic relationship: The least niche ads tend to appear more on the moderately-narrow webpages. Niche firms tend to value advertising on fine segments of consumers and mass advertising more than general products. Perhaps this comes from the set up of the auction where adjusting bids is based on the click-through rate or because only the more general firms can do market research on broad segments of consumers. In future research, I plan on addressing this issue by analyzing the affect of webpage narrowness on ad revenue.

prices of ads should answer this question.

Future research should incorporate the regression into the topic model. If we believe that webpage content could affect the creation of the ad's content, then it should be built into the topic model. The only reason that I did not do so here is so that this paper can serve as an example of the usefulness of topic model in economics that others can build upon to further the understanding of this dynamic field.

## References

- Agarwal, Nikhil, Susan Athey, and David Yang**, "Skewed Bidding in Pay-per-Action Auctions for Online Advertising," *American Economic Review*, 2009, 99 (2), 441–47.
- Agirgas, Cagdas**, "What Drives Media Bias? A Panel Study of Newspaper Archives: 1990-2009," *Job Market Paper*, 2011.

- Anand, Bharat and Ron Shachar**, "Targeted advertising as a signal," *Quantitative Marketing and Economics*, 2009, 7, 237–266.
- Anderson, Simon P. and André de Palma**, "Shouting to be heard in advertising," *Working Paper*, 2010.
- **and Régis Renault**, "Pricing, Product Diversity, and Search Costs: A Bertrand-Chamberlin-Diamond Model," *The RAND Journal of Economics*, 1999, 30 (4), 719–735.
- **and —**, "Advertising Content," *The American Economic Review*, 2006, 96 (1), 93–113.
- **, André de Palma, and Jacques-Francois Thisse**, *Discrete Choice Theory of Product Differentiation*, MIT Press, 1992.
- Athey, Susan and Joshua S. Gans**, "The Impact of Targeting Technology on Advertising Markets and Media Competition," *American Economic Review*, 2010, 100 (2), 608–13.
- Baeza-Yates, R. and B. Ribeiro-Neto**, *Modern Information Retrieval*, New York: ACM Press, 1999.
- Bergemann, Dirk and Alessandro Bonatti**, "Targeting in advertising markets: implications for offline versus online media," *RAND J of Economics*, 2011, 42 (3), 417–443.
- Blei, David M. and John D. Lafferty**, "Dynamic Topic Models," in "ICML" 2006.
- **and —**, "Topic Models," in A. Srivastava and M. Sahami, eds., *Text Mining: Classification, Clustering, and Applications*, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, 2009.

– , **Andrew Ng, and Michael Jordan**, “Latent Dirichlet allocation,” *JMLR*, 2003, 3, 993–1022.

– , **Thomas Griffiths, Michael Jordan, and Joshua Tenenbaum**, “Hierarchical topic models and the nested Chinese restaurant process,” in “NIPS” 2003.

**Center for Media Design and the Sequent Partners**, “Video Consumer Mapping Study,” *Center for Media Design at Ball State University*, 2009.

**Chandra, Ambarish**, “Targeted Advertising: The Role of Subscriber Characteristics in Media Markets,” *The Journal of Industrial Economics*, 2009, 57 (1), 58–84.

**de Cornière, Alexandre**, “Search Advertising,” 2011. Paris School of Economics Job Market Paper.

**Diamond, Peter A**, “A Model of Price Adjustment,” *Journal of Economic Theory*, 1971, 3 (2), 156–68.

**Dickey, James M.**, “Multiple Hypergeometric Functions: Probabilistic Interpretations and Statistical Uses,” *Journal of the American Statistical Association*, 1983, 78 (383), 628–637.

**Dixit, Avinash and Joseph Stiglitz**, “Monopolistic Competition and Optimum Product Diversity,” *American Economic Review*, 1977, 67 (3), 297–308.

**Esteban, Lola, Agustín Gil, and José M. Hernández**, “Informative Advertising and Optimal Targeting in a Monopoly,” *The Journal of Industrial Economics*, 2001, 49 (2), 161–180.

– , **José M. Hernández, and José Luis Moraga-González**, “Customer Directed Advertising and Product Quality,” *Journal of Economics & Management Strategy*, 2006, 15 (4), 943 – 968.

**Esteves, Rosa Branca**, "Customer Poaching and Advertising," *The Journal of Industrial Economics*, 2009, 57 (1), 112–146.

**Galeotti, Andrea and José Luis Moraga-González**, "A Model of Strategic Targeted Advertising," May 2004. CESifo Working Paper No. 1196.

**Gentzkow, Matthew and Jesse M. Shapiro**, "Media Bias and Reputation," *Journal of Political Economy*, 2006, 114 (2), 280–316. Date revised - 2006-08-01; Language of summary - English; Pages - 280-316; ProQuest ID - 56516165; Corporate institution author - Gentzkow, Matthew; Shapiro, Jesse M; DOI - econlit-0859377; 0859377; 0022-3808.

**Ghose, Anindya and Sha Yang**, "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets," *Management Science*, 2009, 55 (10), 1605–1622.

**Goldfarb, A. and C. Tucker**, "Search Engine Advertising: Channel Substitution When Pricing Ads to Context," *Management Science*, 2011.

**Griffiths, Thomas L. and Mark Steyvers**, "Finding Scientific Topics," *PNAS*, 2004, 101 (suppl. 1), 5228–5235.

**Hofmann, Thomas**, "Probilistic latent semantic analysis," in "UAI" 1999.

**Iyer, Ganesh, David Soberman, and J. Miguel Villas-Boas**, "The Targeting of Advertising," *Marketing Science*, 2005, 24 (3), 461 – 476.

**Johnson, Justin P.**, "Targeted Advertising and Advertising Avoidance," 2013. forthcoming RAND J of Economics.

**Larcinese, Valentino, Riccardo Puglisi, and Jr Snyder James M.**, "Partisan Bias in Economic News: Evidence on the Agenda-Setting Behavior of U.S. Newspapers," *Journal of Public Economics*, 2011, 95 (9-10), 1178–1189. Date revised - 2011-09-01;

Language of summary - English; Pages - 1178-1189; ProQuest ID - 896012840; Corporate institution author - Larcinese, Valentino; Puglisi, Riccardo; Snyder, James M, Jr; DOI - econlit-1255076; 1255076; 10.1016/j.jpubecon.2011.04.006; 0047-2727.

**Lewis, Randall A. and David H. Reiley**, “Does Retail Advertising Work? Measuring the Effects of Advertising on Sales via a Controlled Experiment on Yahoo!,” *Working Paper*, 2011.

— **and** —, “Advertising Effectively Influences Older Users: How Field Experiments Can Improve Measurement and Targeting,” *Working Paper*, 2012.

**Lowy, Lisa Sharkis**, “The Existence and Characteristics of Dayparts on the Internet,” *The OPA White Papers*, 2003, 1 (3).

**Minka, Thomas and John Lafferty**, “Expectation-Propagation for the Generative Aspect Model,” *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, 2002, pp. 352–359.

**PricewaterhouseCoopers**, “IAB Internet Advertising Revenue Report: 2009 Full-Year Results,” *Interactive Advertising Bureau (IAB)*, April 2010.

**Rosen-Zvi, Michal, Tom Griffiths, Mark Steyvers, and Padhraic Smyth**, “The Author-Topic Model for Authors and Documents,” in “UAI” 2004.

**Shiman, Daniel R.**, “The Impact of Firms’ Increased Information about Consumers on the Volume and Targeting of Direct Marketing,” 1997. SSRN Working Paper.

**Steyvers, Mark and Tom Griffiths**, “Probabilistic Topic Models,” in T. Landauer, D. Mcnamara, S. Dennis, and W. Kintsch, eds., *Latent Semantic Analysis: A Road to Meaning.*, Laurence Erlbaum, 2006.

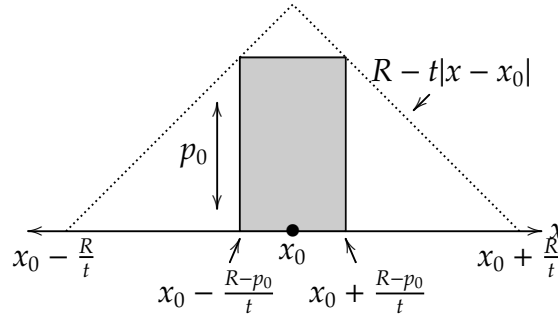
**Teh, Yee-Whye, David Newman, and Max Welling**, “A Collapsed Variational Bayesian Inference Algorithm for Latent Dirichlet Allocation,” in “NIPS” 2006.



- Teh, Yee Whye, Michael I. Jordan, Matthew J. Beal, and David M. Blei**, “Hierarchical Dirichlet Processes,” *JASA*, 2006, 101.
- Villas-Boas, J. M.**, “Dynamic competition with customer recognition,” *RAND J. Econom*, 1999, 30, 604-631.
- , “Price cycles in markets with customer recognition,” *RAND J. Econom*, 2004, 35, 486-501.
- Wei, Xing and Bruce Croft**, “LDA-based document models for ad-hoc retrieval,” in “SIGIR” 2006.
- Willmore, Christopher**, “Targeted persuasive advertising.” PhD dissertation, The University of British Columbia 2008.
- Wolinsky, Asher**, “True Monopolistic Competition as a Result of Imperfect Information,” *The Quarterly Journal of Economics*, 1986, 101 (3), 493–512.

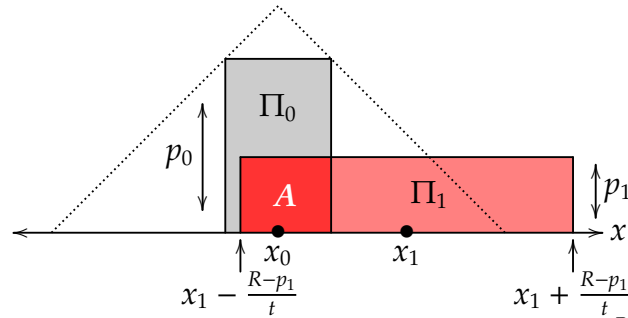
## Figures and Tables

Figure 1: Standard Hotelling Model



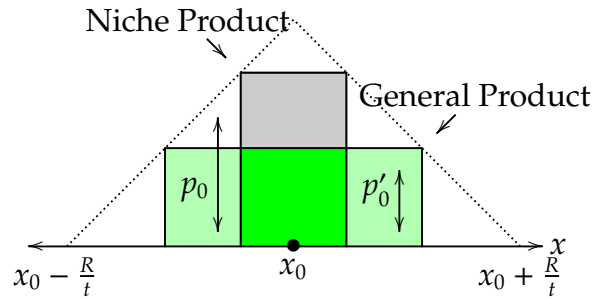
Note: For now price is exogenous (I relax this in A). The utility consumer  $x$  gets from buying from firm 0 is  $u_0(x) \equiv R - t|x - x_0| - p_0$ . Consumers in  $[x_0 - \frac{R-p_0}{t}, x_0 + \frac{R-p_0}{t}]$  buy from firm 0. Firm 0's profit is  $p_0 * 2\frac{R-p_0}{t}$ .

Figure 2: Hotelling Auction Duopoly: Example



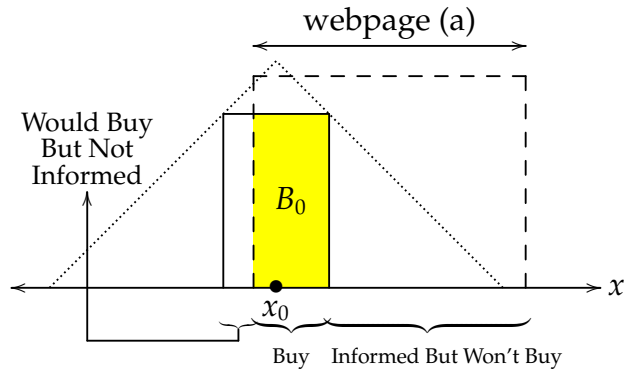
Note:  $A$  = advertising cost to firm 0. Profit of firm 0 is  $p_0 * 2\frac{R-p_0}{t} - A = \Pi_0$ .

Figure 3: Product Niche-ness vs. Price



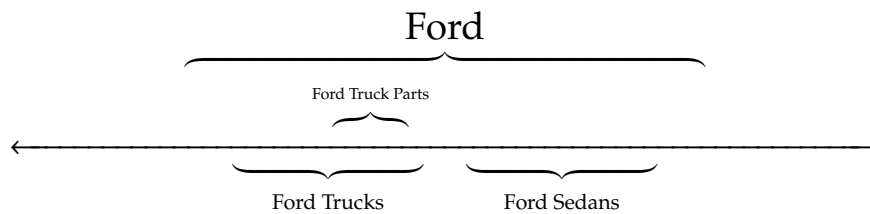
*Caveat: In a full model, niche-ness is a function of consumer preferences, and price is a result of niche-ness. Here, price is a proxy of niche-ness.*

Figure 4: Webpage Monopoly



*Note: Firm 0's advertising bidding function  $b_0$  is restricted to be constant for each webpage interval. This is the same as one bid:  $B_0 = \int b_0(x) 1\{x \in \text{webpage interval}\} dx$ . Further note: The profit for firm 0 is  $B_0$ , because it is a monopolist.*

Figure 5: Webpage Title Segmentation

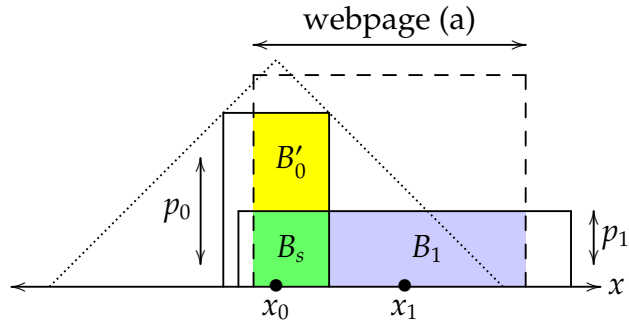


*The number line  $\mathbb{R}$  is now split into webpage intervals. A 'Ford' webpage delivers a larger segment of consumers than a 'Ford Trucks' webpage. A 'Ford Trucks' webpage is narrower than a 'Ford' webpage.*

*Note: Webpage narrow-ness is not related to product niche-ness. They can be different segments along the same number line.*

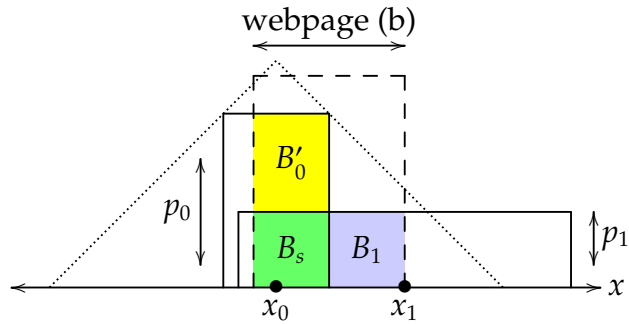
Figure 6: Hotelling Webpage Duopoly: Example

(a) General Product wins General Webpage



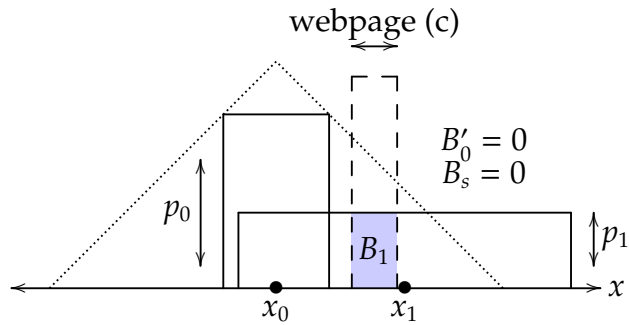
Note: Firm 0 is bidding  $B'_0 + B_s = B_0$ . Firm 1 is bidding  $B_1 + B_s$ .  
 $B_1 > B'_0$ , so firm 1 wins the auction.

(b) Narrower Webpage



Note:  $B'_0 > B_1$ , so firm 0 wins the auction.

(c) General Product wins Narrowest Webpage



$B_1 > B'_0$ , so firm 1 wins the auction.

Figure 7: Example Webpage.

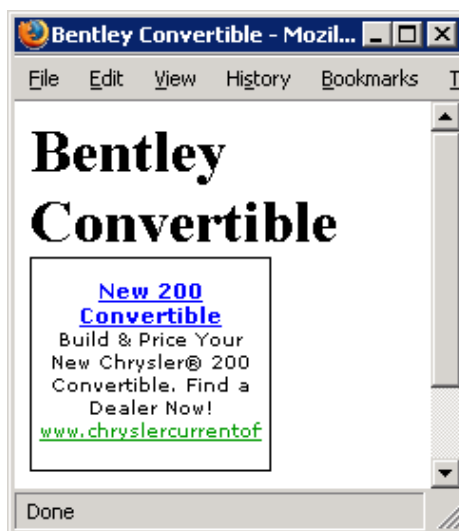


Table 8: Sample Data

My Program.		From Flippa	Determined by Advertisers	
Time and Date	Webpage Content	Price	Ad URL	Ad Content
...	...	...	...	...
7/9/2013 3:07 PM	configuration diesels reference guide and turbocharger ...	\$49	trucksmartsales.com	TruckSmart Service Center Light and Medium Duty Truck Service Diagnostics, Repairs, Inspections.
7/9/2013 3:07 PM	contact we pc market privacy custom this magazine ...	\$499	davishyundai.com	Used Car Sale Trenton NJ Over 700 Vehicles In Our Inventory Come Find Yours Today.
7/9/2013 3:07 PM	citroen home mitsubishi ford daihatsu calculator ...	\$130	davishyundai.com	Used Car Sale Trenton NJ Over 700 Vehicles In Our Inventory Come Find Yours Today.
...	...	...	...	...
7/10/2013 11:09 AM	configuration diesels reference guide and turbocharger ...	\$49	wholesaletransrepair.com	Transmissions Fixed Cheap 732-738-3834, Free Diagnostic. Financing Available, Free Loner Car.
...	...	...	...	...

*Note: I observe different ads on the same webpage at different times.  
Further note: I also observe the same ad on different webpages.*

Table 9: Top Words from 5 Topics Estimated by a Latent Dirichlet Allocation Algorithm

Topic 1: Computing & Servers	Topic 2: Tires & Cars	Topic 3: Shipping, Valuation, Education, & Cameras	Topic 4: Social Media & Hotrods	Topic 5: German
1. cloud 2. hosting 3. trial 4. netsuite 5. 15t 6. pavilion 7. ftp 8. dv6t 9. dns 10. hp 11. 17t 12. free 13. server 14. quad 15. envy 16. crm 17. vps 18. backup 19. marketing 20. tego 21. management 22. software 23. sampling 24. odbc 25. access 26. i5-3230m 27. ghz 28. manage 29. franchise 30. download	1. manuals 2. tire 3. tires 4. barbie 5. michelin 6. 4-7 7. wheels 8. honda 9. rims 10. \$69 11. mpg 12. chip 13. mazda 14. bmw 15. selector 16. horsepower 17. toyota 18. subaru 19. cadillac 20. bfgoodrich 21. manual 22. nissan 23. freightliner 24. dealer 25. suzuki 26. volvo 27. delmarva 28. prices 29. dart 30. rover	1. shipping 2. erase 3. vin 4. 60% 5. adorama 6. \$100,000 7. valuation 8. cordon 9. bleu 10. rebel 11. tablets 12. 3)\$1 13. 2)free 14. car's 15. 40% 16. buys 17. parts 18. educator 19. dslr 20. campus-enroll 21. dns 22. reliability 23. 5s 24. cameras 25. 35% 26. skins 27. jewelry 28. bags 29. cancer 30. headphones	1. they 2. speedyrock 3. wpturbo 4. you 5. stroker 6. wordpress 7. facebook 8. \$190 9. blogging 10. donate 11. mastermind 12. wiseco 13. april 14. plugin 15. bluehost 16. empower 17. esb 18. paving 19. piston 20. hotrods 21. cobra 22. gasket 23. 2,000 24. salvation 25. table 26. nobody 27. still 28. bore 29. hosting 30. kibblewhite	1. der 2. ford 3. charting 4. und 5. emr 6. auf 7. sie 8. oder 9. degree 10. edd 11. diese 12. unlock 13. zu 14. umuc 15. torrent 16. dealer 17. zum 18. eine 19. von 20. sind 21. werden 22. taurus 23. extendd 24. den 25. nicht 26. inhalte 27. f-150 28. seiten 29. anfang 30. weiter

Topic names are created by the econometrician from looking at the top words in that topic. Top words are not the most probable words in a toipic, because the same word can appear in multiple topics. Top words are those with the highest term-score (from Blei and Lafferty, 2009, see (3)).

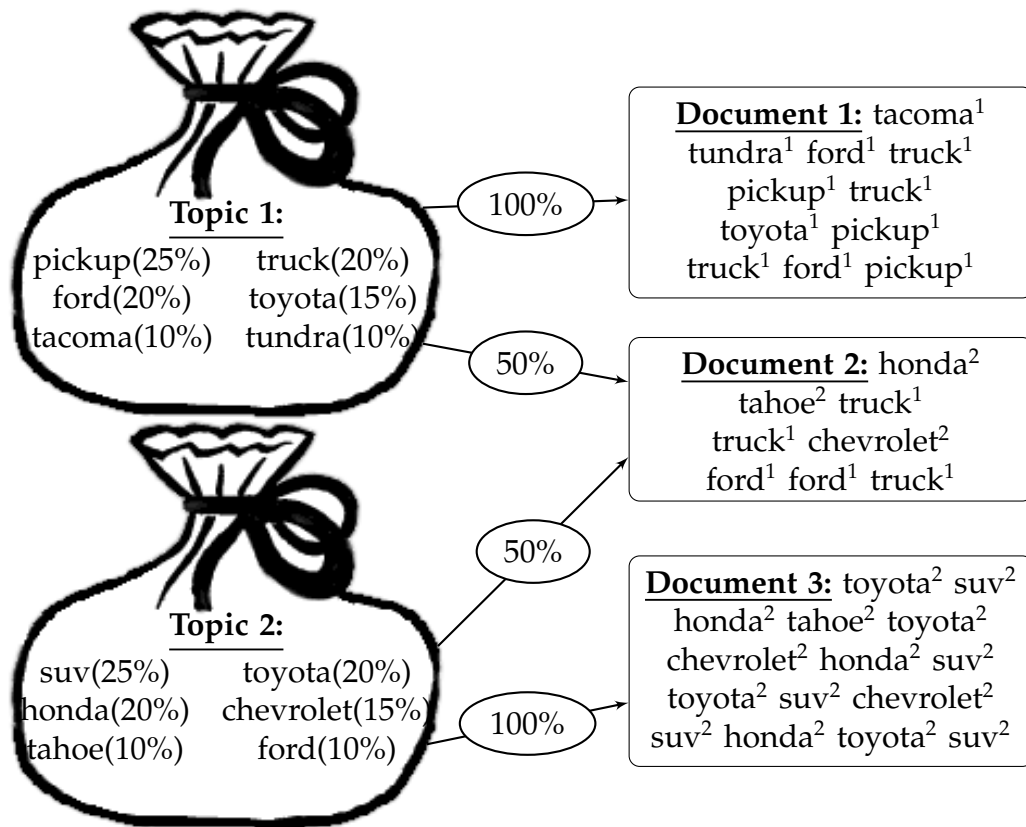
Table 10: Sample Ads with Estimated Topics from a Latent Dirichlet Allocation (LDA) Algorithm

<b>PURE TOPIC 1 (Computing &amp; Servers) AD:</b> <b>Dashboard Analysis Powerful Business Software for Dashboarding and Score-carding.</b>
<b>MOSTLY TOPIC 2 (Tires &amp; Cars) AD:</b> <u>Pre-Owned VW Beetle Search Pre-Owned Inventory. See Special Offers on a VW Beetle.</u>
<b>MIXED TOPIC 2 AND TOPIC 3 WEBSITE:</b> <u>cash cars we cars cash cash buy for hour cars cars</u>
<b>MIXED AD:</b> <u>Mechanic Ripping You Off? Find Out Now From Our Experts. Ready To Chat. 100% Guaranteed</u>

*In LDA, each word is a latent draw from a single topic, and different words in the same ad may be drawn from different topics.*



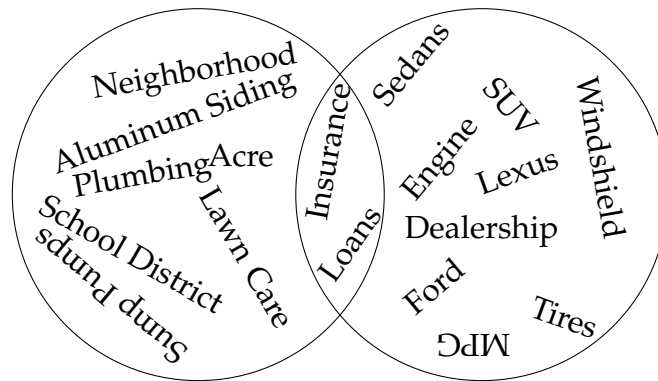
Figure 11: Latent Dirichlet Allocation Example



*The probability of choosing a specific bag of words depends on the document (unique webpage or ad). Words are then selected for each document by first drawing a random bag of words, and then drawing a word from the bag of words.*

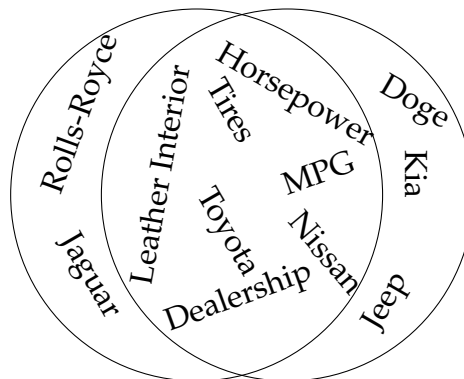
Figure 12: Identification of  $\bar{\alpha}$ 

(a) Small Intersection Between Topics

Home OwnershipAutomobiles

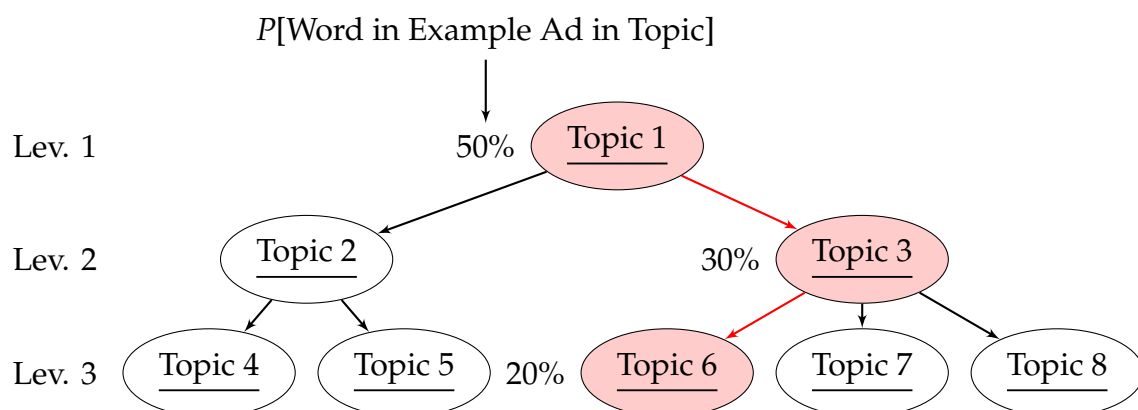
*A small intersection between topics identifies a small  $\bar{\alpha}$ .*

(b) Large Intersection Between Topics

Luxury CarsSUVs

*A large intersection between topics identifies a large  $\bar{\alpha}$ .*

Figure 13: Hierarchical Topics



*The topic path chosen (the red filled in topics) is different from the set of probabilities of words chosen in each ad. The topic path only restricts the possible topics that a word can be drawn from. LDA is run on the smaller set of possible topics for each ad, after its topic path is chosen.*

Table 14: Estimated Topic Allocation by  $\gamma$  (iteration 10,000).

*The  $\gamma$  controls the probability of forming a new topic.*

	$\gamma = .2$				$\gamma = .25$				$\gamma = .3$			
	Topic	Parent	Docs	Words	Topic	Parent	Docs	Words	Topic	Parent	Docs	Words
Level 0:	1	—	16,179	238,794	1	—	16,179	234,725	1	—	16,179	239,223
Level 1:	2	1	10,228	13,990	2	1	13,686	21,472	2	1	8,553	11,796
	3	1	1,633	2,943	3	1	727	8,424	3	1	2,745	8,402
	4	1	958	7,359	4	1	586	2,472	4	1	1,271	6,905
	5	1	924	3,048	5	1	407	2,862	5	1	1,270	2,465
	6	1	906	3,762	6	1	350	737	6	1	1,218	3,322
	7	1	849	2,101	7	1	247	2,213	7	1	706	1,210
	8	1	510	2,239	8	1	89	2,557	8	1	212	1,057
	9	1	133	2,552	9	1	65	127	9	1	155	2,594
	10	1	23	213	10	1	21	200	10	1	32	241
	11	1	8	23	11	1	1	1	11	1	8	13
	12	1	3	3					12	1	4	5
	13	1	3	4					13	1	4	3
	14	1	1	2					14	1	1	3
	Total:		16,179	38,239	Total:		16,179	41,065	Total:		16,179	38,016
Level 2:	15	2	10,172	1,698	12	2	13,116	3,063	15	2	8,440	1,377
	16	2	23	2	13	2	551	128	16	2	40	0
	17	2	8	5	14	2	11	1	17	2	39	2
	18	2	7	0	15	2	8	3	18	2	33	4
	19	2	6	1	16	3	350	237	19	2	1	0
	20	2	3	2	17	3	257	183	20	3	2,641	499
	21	2	3	1	18	3	81	43	21	3	68	11
	22	2	2	0	19	3	31	12	22	3	17	3
	23	2	2	1	20	3	8	6	23	3	17	1
	24	2	1	1	21	4	461	200	24	3	1	4
	25	2	1	1	22	4	120	43	25	3	1	0
	26	3	1,605	331	23	4	3	1	26	4	1,261	311
	27	3	16	3	24	4	2	2	27	4	10	1
	28	3	10	2	25	5	347	170	28	5	578	91
	29	3	2	1	26	5	46	68	29	5	535	108
	30	4	941	289	27	5	6	4	30	5	126	29
	31	4	8	0	28	5	6	1	31	5	11	1
	32	4	8	4	29	5	2	0	32	5	6	1
	33	4	1	0	30	6	104	27	33	5	6	3
	34	5	866	232	31	6	95	29	34	5	5	0
	35	5	29	9	32	6	94	36	35	5	2	0
	36	5	21	9	33	6	57	17	36	5	1	0
	37	5	5	3	34	7	151	49	37	6	1,123	261
	38	5	3	1	35	7	65	32	38	6	93	17
	39	6	607	109	36	7	16	4	39	6	2	1
	40	6	299	60	37	7	7	3	40	7	323	67
	41	7	680	224	38	7	4	3	41	7	321	68
	42	7	95	24	39	7	2	1	42	7	59	11
	43	7	67	20	40	7	1	1	43	7	3	2
	44	7	6	1	41	7	1	0	44	8	105	14
	45	7	1	1	42	8	30	8	45	8	96	53
	46	8	326	76	43	8	21	10	46	8	10	0
	47	8	165	22	44	8	20	4	47	8	1	1
	48	8	16	3	45	8	18	8	48	9	91	9
	49	8	3	0	46	9	54	15	49	9	47	6
	50	9	68	19	47	9	8	1	50	9	9	9
	51	9	65	18	48	9	1	0	51	9	4	0
	52	10	15	2	49	9	1	0	52	9	3	2
	53	10	8	8	50	9	1	1	53	9	1	0
	54	11	4	1	51	10	16	16	54	10	32	10
	55	11	3	1	52	10	2	1	55	11	8	4
	56	11	1	1	53	10	2	1	56	12	4	0
	57	12	2	0	54	10	1	0	57	13	2	2
	58	12	1	0	55	11	1	0	58	13	1	0
	59	13	2	1					59	13	1	0
	60	13	1	0					60	14	1	0
	61	14	1	2								
	Total:		16,179	3,189	Total:		16,179	4,432	Total:		16,179	2,983

$\gamma = .25$  produces few low-count topics, which means it manages to fit more of the ads to the common topics. In addition, it produces a good balance between number of words in each level and a more manageable number of topics.

Table 15: Number of Topics in Each Level with Low Word Count by  $\gamma$  (iteration 10,000)

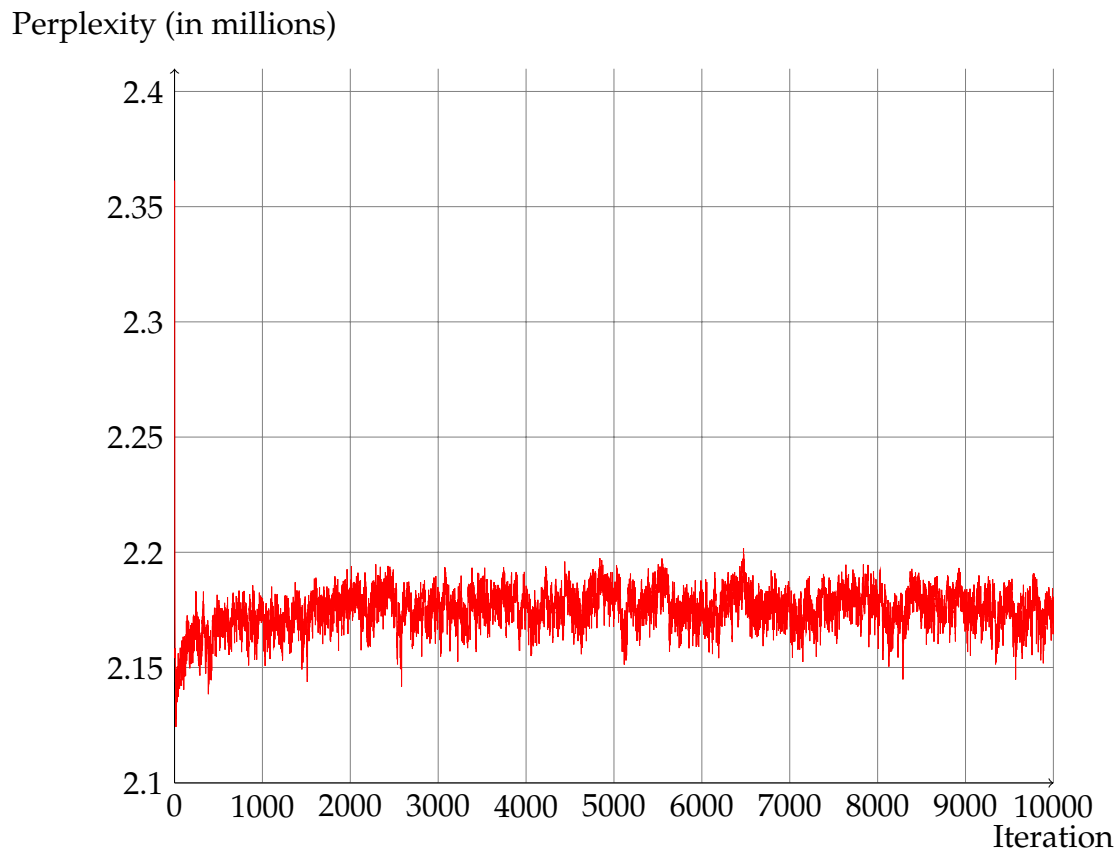
*The  $\gamma$  controls the probability of forming a new topic.*

Level	Number of Topics with Word Count $\leq$	$\gamma = 0.01$	$\gamma = 0.05$	$\gamma = 0.1$	$\gamma = 0.15$	$\gamma = 0.2$	$\gamma = 0.25$	$\gamma = 0.3$	$\gamma = 0.35$	$\gamma = 0.4$
0	$\infty$	1	1	1	1	1	1	1	1	1
	50	0	0	0	0	0	0	0	0	0
	20	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
1	$\infty$	14	11	14	14	13	10	13	15	12
	50	5	2	5	5	4	1	4	3	5
	20	5	1	4	3	3	1	4	2	5
	10	4	1	4	1	3	1	3	2	4
	5	2	0	1	1	3	1	3	2	4
	1	1	0	1	1	0	1	0	1	0
	0	0	0	1	0	0	0	0	0	0
2	$\infty$	41	40	52	43	47	44	46	52	43
	50	35	32	44	37	39	37	37	45	34
	20	33	28	40	34	37	30	36	39	29
	10	28	24	38	31	34	26	32	33	25
	5	27	19	32	28	31	22	28	28	21
	1	19	13	22	22	21	15	19	22	12
	0	14	9	13	15	8	6	13	15	9

Word Count was calculated by taking the number of estimated words drawn from a topic (in each unique ad).

Note:  $\gamma = .25$  produces few low count topics, which means it manages to fit more of the ads to the common topics. In addition, it produces a more manageable number of topics.

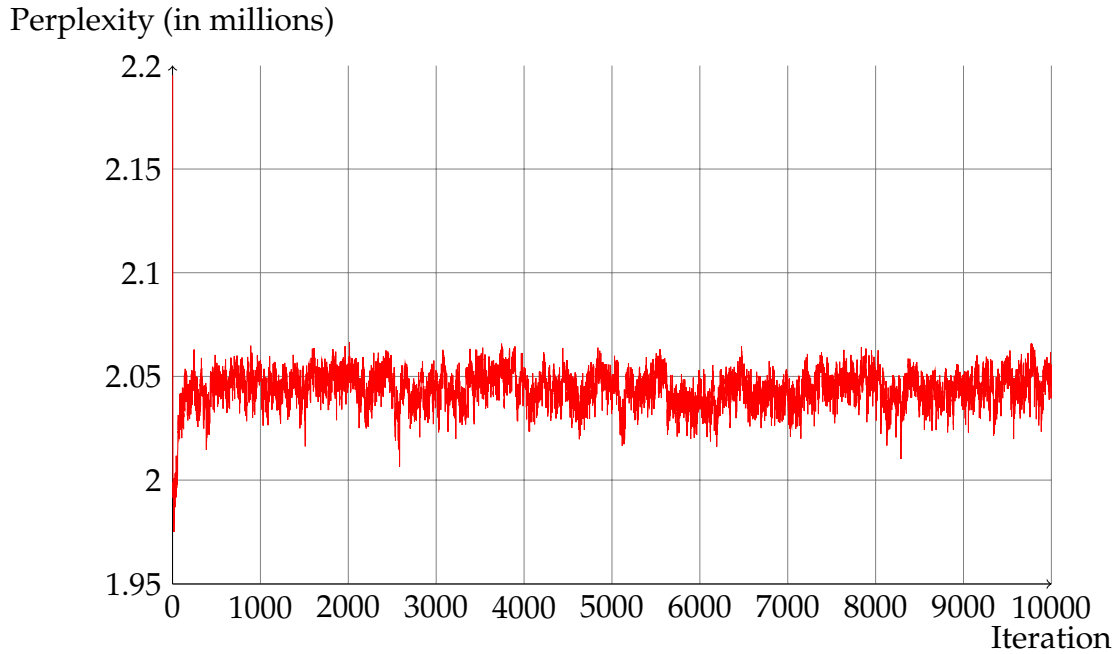
Figure 16: Convergence of Perplexity



*In language modeling, it is conventional to use perplexity instead of likelihood (see: Blei et al., 2003a). Perplexity is monotonically decreasing in the likelihood, and is equivalent to the inverse of the geometric mean of the per-word likelihood. A lower perplexity indicates a higher likelihood, or equivalently better performance.*

Figure 17: Convergence of Estimates of  $\vec{\eta}$ 

$\eta_l$  is the symmetric Dirichlet parameter that determines the random draw of a probability  $\beta_{k,v}$  of seeing a word  $v$  in a topic  $k$  from level  $l$ .

(a)  $\vec{\eta}$  Perplexity

This is the perplexity conditional on the  $\vec{\eta}$  parameters being true. In language modeling, it is conventional to use perplexity instead of likelihood (see: Blei et al., 2003a; Rosen-Zvi et al., 2004). Perplexity is monotonically decreasing in the likelihood, and is equivalent to the inverse of the geometric mean of the per-word likelihood. A lower perplexity indicates a higher likelihood, or equivalently better performance.

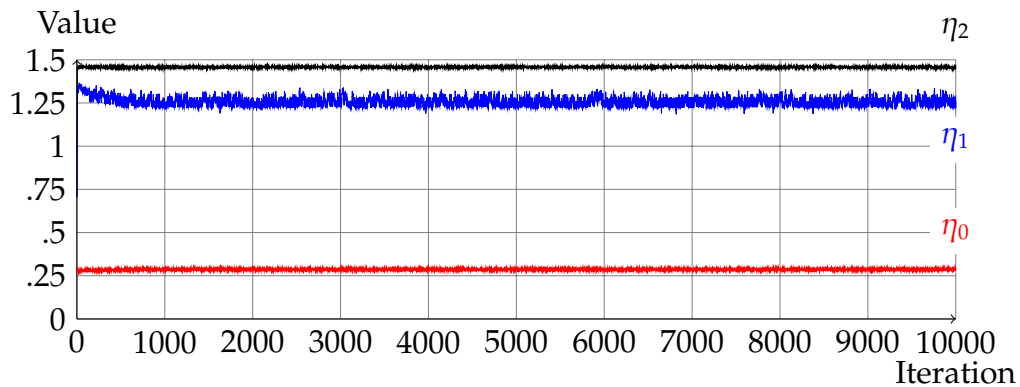
(b)  $\vec{\eta}$  Values

Table 18: Sample Frequency of Word Level Conditional on the Word Level of another Given Word in the Same Ad or Webpage

(a) Ads (All observations)

		Sample Frequency of Word (in ad)		
		Level 0	Level 1	Level 2
Conditional on a given word in same ad being	Unconditional	88.60%	10.13%	1.27%
	Level 0	89.82%	9.06%	1.15%
	Level 1	83.23%	15.07%	1.71%
	Level 2	83.28%	13.50%	3.26%

(b) Ads (Unique Ads)

		Sample Frequency of Word (in ad)		
		Level 0	Level 1	Level 2
Conditional on a given word in same ad being	Unconditional	86.05%	12.11%	1.85%
	Level 0	87.89%	10.54%	1.59%
	Level 1	78.99%	18.41%	2.61%
	Level 2	78.39%	17.25%	4.38%

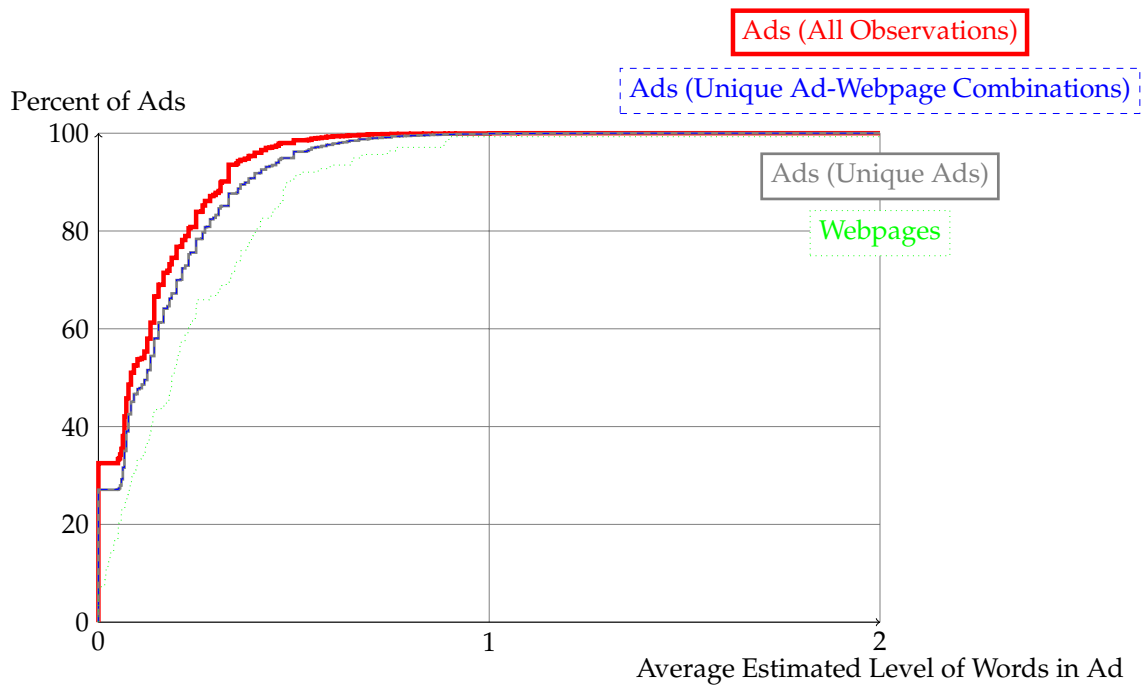
(c) Webpages

		Sample Frequency of Word (in webpage)		
		Level 0	Level 1	Level 2
Conditional on a given word in same webpage being	Unconditional	77.56%	20.19%	2.25%
	Level 0	82.85%	15.77%	1.38%
	Level 1	56.75%	41.31%	1.94%
	Level 2	69.27%	27.10%	3.62%

Note: These are data across unique webpage. Frequencies are the same (or very close) across all observations because there are roughly the same number of each webpage.

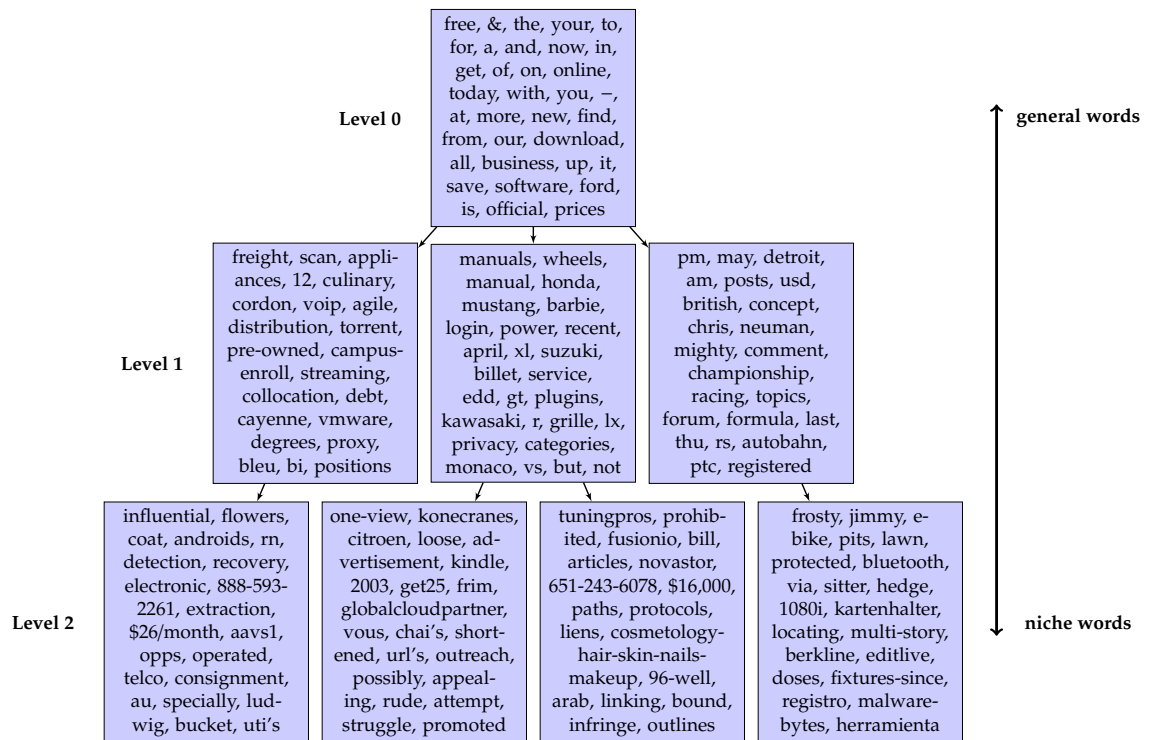


Figure 19: Cumulative Distribution Function of Average Estimated Level of Words in Ad



Levels of words were calculated by taking the mode of the estimated level from iterations 1,000, 2,000, 3,000, ..., and 10,000 of Blei et al. (2003b)'s Gibbs sampling algorithm for HLDA.

Figure 20: Top Words from Topics Identified by a Hierarchical Latent Dirichlet Allocation Algorithm



Each box represents a different topic identified by HLDA. Words in clusters or topics further down the tree are more specific; therefore, they are more niche.

*Note: Only the most common topics, or topics that occur frequently, are shown. There is only one level zero topic; it is shown. There are ten level one topics; the three topics shown are the only topics estimated to cover 2,862 or more words in my data (unique ads and webpages). There are forty-four level two topics; the four topics shown are the only topics that satisfy both: (1) have a parent topic of the three most common level one topics and (2) cover 170 or more words in my data.*

Table 21: Results Using HLDA Measure of Niche-ness

*Average level of words in ad estimated by the HLDA Algorithm.*

(a) Sample Statistics on Ad Niche-ness & Webpage Narrowness

	Mean	StDev	Min	Max	N
Ad Niche-ness					
All Observations	.127	.136	0	1.333	893,614
Unique Ads	.158	.165	0	1.333	15,970
Unique Ad-Webpage Combinations	.123	.146	0	1.333	77,507
Webpage Narrowness					
All Observations	.247	.253	0	2	893,614
Unique Webpages	.247	.253	0	2	138
Unique Ad-Webpage Combinations	.251	.256	0	2	77,507

(b) Tobit Regressions of Webpage Narrowness on Ad Niche-ness

		All Observations				Unique Ad-Webpage Combinations				
		Linear Regressor		Quadratic Regressors		Linear Regressor		Quadratic Regressors		
		(R1)	(R2)	(R3)	(R4)	(R1a)	(R2a)	(R3a)	(R4a)	
Time of Day	Webpage	Narrowness	0.012*** (0.001)	0.012*** (0.001)	-0.119*** (0.002)	-0.119*** (0.002)	-0.007* (0.003)	-0.006* (0.003)	-0.042*** (0.006)	-0.041*** (0.006)
		Narrowness <sup>2</sup>	—	—	0.103*** (0.001)	0.103*** (0.001)	—	—	0.027*** (0.004)	0.027*** (0.004)
	Time of Day	Early Morning	—	0.004 (0.002)	—	0.003 (0.002)	—	0.014 (0.008)	—	0.014 (0.008)
		Evening	—	-0.012*** (0.001)	—	-0.012*** (0.001)	—	0.013* (0.005)	—	0.013* (0.005)
		Late Night	—	-0.004*** (0.001)	—	-0.004** (0.001)	—	0.001 (0.005)	—	0.001 (0.005)
		Weekend	—	-0.011*** (0.001)	—	-0.011*** (0.001)	—	-0.009* (0.004)	—	-0.008* (0.004)
		Constant	0.084*** (0.000)	0.094*** (0.001)	0.104*** (0.000)	0.114*** (0.001)	0.074*** (0.001)	0.080*** (0.004)	0.080*** (0.001)	0.085*** (0.004)
		$\sigma$	0.186*** (0.000)	0.186*** (0.000)	0.185*** (0.000)	0.185*** (0.000)	0.207*** (0.001)	0.207*** (0.001)	0.207*** (0.001)	0.207*** (0.001)
		p	0.0000	0.0000	0.0000	0.0000	0.0309	0.0000	0.0000	0.0000
		N	893,614	893,614	893,614	893,614	77,507	77,507	77,507	77,507

\* = 5% significance, \*\* = 1% significance, and \*\*\* = .1% significance.

*Note: Results were similar for OLS regressions.*

Table 22: Residual Sample Statistics for Regressions

	<u>Mean</u>	<u>St. Dev.</u>	<u>Min.</u>	<u>Max.</u>
<b>All Observations</b> (N = 893,614)				
"True Value"	0.127	0.136	0.000	1.333
Estimate (R1)	0.126	0.002	0.124	0.140
Estimate (R2)	0.126	0.003	0.122	0.150
Estimate (R3)	0.126	0.016	0.114	0.284
Estimate (R4)	0.126	0.016	0.112	0.296
Residual (R1)	0.001	0.136	-0.140	1.209
Residual (R2)	0.001	0.136	-0.150	1.210
Residual (R3)	0.001	0.135	-0.284	1.220
Residual (R4)	0.001	0.135	-0.296	1.216
<b>Ad-Webpage Combinations</b> (N = 77,507)				
"True Value"	0.123	0.146	0.000	1.333
Estimate (U1)	0.124	0.001	0.117	0.125
Estimate (U2)	0.124	0.004	0.115	0.138
Estimate (U3)	0.124	0.004	0.118	0.146
Estimate (U4)	0.124	0.005	0.117	0.159
Residual (U1)	-0.001	0.146	-0.125	1.211
Residual (U2)	-0.001	0.146	-0.138	1.211
Residual (U3)	-0.001	0.146	-0.146	1.215
Residual (U4)	-0.001	0.146	-0.159	1.213

For regressions see table 21 (b). "True Value" refers to the value of ad niche-ness estimated by the HLDA algorithm. Estimate (x) refers to the estimated value of ad niche-ness and Residual (x) refers to the residual using regression (x).

## Appendix

### A Many Firms / Endogenous Prices

Now let's consider a case of where there are  $N$  firms bidding in an auction for an auction for a single webpage. Let it be common knowledge that the center of

the webpage interval be located at  $x_w = 0$  and the length of its interval be  $\ell > 0$ . Let each firm  $j = 1, \dots, N$  receives a random, privately-known, horizontal product characteristic or location of  $x_j \sim U[-M, M]$ , where  $M > \ell$ . Then the quantity  $q_j$  of consumers delivered by the webpage that would buy product  $j = 1, 2, 3, \dots, N$  at the price  $p_j$  would be given by equation (11). Giving a first-order pricing condition of equation (12).

$$q_j(p_j, \ell, |x_j|) = \begin{cases} 0 & \text{if } |x_j| \geq \ell + \frac{R-p_j}{t} \\ 2\ell & \text{if } |x_j| \leq \frac{R-p_j}{t} - \ell \\ 2\frac{R-p_j}{t} & \text{if } |x_j| \leq \ell - \frac{R-p_j}{t} \\ \frac{R-p_j}{t} + \ell - |x_j| & \text{otherwise} \end{cases} \quad (11)$$

$$p_j^*(\ell, |x_j|) = \begin{cases} \frac{R}{2} & \text{if } |x_j| \leq \ell - \frac{R}{2t} \\ R - t(\ell - |x_j|) & \text{if } \ell - \frac{R}{2t} \leq |x_j| \leq \ell - \frac{R}{3t} \\ \frac{R+t(\ell-|x_j|)}{2} & \text{if } \ell - \frac{R}{3t} \leq |x_j| \leq \ell = \frac{R}{t} \\ \text{anything} & \text{if } |x_j| \geq \ell = \frac{R}{t} \end{cases} \quad (12)$$

It is straight forward to show using equation (12) that the closest firm to  $x_w = 0$  would win the auction. From this result and equation (12), the closest firm (and therefore the winner of the advertising auction) may be a lower-priced, more general product or a higher-priced, more niche product. In general, we can simply say that firm  $j$  is niche if  $\frac{R-p_j}{t} > \frac{R-p}{t}$  or equivalently if  $p_j > \underline{p}$ . Let  $\text{underbarp} > R/2$ .<sup>48</sup> It follows that the winning advertiser is niche when  $\ell - \frac{R-p}{t} < |x_j| < \ell = \frac{R-2p}{t}$ ; therefore, the closest and furthest firms from the webpage sell general products, while the niche products are sold in a band around  $x_w = 0$ .

The probability that at least one firm is within  $x \in [0, M]$  of  $x_w = 0$  is  $1 - (1 - \frac{x}{M})^N$ . I consider the limit case where there are many firms and  $M$  is huge; I take  $M$  and  $N$

---

<sup>48</sup> If  $\text{underbarp} < R/2$ , then the winning advertiser is niche when  $|x_j| < \ell - \frac{R-2p}{2t}$ ; therefore the closest firms sell the niche products. This is an uninteresting case.

to infinity and hold  $\lambda \equiv \frac{N}{M}$  constant. This probability that at least one firm is within  $x \in [0, M]$  of  $x_w = 0$  becomes  $1 - e^{-\lambda x}$ . Therefore, the probability density function of the closest firm to  $x_w = 0$  (and therefore the winner of the advertising auction) is  $\lambda e^{-\lambda x}$ . The probability that the closest firm to  $x_w = 0$  (and therefore the winner of the advertising auction) is a niche product is given by equation (13).

$$Pr(\text{Niche Firm Wins}) = \begin{cases} e^{-\lambda \ell} (e^{\lambda \frac{R-p}{t}} - e^{-\lambda \frac{R-2p}{t}}) & \text{if } \ell > \frac{R}{2t} \\ 1 - e^{-\lambda(\ell + \frac{R-2p}{t})} & \text{if } \ell < \frac{R}{2t} \end{cases} \quad (13)$$

Notice when  $\ell > \frac{R}{2t}$  or equivalently when the webpage is generally-focused, as a webpage becomes more narrowly-focused (as  $\ell$  shrinks), a niche product would have an increasing chance of winning the auction. Additionally when  $\ell < \frac{R}{2t}$  or equivalently when the webpage is narrowly-focused, as a webpage becomes more narrowly-focused (as  $\ell$  shrinks) a niche product would have a decreasing chance of winning the auction. This shows that the relationship between the niche-ness of the advertised product and the narrowness of the webpage is not necessarily monotonic, and therefore is an empirical question.

To empirically test which kind of firm win the auction for advertising on a webpage, I regress ad niche-ness (my independent variable) on webpage narrowness (my dependant variable). If I knew the webpage narrowness and the relative locations of the firms and webpages, then I could predict perfectly which kind of firm (niche or general) would win the auction to advertise on a webpage, because I would be able to solve the firm's problem for pricing, ignoring any multiple equilibria or the influence of outside markets on price. However, because I do not know locations and pricing, I have an error in the kind of firm that wins the auction for advertising, so I need to investigate which type of firm tends to win the bid for advertising on which type of webpage.

## Chapter II

---

# Imperfect Targeting of Advertising and Privacy Regulations

### 1 Introduction

When a firm gathers more information on a consumer, it has a more accurate signal on that consumer's taste. I will refer to this as *signal accuracy*. Firms use these signals to determine which consumers to target their ads. *Targeted advertising* is the personalizing of advertisements to fit consumers' tastes. Consumers, including myself, appreciate this because they get informed of a greater selection of more personalized products. I appreciate getting ads on products like DVDs and books, instead of ads on baby food or senior living, because I am more likely to buy DVDs and books.

Yet if firms gather too accurate information on consumer taste, then consumers are pestered with too many annoying ads. If a consumer buys a general product, then he will not gain much attention, but if a consumer buys a niche product, then he will be deluged with ads. If a consumer gives to the United Way, then he will not gain much attention, but if a consumer gives to the Charlottesville Fire Department, then he will receive tons of heart-throbbing letters from various emergency services around the Charlottesville and Richmond area. If the viewers of a television program are very homogeneous, then they will be pestered with more advertisements.

The Federal Trade Commission (FTC) and the European Parliament have struggled with how to regulate targeted advertising to protect consumers' personal information. Goldfarb and Tucker (2011) showed that these privacy regulations in the European Union decreased the effectiveness of online ads. If better targeting of advertising is beneficial to consumers, then why would we want to protect our information?

It could be that people value privacy because they want to protect against the criminal use of their information. Yet even with identity theft protection, regulation, and insurance, people feel uncomfortable with businesses knowing too much of their personal information. It could be that people value privacy because it contains some psychological benefit. This would make privacy a final good, instead of an intermediate. Yet Google has consumers opt out of their targeted advertising program, when the personal information used to target is stored on the consumer's personal computer. It could be that people value privacy because they want to avoid price discrimination. This possibility was explored when firms target advertise by Bergemann and Bonatti (2011). In this paper, I propose another explanation for why consumers value privacy with advertisement annoyance. When firms get too much information about consumers, too many firms enter the market, pestering consumers with too many annoying ads. Personally, I dislike receiving junk mail and spam email more than I am worried about companies misusing my credit card numbers.<sup>49</sup> Especially, because my credit card provider protects me with account monitoring, password protection, and a guarantee to cover stolen funds.

In this paper, I explore the effect of increasing signal accuracy on consumer welfare. In my model, firms will face free entry, and therefore make no profit. I do not model any welfare from information generation or ad platform profit, so consumer welfare and social welfare may not be equivalent. Nor do I model any

---

<sup>49</sup> Don't tell the criminals.



effect from the product price. Although product prices are endogenous, the form of the demand curve is such that firms will always choose to set the same price. There are plenty of other papers that already explore targeting and pricing (see for example: Bergemann and Bonatti, 2011).

As each firm gets more accurate information on each consumer's taste, firms switch from advertising to all consumers, or equivalently *mass advertising*, to targeted advertising. Consumers do not benefit from this switch to targeted advertising through the selection of products offered to them. I find that the initial cost of being offered fewer products equals the initial benefit from being offered better matching products, because the signal accuracy is so low. Yet consumers do benefit from this switch to targeted advertising through a reduction in annoying ads. Although targeted advertising induces more firms to enter the market, fewer firms advertise to each consumer, because each firm is advertising to a smaller, targeted segment of the market.

When firms target advertise, signal accuracy has an ambiguous effect on consumer welfare. A higher signal accuracy increases welfare by inducing firms to offer each consumer more products, which are better matched to that consumer's taste. I refer to this as the *product selection benefit*. Yet a higher signal accuracy decreases welfare by inducing more firms to enter the market and pester consumers with a greater number of annoying ads. I refer to this as the *ad annoyance cost*.

I find that the product selection benefit is increasing and convex in signal accuracy. This means that as firms receive better signals about each consumer's taste, the benefit to consumers from the products they buy increases at an increasing rate. Because of this result, I find that if the ad annoyance is convex enough in the number of ads a consumer sees, then consumer welfare has an inverted-u-shape with respect to signal accuracy. This means that if the marginal annoyance of each additional ad is increasing fast enough, then welfare could be optimized with an interior signal accuracy. Otherwise consumer welfare has a u-shape with respect

to signal accuracy. Then welfare is maximized with either firms getting a perfect signal on consumers' tastes or firms getting a signal just accurate enough to induce targeted advertising.

In section 6, I extend my model to consider the case where each firm can pay to get a more accurate signals about consumers' tastes. I call this the *marketing cost*. Firms collect information on consumers through surveys, tracking behavior, locational characteristics, credit scores, etc. It is costly for firms to collect more information on consumers and receive better signals about each consumer's taste.

In the targeted advertising model of Iyer et al. (2005), firms pay a fixed marketing cost to have the ability to target advertise, instead of mass advertising. I extend this by making marketing cost an endogenous function of signal accuracy. I find that when firms choose to mass advertise, each firm chooses not to pay for any information on consumers, giving all firms completely noisy and meaningless signals on each consumer's taste. And when firms choose to target advertise, each firm chooses to pay for the same amount of information on each consumers, giving all firms equally accurate signals on each consumer's taste. This confirms the assumption of Iyer et al. (2005) of a fixed marketing research cost for targeted advertising.

I find that the marketing cost adds to the entry cost, discouraging some firms from entering the market. This decreases consumer welfare because each consumer gets informed about a smaller, less-targeted selection of products. Yet this increases consumer welfare because each consumer is pestered by fewer ads. I investigate the idea that tougher privacy regulations induce lower signal accuracies by making it more expensive for firms to gather information about consumers, instead of directly setting signal accuracy. I find that the equilibrium signal accuracy could be greater or less than the optimal signal accuracy under no or fixed marketing costs, depending on the convexity of the ad annoyance function. This means that how privacy laws restrict the signal accuracy matters. Sometimes it is more optimal for

us to set signal accuracy by picking the information that firms can collect. Other times it is more optimal for us to induce a signal accuracy by changing firms' costs of collecting information.

In section 7, I extend my model to investigate how signal accuracy affects consumer welfare through an endogenous price of ads. For a low quantity and high price of ads, firms target advertise. For an intermediate quantity and cutoff price of ads, firms *mix advertise*, or equivalently firms mix between targeted advertising and mass advertising. Either all firms advertise to some of the consumers not in their targeted segment, or some firms target advertise and some firms mass advertise. For a high quantity and low price of ads, firms will mass advertise. Increasing the signal accuracy, lowers the cutoff price, inducing more targeting. I find that this unambiguously increases consumer welfare. When firms target advertise, an increase in signal accuracy increases the demand for advertising, which increases the price of ads. Higher ad prices have an additional effect on consumer welfare by discouraging firm entry. Therefore consumers forfeit the opportunity to buy those products, decreasing welfare, and avoid the costs of being pestered by those ads, increasing welfare.

In section 8, I extend my model to include ad retention. Consumers don't always see or remember every ad sent to them. In an extension to this, I will allow consumers to choose ad retention through paying for ad avoidance. Consumers may choose to block an increasing number of ads by paying the costs and opportunity costs involved in no-call-or-email lists, spam filters, and driving down roads with fewer billboards. The more that a consumer participates in these programs, the more expensive it becomes and the more ads he blocks. This differs from the ad avoidance tool in the targeted advertising model of Johnson (2013), because he only allows consumers to choose to avoid all ads or not avoid any ads. In another extension, I allow firms to choose to increase ad retention by sending multiple ads to each consumer, similar to Anderson and de Palma (2010).

In section 9, I will consider the case where firms receive the same common signal about each consumer's type to investigate how information sharing affects my model. Although I get a different, more complicated algebraic result, I get mostly the same conceptual result. In section 10, I conclude.

## 2 Literature Review

This paper is most related to Johnson (2013). In Johnson (2013), improving the accuracy of the signal firms get on consumer tastes increases product selection and ad annoyance faced by consumers. The big difference between our models is that in Johnson (2013) this effect is intertwined with an ad avoidance or ad blocking effect. In Johnson (2013), an improvement in signal accuracy encourages firms to advertise to more consumers, which encourages more consumers to block advertisements. I explore ad blocking and compare it to a basic model without ad blocking in 8.4.

This paper is also related to Bergemann and Bonatti (2011). Both our papers analyze how improving the accuracy of the signal consumers get on consumer tastes improves consumer welfare through offering them a better selection of products. The biggest difference between our models is that Bergemann and Bonatti (2011) has no ad annoyance. In his model, consumers are delivered a fixed number of ads. Firms buy these ads through perfect competition or from a monopolist publisher. This induces firms to advertise in fewer markets and increase the price of their products. While the product pricing effects are beyond the scope of this paper,<sup>50</sup> When I allow for endogenous advertisement costs in section 7, I show an additional effect on welfare through product selection (similar to Bergemann and Bonatti, 2011) and through ad annoyance. For a complete discussion on the product pricing effect, refer to Bergemann and Bonatti (2011).

This paper is also related to the preliminary work of Shiman (1997). Although

---

<sup>50</sup> there are plenty of papers that examine how consumer information can be used to price discriminate (see for example: Esteves, 2009; Goldfarb and Tucker, 2011; Villas-Boas, 1999, 2004), and several papers that extend this to targeted advertising (see for example: Galeotti and Moraga-González, 2004; Iyer et al., 2005).

Shiman abandoned the project, because there was very little interest at the time for targeted advertising in economics and the mathematics became too difficult, it is the first paper to explore how improving the accuracy of the signal firms get on consumer tastes increases product selection and ad annoyance faced by consumers. The biggest differences between our papers is in the purchasing decision. In Shiman (1997), a consumer's purchase decision is independent of what other products he bought. In this paper, I only allow a consumer to buy one product. In addition, I am able to explore advertisement prices, marketing costs, and the sharing of information between firms.

Section 6 is also related to Iyer et al. (2005). Both analyze firms decisions to either mass advertise or target advertise.<sup>51</sup> In Iyer et al. (2005), the cost of targeted advertising is some exogenous constant, and in section 6 I extend my model to allow for a marketing cost of gathering information (a better signal). Firms would pay for an equilibrium level of information, if they are using it to target, and not pay for any information, if they are mass advertising. In this way I am making the fixed targeting cost of Iyer et al. (2005) an endogenous constant.

Both Iyer et al. (2005) and this paper find that for a low per-consumer advertising cost or a high marketing cost, firms will mass advertise, and for a high per-consumer advertising cost or a low marketing cost, firms will target advertise.<sup>52</sup> I extend this to show that for a low signal accuracy, firms mass advertise, and for a high signal accuracy, firms mass advertise.

This paper is also related to Esteban et al. (2001). In their paper, the specialization of a magazine or other advertising medium is equivalent to signal accuracy.

---

<sup>51</sup> Iyer et al. (2005) analyzes the case of a duopoly, and I analyze the case of monopolistic competition.

<sup>52</sup> Iyer et al. (2005) also show that for an intermediate per-consumer advertising cost or an intermediate marketing cost, one firm will target advertise and the other will mass advertise. These results depend on the fact that consumers who would be willing to buy either product (the comparison shoppers) have the same value  $r$  for both products. If both firms were to advertise to the comparison shoppers, then firms would both set a product price of zero in Bertrand style price competition.

Consumers reading more specialized magazines are more likely to buy the product. Esteban et al. (2001) show that if a social planner picks product price and degree of advertisement specialization, then targeted advertising is preferable to mass advertising. I get a similar result. I find that if a social planner picks the signal accuracy, then targeted advertising is consumer welfare improving over mass advertising.

Esteban et al. (2001) also show that a firm might choose to specialize its advertising too much. This might make targeted advertising less desirable than mass advertising, because the firm may choose to increase its prices too much.<sup>53</sup> My paper extends this discussion by the addition of ad annoyance costs. I find that if firms are getting too accurate or too inaccurate information on consumers, then targeted advertising might be less desirable than mass advertising, because consumers face additional advertisement annoyance costs.

### 3 Game

There is a sufficiently large number of profit maximizing firms. Each firm may potentially enter the market at an entry cost  $F > 0$  and produce a product at a constant marginal cost normalized to zero. There is a unit mass of utility maximizing consumers. Half of the consumers are type 0 and the other half of the consumers are type 1.

Initially, a firm knows the aggregate distribution of types, but does not know the types of individual consumers. If a firm chooses to enter, then it receives a private, i.i.d. signal about each consumer's type that is true with probability  $\frac{1+\theta}{2}$  and false with a probability  $\frac{1-\theta}{2}$ , where  $\theta \in [0, 1]$ . I interpret  $\theta$  as the signal accuracy. A  $\theta$  of one would be a perfect signal about consumers' types and a  $\theta$  of zero would be a meaningless signal about consumers' types.

Each entrant  $j$  simultaneously chooses its product type (type 0 or a type 1), which consumers it will advertise to, and single product price  $p_j$  to maximize its

---

<sup>53</sup> He gets this result when mass advertising wastes few ads and when product demand is sufficiently inelastic.

own profit  $\Pi_j$ . For each consumer whom it chooses to advertise to, entrant  $j$  pays a constant per consumer advertising cost  $c$ . These choices are observed by all consumers. If firm  $j$  enters, advertises to  $M_j$  consumers, and sells  $Q_j$  units of its good then it gets a profit of  $\Pi_j \equiv p_j Q_j - c M_j - F$ .

Each consumer  $i$  may buy one unit of his choice from the goods advertised to him, or he may choose to take an outside option.<sup>54</sup> The utility consumer  $i$  would get from buying good  $j$  is  $u_{ij} = R + b\lambda_{ij} - p_j + \epsilon_{ij} - A(N)$ , where  $R > 0$  is the benefit from consuming any advertised good,  $b > 0$  is the benefit from buying a good personalized to a consumer's type,  $\lambda_{ij}$  is one if consumer  $i$ 's type matches product  $j$ 's type and zero otherwise,  $p_j$  is the price of good  $j$ ,  $\epsilon_{ij}$  is an i.i.d. stochastic shock to consumer  $i$ 's value for good  $j$  with a c.d.f. of  $F(\epsilon) = e^{-e^{-\epsilon/\mu}}$ , and  $A(N)$  is the ad annoyance suffered by a consumer from seeing ads from  $N$  different firms. I assume that for all  $N \geq 0$ , I have  $A'(N) > 0$ . The utility consumer  $i$  would get from buying the outside option is  $u_{i0} = \epsilon_{i0} - A(N)$ , where  $\epsilon_{i0}$  is an i.i.d. stochastic shock to consumer  $i$ 's value for the outside option with a c.d.f. of  $F(\epsilon) = e^{-e^{-\epsilon/\mu}}$ .

## 4 Equilibrium

In this section, I find a pure strategy monopolistic competitive equilibrium that is symmetric in pricing and advertising with an equal number of firms of each type. I start by finding the equilibrium prices and number of entrants in each market for any given signal accuracy  $\theta \in [0, 1]$ . Then I examine how consumer welfare changes with signal accuracy  $\theta$ .

If entrant  $j$  expects a profit from advertising to a consumer with a signal of type  $l \in \{0, 1\}$ , then entrant  $j$  expects a profit from advertising to each and every consumer with a signal of type  $l$ . Therefore a firm only chooses one of four possibilities: (1) advertise to all consumers (mass advertise), (2) advertise to all consumers with a signal of its type and only some of the other consumers (mixed advertise), (3)

---

<sup>54</sup> This may simply be the option not to buy any good.

advertise to only those consumers with a signal of its type (target advertise), and (4) not advertise to any consumers (not enter the market).

If entrant  $j$  of type  $k \in \{0, 1\}$  expects a profit from advertising to a consumer with a signal of type  $l \in \{0, 1\}$ , then every entrant of type  $k$  expects a profit from advertising to each and every consumer with a signal of type  $l$ , and by symmetry, every entrant of type  $1 - k$  expects a profit from advertising to each and every consumer with a signal of type  $1 - l$ . Therefore I have Proposition 1.

**Proposition 1.** *Either:*

- (a) *all entrants advertise to all consumers (the mass advertising equilibrium),*
- (b) *all entrants target their advertising (the targeted advertising equilibrium), or*
- (c) *entrants are indifferent between mass advertising and targeted advertising (the mixed advertising equilibrium).*

In section 4.1, I explore the mass advertising equilibrium. In section 4.2, I explore the targeted advertising equilibrium. And in 4.3, I show that all firms choose to mass advertise for low  $\theta$  and all firms choose to target advertise for high  $\theta$ . I show that the mixed advertising equilibrium only occurs for the threshold value of  $\theta$ .

## 4.1 Mass Advertising Equilibrium

In this section I find the consumer welfare when all firms mass advertise. I: find the demand for an individual firm (Step 1), find the product price (Step 2), find the equilibrium number  $N$  of entrants in each submarket (Step 3), and find the aggregate consumer welfare (Step 4). I repeat these steps throughout the paper to find consumer welfare for different equilibrium and adaptations of the model.

Step 1. Firm Demand: If firm  $j$  of type  $k \in \{0, 1\}$  advertises to all consumers, then half of the consumers it advertises to will be of type  $k$  and half of the consumers it advertises to will be of type  $1 - k$ . In addition, if all firms mass advertise, then half of the firms that advertise to a consumer will be of his type, and half of the



firms that advertise to a consumer will not be of his type. It is straight-forward to show that if  $2N$  firms enter the market (including firm  $j$ ) and if all firms set a price of  $p$ , firm  $j$  will sell to a share  $e^{\frac{R+b-p_j}{\mu}}/K_{MA}$  of the type  $k$  consumers and to a share  $e^{\frac{R-p_j}{\mu}}/K_{MA}$  of the type  $1-k$  consumers, where  $K_{MA} \equiv 1 + N[e^{\frac{R+b-p}{\mu}} + e^{\frac{R-p}{\mu}}]$ .<sup>55</sup> Therefore firm  $j$ 's quantity  $Q_j$  sold as a function of its price  $p_j$  is given by (1).

$$Q_j = \frac{\frac{1}{2}[e^{\frac{R+b-p_j}{\mu}} + e^{\frac{R-p_j}{\mu}}]}{K_{MA}} \quad (1)$$

**Step 2. Product Price:** In this paper I analyze the case of monopolistic competition. Similar to the basic logit monopolistic competition model presented in Anderson et al. (1992, p. 221-226), this market structure is the limit case where there are so many firms that an individual firm's decisions do not impact the market variable  $K_{MA}$ , or an individual firm does not consider its impact on the the market variable  $K_{MA}$  (as in Dixit and Stiglitz, 1977). This market structure not only makes my model more tractable, but more realistically reflects many advertising markets, including online advertising, billboard advertising, and telemarketing.

Optimizing firm  $j$ 's profit over its price  $p_j$  and using symmetry, I find that each advertiser sets a price of  $p = \mu$  and sells the same quantity  $Q$  of their good. In the same manner I will find  $p = \mu$  for all the equilibria and adaptations of this model. This is not a limitation of the model, but a feature. In this paper, I examine the effects on the advertising side of the market. There are plenty of papers that examine product pricing effects of targeted advertising (see for example: Bergemann and Bonatti, 2011; Galeotti and Moraga-González, 2004; Iyer et al., 2005).

**Step 3. Number of Entrants:** Because firm  $j$  would advertise to all of the consumers, it would pay an advertising cost  $c$ . By free entry, I have the zero profit condition  $F + c = pQ$ , which solving for  $N$  becomes (2).

---

<sup>55</sup> These market shares are found in Anderson et al. (1992, p. 39-40) for a more general framework. My addition is the separation into two market types.

$$N = \frac{\mu}{2(F+c)} - \frac{1}{e^{\frac{R+b-\mu}{\mu}} + e^{\frac{R-\mu}{\mu}}} \equiv N_{MA} \quad (2)$$

Here  $\frac{\mu}{2(F+c)}$  is the number of firms that would enter each sub-market if there were no outside option. It is also the effect of the entry cost  $F$  and the advertising cost  $c$  on the entry. And  $[e^{\frac{R+b-\mu}{\mu}} + e^{\frac{R-\mu}{\mu}}]^{-1}$  is the number of firms discouraged from entering each sub-market due to the outside option.

Step 4. Consumer Welfare: The total consumer welfare is the sum of the aggregate consumer surplus from sales minus the aggregate ad annoyance cost. From Anderson et al. (1992, p. 60-61), I have that the consumer surplus for this type of model (a monopolistic competitive logit model) from the sales of the product is  $CS = \mu \ln K$ . Therefore the consumer welfare  $CW$  is given by (3).

$$CW = R - \mu + \mu \left[ \ln(e^{b/\mu} + 1) + \ln\left(\frac{\mu}{2(F+c)}\right) \right] - A(2N_{MA}) \equiv CW_{MA} \quad (3)$$

Here  $R - \mu + \mu[\ln(e^{b/\mu} + 1) + \ln(\frac{\mu}{2(F+c)})]$  is the aggregate consumer surplus gained from sales, while  $A(2N_{MA})$  is the aggregate ad annoyance cost.<sup>56</sup> Higher entry costs  $F$  or advertising costs  $c$  would hurt consumers through a worse product selection by way of the term  $\mu \ln(\frac{\mu}{2(F+c)})$  and would benefit consumers through less ad annoyance  $A(2N_{MA})$  through less ad annoyance  $A(2N_{MA})$  by way of less entry, see equation (2).

## 4.2 Targeted Advertising Equilibrium

Step 1. Firm Demand: If firm  $j$  of type  $k \in \{0, 1\}$  advertises to only those consumers with a signal of  $k$ , then firm  $j$  will only advertise to half of the consumers:  $\frac{1+\theta}{2}$  of whom will be of type  $k$ , and  $\frac{1-\theta}{2}$  of whom will be of type  $1 - k$ . In addition, if all firms target advertise, then  $\frac{1+\theta}{2}$  of the firms that advertise to a consumer will be of his type, and  $\frac{1-\theta}{2}$  of the firms that advertise to a consumer will be not of his type. It

<sup>56</sup> Note that consumers see ads from both types of firms, so they receive ads from  $2N$  firms.

is straight-forward to show that if  $2N$  firms enter the market and if all firms sets a market price of  $p$ , firm  $j$  will sell to a share  $e^{\frac{R+b-p_j}{\mu}}/K_{TA}$  of the type  $k$  consumers and a share  $e^{\frac{R-p_j}{\mu}}/K_{TA}$  of the type  $1-k$  consumers, where  $K_{TA} \equiv 1 + N[\frac{1+\theta}{2}e^{\frac{R+b-p}{\mu}} + \frac{1-\theta}{2}e^{\frac{R-p}{\mu}}]$ .<sup>57</sup> Therefore firm  $j$ 's quantity  $Q_j$  sold as a function of its price  $p_j$  is given by (4).

$$Q_j = \frac{\frac{1}{2}[\frac{1+\theta}{2}e^{\frac{R+b-p_j}{\mu}} + \frac{1-\theta}{2}e^{\frac{R-p_j}{\mu}}]}{K_{TA}} \quad (4)$$

Note the difference between equations (1) and (4). Under targeted advertising, firm  $j$  would advertise to half the consumers than it would under mass advertising, which explains the differences in the numerators of equations (1) and (4). Yet under targeted advertising, a consumer would only see half of the ads, while under mass advertising, a consumer would see all of the ads. In addition under targeted advertising, a consumer would see  $\frac{1+\theta}{2}$  of its ads from products of its type and  $\frac{1-\theta}{2}$  of its ads from products not of its type. This explains the difference between  $K_{MA}$  and  $K_{TA}$ .

Step 2. Product Price: Similar to section 4.1, by optimizing firm  $j$ 's profit over its price  $p_j$  and using symmetry, I find that each advertiser sets a price of  $p = \mu$  and sells the same quantity  $Q$  of their good.

Step 3. Number of Entrants: Because firm  $j$  would advertise to half of the consumers, it would pay an advertising cost  $c/2$ . By free entry, I have the zero profit condition  $F + c/2 = pQ$ , which solving for  $N$  becomes (5).

$$N = \frac{\mu}{2(F + c/2)} - \frac{1}{\frac{1+\theta}{2}e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2}e^{\frac{R-\mu}{\mu}}} \equiv N_{TA}(\theta) \quad (5)$$

Here  $\frac{\mu}{2(F+c/2)}$  is the number of firms that would enter each sub-market if there were no outside option. It is also the effect of the entry cost  $F$  and the advertising cost  $c/2$  on the entry. And  $[\frac{1+\theta}{2}e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2}e^{\frac{R-\mu}{\mu}}]^{-1}$  is the number of firms discouraged from entering each sub-market due to the outside option. Note that, unlike (2),

<sup>57</sup> These market shares are found in Anderson et al. (1992, p. 39-40) for a more general framework.

this term depends on the signal accuracy  $\theta$ , because here firms are using their signals to decide which consumers to show their ads. Coincidentally the fact that  $\theta$  only affects entry through the outside option is a result of the form of the demand function. Also note that the first and second derivative of (5) uphold the finding in Shiman (1997) that increasing the signal accuracy induces firms to advertise more at a decreasing rate.

Step 4. Consumer Welfare: The total consumer welfare is the sum of the aggregate consumer surplus from sales minus the aggregate ad annoyance cost. From Anderson et al. (1992, p. 60-61), I have that the consumer surplus for this type of model (a monopolistic competitive logit model) is  $CS = \mu \ln K$ . Therefore the consumer welfare  $CW$  is given by (6).

$$CW = R - \mu + \mu \left[ \ln\left(\frac{1+\theta}{2} e^{b/\mu} + \frac{1-\theta}{2}\right) + \ln\left(\frac{\mu}{2(F+c/2)}\right) \right] - A(N_{TA}) \equiv CW_{TA}(\theta) \quad (6)$$

Here  $R - \mu + \mu \left[ \ln\left(\frac{1+\theta}{2} e^{b/\mu} + \frac{1-\theta}{2}\right) + \ln\left(\frac{\mu}{2(F+c/2)}\right) \right]$  is the aggregate consumer surplus gained from sales, while  $A(N_{TA})$  is the aggregate ad annoyance cost. In the mass advertising equilibrium, the signal accuracy  $\theta$  doesn't matter, because firms advertise to all consumers anyway. Here, in the targeted advertising equilibrium, the signal accuracy  $\theta$  does matter, because firms only advertise to those consumers with signals that match their product characteristic. Higher signal accuracy  $\theta$  would benefit consumers through better product selection by way of the term  $\mu \ln\left(\frac{1+\theta}{2} e^{b/\mu} + \frac{1-\theta}{2}\right)$  and would hurt consumers through higher ad annoyance  $A(N_{TA})$  by way of more entry, see equation (5).

### 4.3 Equilibrium Determination Condition

In this section, I show that for most signal accuracies  $\theta$  there is only one possible equilibrium. I show that there exists a threshold signal accuracy  $\widehat{\theta}$  such that for  $\theta < \widehat{\theta}$  I have the mass advertising equilibrium and for  $\theta > \widehat{\theta}$  I have a targeted

advertising equilibrium.

Under the mass advertising equilibrium, a firm will sell to some consumers not of its signal type. This quantity is given by (7).

$$q_{MA} \equiv \frac{\frac{1}{2} \left[ \frac{1-\theta}{2} e^{\frac{R+b-p}{\mu}} + \frac{1+\theta}{2} e^{\frac{R-p}{\mu}} \right]}{K_{MA}} \quad (7)$$

A firm must be profiting on these consumers, in order to choose mass advertising. Therefore I have  $c/2 \leq pq_{MA}$ , which reduces to  $\theta \leq \widehat{\theta}$  where  $\widehat{\theta}$  is given by (8).

$$\widehat{\theta} \equiv \frac{F}{F+c} \frac{e^{b/\mu} + 1}{e^{b/\mu} - 1} \quad (8)$$

Under the targeted advertising equilibrium, a firm could potentially sell to consumers not of its signal type. This quantity is given by (9).

$$q_{TA} \equiv \frac{\frac{1}{2} \left[ \frac{1-\theta}{2} e^{\frac{R+b-p}{\mu}} + \frac{1+\theta}{2} e^{\frac{R-p}{\mu}} \right]}{K_{TA}} \quad (9)$$

A firm must not be able to profit on these consumers, in order to choose targeted advertising.

Therefore I have  $c/2 \geq pq_{TA}$ , which reduces to  $\theta \geq \widehat{\theta}$  where  $\widehat{\theta}$  is given by (8).

Therefore  $\widehat{\theta}$  is the minimum signal accuracy  $\theta$  needed for targeted advertising to be profitable. When  $\theta < \widehat{\theta}$  I have the mass advertising equilibrium and when  $\theta > \widehat{\theta}$  I have the targeted advertising equilibrium. When  $\theta = \widehat{\theta}$ , I can have either the mass advertising equilibrium, the targeted advertising equilibrium, or the mixed advertising equilibrium. The mixed equilibrium is only feasible when  $\theta = \widehat{\theta}$ , because firm  $j$  with a product characteristic  $k$  needs to be indifferent between advertising and not advertising to consumers with signal  $1 - k$ .

From (8), I have that the threshold signal accuracy  $\widehat{\theta}$  increases with a higher entry cost  $F$ , a lower per-person advertising cost  $c$ , a lower benefit  $b$  from consuming

a product that matches your type, and a lower variation  $\mu$  in consumers' tastes. I interpret this as if entry costs are high enough, advertising is cheap enough, the benefit from buying a good matching a consumer's type is low enough, and consumers' tastes are varied enough, then an entrant might as well sell to the less profitable group of consumers, those consumers with signals that don't match its product type.

Note that choosing whether or not to target advertise does not change the market variable  $K_{MA}$  in (7) and the market variable  $K_{TA}$  in (9), because I am analyzing the case of monopolistic competition (see page 80). This extends the standard logit monopolistic competition assumption that there are so many firms that an individual firm's pricing decision does not impact the market variable  $K$  (as in Anderson et al., 1992, p. 221-226), or an individual firm does not consider the impact of its pricing on the the market variable  $K$  (as in Dixit and Stiglitz, 1977), to the firms' advertising decisions. This makes the monopolistic competition framework particularly desirable for logit demand models, such as my model.

## 5 Impact of Signal Accuracy on Welfare

In this section I consider how changing the signal accuracy  $\theta$  affects the equilibrium found in section 4. In particular I am concerned with finding the effect of  $\theta$  on consumer welfare  $CW$ , because I interpret stricter privacy laws as creating noisier signals about consumers' tastes. This section would be useful in determining how much we should protect the privacy of personal information to maximize a consumer welfare.

I start by analyzing the welfare effect of inducing the switch between mass and targeted advertising, through changing signal accuracy  $\theta$  (see section 5.1). Then I analyze the welfare effects of changing signal accuracy  $\theta$  under targeted advertising (see section 5.2). When firms mass advertise, there is no welfare effect of changing  $\theta$ , because firms are not using the signal on consumer's tastes.

## 5.1 Effect near the Threshold

In 4.3, I showed how signal accuracy  $\theta$  impacts whether we are in the mass advertising equilibrium or a targeted advertising equilibrium. I showed that for  $\theta < \widehat{\theta}$  I have the mass advertising equilibrium and for  $\theta > \widehat{\theta}$  I have a targeted advertising equilibrium, where  $\widehat{\theta}$  is given by (8). Here I show that when  $\widehat{\theta} \leq 1$ , there exists a targeted advertising equilibrium consumer welfare improving to the mass advertising equilibrium.<sup>58</sup>

I do this by considering the targeted advertising equilibrium for  $\theta = \widehat{\theta}$ . This is the case where entrants are indifferent between targeting their advertisements and not. Putting  $\theta = \widehat{\theta}$  into (5), I have the relation between the number  $2N_{TA}$  of entrants under targeted advertising and the number  $2N_{MA}$  of entrants under mass advertising, given by (10).

$$N_{TA}(\widehat{\theta}) = \frac{F + c}{F + c/2} N_{MA} \quad (10)$$

Putting  $\theta = \widehat{\theta}$  into (6), I have that the aggregate consumer surplus from sales under targeted advertising is equal to the aggregate consumer surplus from sales under mass advertising. I find that the initial cost of being offered fewer products equals the initial benefit from being offered better matching products, because the signal accuracy is so low. Therefore I have that the consumer welfare  $CW_{TA}$  from targeted advertising is greater than the consumer welfare  $CW_{MA}$  from mass advertising, as shown in (11).

$$\begin{aligned} CW_{TA}(\widehat{\theta}) &= R - \mu + \mu \left[ \ln(e^{b/\mu} + 1) + \ln\left(\frac{\mu}{2(F + c)}\right) \right] - A\left(\frac{F + c}{F + c/2} N_{MA}\right) \\ &> R - \mu + \mu \left[ \ln(e^{b/\mu} + 1) + \ln\left(\frac{\mu}{2(F + c)}\right) \right] - A(2N_{MA}) = CW_{MA} \end{aligned} \quad (11)$$

---

<sup>58</sup> When  $\widehat{\theta} > 1$ , only the mass advertising equilibrium is possible.

Therefore I have Proposition 2.

**Proposition 2.** *As long as targeted advertising is possible (i.e.  $\widehat{\theta} \leq 1$ ), there exists at least one value of  $\theta$  that induces firms to target advertise and makes consumers better off than mass advertising.*

This result is similar to the result in Esteban et al. (2001) that if a social planner picks product price and degree of advertisement specialization, then targeted advertising is welfare improving to mass advertising. Here I show that if a social planner picks the signal accuracy, then it can induce a preferable targeted advertising equilibrium to mass advertising.

## 5.2 Effect on the Targeted Advertising Equilibrium

In 5.1, I showed that targeted advertising (for some values of  $\theta$ ) would be consumer welfare improving to mass advertising. Here I consider how changing the signal accuracy  $\theta$  affects the targeted advertising equilibrium. Because I interpret stricter privacy laws as creating noisier signals about consumers' tastes, this subsection would be useful in determining how much we should protect the privacy of personal information, while still allowing firms to collect enough information to target their advertisements.

Differentiating the consumer welfare under the targeted advertising equilibrium (given by (6)) by the signal accuracy  $\theta$ , I have (12).

$$CW_{TA}'(\theta) = N'_{TA}(\theta) \left[ \mu \left( \frac{1+\theta}{2} e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2} e^{\frac{R-\mu}{\mu}} \right) - A'(N_{TA}(\theta)) \right] \quad (12)$$

$$\text{where } N'_{TA}(\theta) = \frac{\frac{1}{2}(e^{\frac{R+b-\mu}{\mu}} - e^{\frac{R-\mu}{\mu}})}{\left( \frac{1+\theta}{2} e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2} e^{\frac{R-\mu}{\mu}} \right)^2} > 0$$

Here  $N'_{TA}(\theta) \mu \left( \frac{1+\theta}{2} e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2} e^{\frac{R-\mu}{\mu}} \right) > 0$  is the aggregate consumer surplus gained from increasing  $\theta$  through more goods being offered to consumers (from a higher



number of firms) and through an increase in the chance of products matching consumers' tastes (from more accurate signals), and  $N'_{TA}(\theta)A'(N_{TA}(\theta)) > 0$  is the aggregate ad annoyance gained from increasing  $\theta$  through a higher number of advertisers. These two forces, the product selection effect and the ad annoyance effect, can make increasing signal accuracy  $\theta$  either increase or decrease total consumer welfare. Furthermore by taking the derivative of (12) and evaluating it when it is equal to zero, I have Proposition 3.

**Proposition 3.** *If  $\theta^{opt}$  solves  $CW_{TA}'(\theta^{opt}) = 0$  and:*

(a) *if  $A''(N_{TA}(\theta^{opt})) > \mu(\frac{1+\theta^{opt}}{2}e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta^{opt}}{2}e^{\frac{R-\mu}{\mu}})$*   
*then  $\theta^{opt}$  is a local maximum of  $CW_{TA}(\theta)$*

(b) *if  $A''(N_{TA}(\theta^{opt})) < \mu(\frac{1+\theta^{opt}}{2}e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta^{opt}}{2}e^{\frac{R-\mu}{\mu}})$*   
*then  $\theta^{opt}$  is a local minimum of  $CW_{TA}(\theta)$*

Proposition 3 shows how the shape of the advertising annoyance function influences the shape of the consumer welfare function  $CW_{TA}$ . If the ad annoyance function  $A(N)$  satisfies (a) for critical value(s) in  $[\widehat{\theta}, 1]$ , then the consumer welfare function is an inverted-u-shape (see the example given in 5.2.2). Then consumers would be better off with firms getting somewhat noisy information about the consumer characteristic (i.e.  $\theta \in (\widehat{\theta}, 1)$ ). Yet if the ad annoyance function  $A(N)$  satisfies (b) for critical value(s) in  $[\widehat{\theta}, 1]$ , then the consumer welfare function is a u-shape (see the example given in 5.2.1). Then consumers would be better off with firms getting either an accurate signal (i.e.  $\theta = 1$ ) or a noisy signal barely accurate enough to encourage targeted advertising (i.e.  $\theta = \widehat{\theta}$ ).

### 5.2.1 Example: Linear Ad Annoyance

For example, suppose the ad annoyance function were of the form  $A(N) = aN$ , where  $a > 0$  is the additional annoyance cost to a consumer per ad. By Proposition 3, any  $\theta^{opt}$  that solves  $CW_{TA}'(\theta^{opt}) = 0$  would be a minimum. Therefore  $CW_{TA}(\theta)$  is

a u-shape. Therefore consumers would be better off with firms getting either an accurate signal (i.e.  $\theta = 1$ ) or a noisy signal barely accurate enough to encourage targeted advertising (i.e.  $\theta = \widehat{\theta}$ ). Comparing  $CW_{TA}(1)$  to  $CW_{TA}(\widehat{\theta})$  using (6), I have Corollary 1.

**Corollary 1.** *If ad annoyance is linear (i.e.  $A(N) = aN$ ), if targeted advertising is possible (i.e.  $\widehat{\theta} \leq 1$ ), and:*

(a) *if  $a > \widehat{a}$  then the consumer welfare maximizing  $\theta$  is one,*

(b) *if  $a < \widehat{a}$  then the consumer welfare maximizing  $\theta$  is  $\widehat{\theta}$ ,*

where  $\widehat{a} \equiv \mu e^{\frac{R+b-\mu}{\mu}} (e^{b/\mu} + 1) \ln[\frac{e^{b/\mu}+1}{e^{b/\mu}} \frac{F+c/2}{F+c}]$ .

### 5.2.2 Example: Quadratic Ad Annoyance

For example, suppose the ad annoyance function were of the form  $A(N) = aN^2$ , where  $a > 0$  is some ad annoyance parameter. Then Proposition 3 shows that critical values could be either local minima or local maxima. Under sufficient and reasonable parameter restrictions given by Corollary 2, a quadratic ad annoyance function has an interior optimum.

**Corollary 2.** *If ad annoyance is quadratic, if targeted advertising is possible (i.e.  $\widehat{\theta} \leq 1$ ), and if  $a \in (a_1, a_2)$ , then the consumer welfare maximizing  $\theta \in (\widehat{\theta}, 1)$ , for some  $a_1, a_2 > 0$ .*

*Proof.* See Appendix B. □

## 6 Marketing Costs

In this section, I consider the case where firms can pay to get a more accurate signal  $\theta$  about consumers' tastes. Here, each firm may potentially enter the market at an entry cost  $f + M(\theta)$ , where  $f > 0$  is the fixed entry cost and  $M(\theta)$  is the cost of getting a signal accuracy of  $\theta$ . I interpret  $M(\theta)$  as the marketing research cost.

I normalize  $M(0) = 0$ . In addition, I assume that (A1)  $M$  is twice differentiable and continuous, (A2)  $M'(0) = 0$ , (A3)  $M'(\theta) > 0$  for all  $\theta \in (0, 1]$ , and (A4)  $M''(\theta) > 0$  for all  $\theta \in [0, 1]$ . I interpret assumptions (A2) to (A4) as a diminishing marginal effect of marketing research. As a firm increases the accuracy of the signal  $\theta$ , additional accuracy gets more expensive.

Note that I have intentionally used different notation for the fixed entry cost  $f$  (in this section) and the entry cost  $F$  (in sections 3 through 5), because there is no reason for  $f$  to equal  $F$ . Some of marketing costs may be considered a component of  $F$  in the basic model. In 6.4, I will use this distinction when I consider what type of privacy restrictions a government should impose: marketing costs or data restrictions.

## 6.1 Equilibrium

The mass advertising equilibrium with marketing costs is the same as the mass advertising equilibrium found in section 4.1 with an entry cost of  $F = f$  and a signal accuracy  $\theta = 0$ , because firms would not pay for information about consumers' tastes if they are going to advertise to all consumers.

The targeted advertising equilibrium with marketing costs is the same as the targeted advertising equilibrium found in section 4.2 with an entry cost of  $F = f + M(\theta^*)$  and a signal accuracy  $\theta = \theta^*$ , where  $\theta^*$  solves a firm's first order condition for  $\theta$ , which reduces to (13).<sup>59</sup>

$$v(\theta^*) = \theta^* + \frac{e^{b/\mu} + 1}{e^{b/\mu} - 1} \quad (13)$$

$$\text{where } v(\theta) \equiv \frac{f + M(\theta) + c/2}{M'(\theta)}$$

Here  $v(\theta)$  is the inverse hazard of the targeted advertising total cost function

---

<sup>59</sup> When the  $\theta^*$  that solves (13) is greater than one, then firms would choose the profit maximizing  $\theta = 1$ .

$f + M(\theta) + c/2$ . My assumptions on the function  $M$ , guarantee a unique value of  $\theta^*$ .

From (13), I have  $\theta^*$  increases with a higher fixed entry cost  $f$  and a higher per-person advertising cost  $c$ . I interpret this as when fewer firms enter the market because of a higher fixed entry cost  $f$  or a higher advertising cost  $c$ , each firm will benefit more from better information about consumers' tastes. In addition, I have  $\theta^*$  increases with a higher benefit  $b$  from consuming a product that matches your type and a lower variation  $\mu$  in consumers' tastes. I interpret this as when each consumer is more likely to prefer a good designed for his type, then firms will benefit more from better information about consumers' tastes.

## 6.2 Equilibrium Determination Condition

For most targeted advertising equilibrium signal accuracies  $\theta^*$  there is only one possible equilibrium. Similar to the  $\widehat{\theta}$  found section 4.3, there exists a threshold  $\tilde{\theta}$  such that for  $\theta^* < \tilde{\theta}$  I have the mass advertising equilibrium and for  $\theta^* > \tilde{\theta}$  I have the targeted advertising equilibrium, where  $\theta^*$  solves (13). Note that the big difference here is that  $\theta^*$  is now an endogenous constant.<sup>60</sup>

$$\tilde{\theta} \equiv \frac{f + 2M(\tilde{\theta}) e^{b/\mu} + 1}{f + c} \frac{e^{b/\mu} - 1}{e^{b/\mu} - 1} \quad (14)$$

Similar to the effect of the entry cost  $F$  on  $\widehat{\theta}$ , increasing the fixed entry cost  $f$  increases  $\tilde{\theta}$ . I interpret this as if fixed entry costs are high enough, then enough firms will be deterred from entering so that an entrant might as well sell to the less profitable group of consumers, those consumers with signals that don't match its product characteristic. Unlike their effects on  $\widehat{\theta}$ , the effects of the advertising cost  $c$ , the benefit  $b$  from consuming a product that matches your type, and the variation  $\mu$  in consumers' tastes on  $\tilde{\theta}$  is ambiguous. Decreasing the advertising cost  $c$ , increasing the benefit  $b$  from consuming a product that matches your type, and decreasing the variation  $\mu$  in consumers' tastes both directly increases  $\tilde{\theta}$  as

---

<sup>60</sup> See Appendix A for the derivation of (14).

shown in (14) and indirectly decreases  $\tilde{\theta}$  through decreasing  $\theta^*$  as shown in (13). See Appendix A for details.

### 6.3 Impact of Marketing Costs on Welfare

Under the mass advertising equilibrium, the impact of changing the function  $M$  on the consumer welfare  $CW_{MA}$  is zero, because firm would not choose the minimal signal accuracy  $\theta = 0$  and I have restricted  $M(0) = 0$ .

Under the targeted advertising equilibrium, increasing the function  $M$  would effect total consumer welfare  $CW_{TA}$  in three ways (see (6)): (1) a negative product selection effect (through increasing the fixed cost  $F = f + M(\theta)$ ), (2) a positive ad annoyance effect (through decreasing the number of firms  $N_{TA}$  advertising to a consumer), and (3) a possibly negative or positive indirect effect by altering the firms' choice of  $\theta$  (see (12), its following discussion, and (13)).

In this paper, I consider the case where marketing cost is of the form  $M(\theta) = m\theta^\sigma$ , where  $\sigma > 1$  to satisfy the condition  $M''(\theta) > 0$ , and where  $m > 0$ . I interpret  $m$  as the frictional cost associated with gathering a more accurate signal. Then by the equilibrium condition given by (13), I have (15).

$$\frac{d\theta^*}{dm} = -\frac{f + c/2}{f + m\theta^{*\sigma} + c/2} \frac{\theta^*}{m(\sigma - 1)} < 0 \quad (15)$$

I interpret (15) as increasing the frictional cost  $m$  of gathering a more accurate signal about consumers' tastes induces firms to invest less in signal accuracy.

At first glance, it might appear that increasing the fixed entry cost  $f$  or the advertising cost  $c$  would decrease  $\frac{\partial \theta^*}{\partial m}$ . This is usually the case. Yet  $\theta^*$  endogenously depends on  $f$  and  $c$ , so I have Proposition 4.

**Proposition 4.** *There exists  $\bar{\sigma}, \bar{\theta} > 0$ , such that:*

- (a) *if either  $\sigma > \bar{\sigma}$  or  $\theta^* > \bar{\theta}$ , then  $\frac{d\theta^*}{dm}$  is increasing in both  $f$  and  $c$*
- (b) *if both  $\sigma < \bar{\sigma}$  and  $\theta^* < \bar{\theta}$ , then  $\frac{d\theta^*}{dm}$  is decreasing in both  $f$  and  $c$*

*Proof.* See Appendix B. □

Differentiating my total consumer welfare  $CW_{TA}$  under the targeted advertising equilibrium by the frictional cost  $m$  of a better signal, I have (16).

$$CW_{TA}'(m) = \frac{d\theta^*}{dm} CW_{TA}'(\theta^*) - \phi(\theta^*)(2(f + m\theta^{*\sigma} + c/2) - A'(N_{TA})) \quad (16)$$

$$\text{where } \phi(\theta) \equiv \frac{\mu\theta^\sigma}{2(f + m\theta^\sigma + c/2)^2} > 0$$

Here  $CW_{TA}'(\theta^*)$  is the effect of  $\theta$  on consumer welfare found for the basic model evaluated at  $\theta = \theta^*$  (see (12)). From (16), I have Proposition 5.

**Proposition 5.** *Under targeted advertising, tougher privacy regulations (i.e. increasing the frictional marketing cost  $m$ ) increases the marketing costs of gathering consumer information about consumer tastes, which discourages firms from entering the market. This decreases consumer welfare by giving each consumer a smaller and less targeted selection of products to buy from. And this increases consumer welfare by giving each consumer a smaller number of annoying ads.*

*Proof.* Additional marketing costs from raising  $\theta$  induce fewer firms to enter the market, which decreases the product selection benefit by  $\phi(\theta^*) \cdot 2(f + m\theta^{*\sigma} + c/2) > 0$  and the ad annoyance cost by  $\phi(\theta^*)A'(N_{TA})$ . □

## 6.4 Should We Use Marketing Cost or Data Restrictions?

Privacy regulations can either change signal accuracy  $\theta$  through limiting the data firms can collect on consumers or change  $m$  through making it more costly for firms to gather information. In this section, I explore a consumer welfare maximizing government's choice between these two kinds of privacy regulations.

**Proposition 6.** *If a consumer welfare maximizing government controls entry cost  $f$  (or  $F$ ) and firms target advertise in both equilibria, then data restrictions (choosing  $\theta$ ) is equivalent to increasing marketing costs (choosing  $m$ ).*

*Proof.* We could induce firms to choose any desired  $\theta^*$  through choosing  $m$  and adjusting entry costs.  $\square$

Proposition 6 gives us conditions for when it does not matter whether a government chooses marketing costs or data restrictions. Note that our equilibrium determination condition  $\tilde{\theta} > \widehat{\theta}$ . Therefore firms require less restrictive privacy regulations to choose targeted advertising under marketing costs, but when all firms choose targeted advertising, the equilibria are equivalent.

Let's assume that the government cannot control entry cost  $f$  (or  $F$ ). Then comparing the two models gets difficult, because it depends on how  $f$  is related to  $F$ . I consider two cases: (1) both fixed costs are equivalent (i.e.  $F = f$ ) and (2) marketing costs are part of the existing fixed entry cost (i.e.  $F = f + M(\theta^*)$ ). (1) is similar to Iyer et al. (2005), who assumed that there was an additional (exogenous) fixed marketing cost. In this paper, I find the endogenously derived the fixed marketing cost.

**Proposition 7.** *If both fixed costs are equivalent (i.e.  $F = f$ ), then:*

- (a)  $\tilde{\theta} > \widehat{\theta}$
- (b) *marketing costs lower the product selection benefit*
- (c) *marketing costs lower the ad annoyance costs*

*Proof.* This follows directly from (8), (14), and 5.  $\square$

Proposition 7 tells us that when both fixed costs are equivalent (i.e.  $F = f$ ), the effect of marketing costs on welfare is ambiguous. Marketing costs discourage more firms from entering, so consumer welfare may improve through a reduction

of ad annoyance and possibly better signal accuracy (through an endogenous  $\theta$ ). Yet consumer welfare would also be hurt by a worse product selection (through fewer entrants), possibly a worse signal accuracy (through an endogenous  $\theta$ ), and a bigger incentive to mass advertise. These results vary depending on the ad annoyance function  $A$ .

**Proposition 8.** *If under data restricting privacy regulations consumer welfare is maximized at an interior  $\theta = \theta^{opt}$ , if under marketing cost restricting privacy regulations consumer welfare is maximized at an interior  $m = m^{opt}$ , if marketing costs are part of the existing fixed entry cost (i.e.  $F = f + M(\theta^*(m^{opt}))$ ), and:*

(a) if  $\frac{\mu}{2(F+c/2)} > \frac{1}{\frac{1+\theta^*}{2}e^{\frac{R+b+\mu}{\mu}} + \frac{1-\theta^*}{2}e^{\frac{R+\mu}{\mu}}}$  then  $\theta^{opt} > \theta^*(m^{opt})$

(b) if  $\frac{\mu}{2(F+c/2)} < \frac{1}{\frac{1+\theta^*}{2}e^{\frac{R+b+\mu}{\mu}} + \frac{1-\theta^*}{2}e^{\frac{R+\mu}{\mu}}}$  then  $\theta^{opt} < \theta^*(m^{opt})$

*Proof.* This follows directly from evaluating (16) at the  $m$  such that  $\theta^* = \theta^{opt}$ . By (12),  $\mu\left(\frac{1+\theta^*}{2}e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta^*}{2}e^{\frac{R-\mu}{\mu}}\right) = A'(N_{TA}(\theta^*))$ .  $\square$

Recall that  $\frac{\mu}{2(F+c/2)}$  is the number of firms that would enter each sub-market if there were no outside option. And that  $[\frac{1+\theta}{2}e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2}e^{\frac{R-\mu}{\mu}}]^{-1}$  is the number of firms discouraged from entering each sub-market due to the outside option. Therefore I interpret Proposition 8 as the optimal signal accuracy  $\theta^{opt}$  will be lower than the equilibrium signal accuracy  $\theta^*(m^{opt})$  when the outside option and signal accuracy plays a big role in discouraging firms from entering the market.

## 7 Endogenous Advertising Prices

In the previous sections, I considered a perfectly elastic supply of ad spaces, or in other words, there is a constant per consumer advertising cost  $c$ . This happens when the product market is much smaller than the advertising market. For example, the product market for shampoo is relatively small compared to the television



advertising market, so the quantity of shampoo television ads has a negligible impact on the price of television advertisements.

In this section, I consider a more general supply of ads. After firms enter the market, they buy ad spaces in a perfectly competitive advertising market. For each consumer  $i$  and each possible advertising cost  $c$ , ad suppliers (like newspapers, television channels, and online search engines) supply  $S_i(c)$  ad spaces. I assume that this is symmetric across consumers, so  $S_i(c) = S(c)$ . And I assume  $S'(c) \geq 0$  for all  $c$ .

## 7.1 Equilibrium

Each individual firm takes the equilibrium advertising price  $c = c^*$  as a given constant. Therefore the mass advertising equilibrium and the targeted advertising equilibrium would be the same as the basic model (found in 4.1 and 4.2) evaluated at the equilibrium advertising price  $c = c^*$ . We are already given the per consumer supply  $S(c)$  of ad spaces. Therefore in order to solve the mass advertising equilibrium and the targeted advertising equilibrium, I derive the per consumer demands,  $D_TA(c)$  and  $D_MA(c)$ , of ad spaces.  $D_TA(c)$  and  $D_MA(c)$  are found in a similar fashion to the equilibrium number  $N$  of entrants in each submarket (Step 3).

I find the equilibrium where ad prices are symmetric across consumers and submarkets. Therefore each consumer receives the same number of ads.<sup>61</sup>

Each firm takes  $c$  as the given ad price for all consumers. In equilibrium, the advertising price  $c$  would give firms make zero profit. If  $c$  were too low, firms would make a profit, which would induce firms to enter the market (therefore raising  $c$ ) until they make zero profit. If  $c$  were too high, firms would make a loss, which would induce firms to leave the market (therefore lowering  $c$ ) until they make zero profit.

Under mass advertising,  $2N_{MA}(c)$  firms would advertise to all of the consumers.

---

<sup>61</sup> If there were asymmetry in the ad prices, then firms could take that as an additional signal of a consumer's type. By making this assumption, I avoid considering these types of possible equilibria.

Therefore by (2) the demand  $D_{MA}(c)$  for ad spaces would be given by (17). Under targeted advertising  $N_{TA}(c)$  firms would advertise to half of the consumers, and another  $N_{TA}(c)$  firms would advertise to the other half of the consumers. Therefore by (5) the demand  $D_{TA}(c)$  for ad spaces would be given by (18).

$$D_{MA}(c) = 2 \left[ \frac{\mu}{2(F+c)} - \frac{1}{e^{\frac{R+b-\mu}{\mu}} + e^{\frac{R-\mu}{\mu}}} \right] \quad (17)$$

$$D_{TA}(c) = \frac{\mu}{2(F+c/2)} - \frac{1}{\frac{1+\theta}{2} e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2} e^{\frac{R-\mu}{\mu}}} \quad (18)$$

Note that  $D_{MA}(c)$  and  $D_{TA}(c)$  are both decreasing in the advertising price  $c$  and in entry cost  $F$ . I interpret this as when advertising or entry gets more expensive fewer firms enter the market, which creates a lower quantity demanded of ads.

## 7.2 Equilibrium Determination Condition

In this section, I find a threshold  $\widehat{c}$  such that a threshold ad price  $\widehat{c}$  such that: for  $c < \widehat{c}$ , I have the mass advertising equilibrium; for  $c = \widehat{c}$ , I have the mixed advertising equilibrium; and for  $c > \widehat{c}$ , I have the targeted advertising equilibrium. In doing so, I find one general demand  $D(c)$  for ad spaces. This is similar to the  $\widehat{\theta}$  found section 4.3, with the exception that now mixed advertising is a real possibility, because  $c$  is endogenous.

Under the mass advertising equilibrium, a firm  $j$  with a product characteristic  $k$  must expect a profit from advertising to consumers with signal  $1 - k$ . By the same reasoning in section 4.3, I have the condition  $c/2 \leq pq_{MA}$ , which reduces to  $c \leq \widehat{c}$  where  $\widehat{c}$  is given by (19).

$$\widehat{c} \equiv F \left( \frac{1}{\theta} \frac{e^{b/\mu} + 1}{e^{b/\mu} - 1} - 1 \right) \quad (19)$$

Similarly under the targeted advertising equilibrium, a firm  $j$  with a product

characteristic  $k$  must expect a loss from advertising to consumers with signal  $1 - k$ . Therefore I have the condition  $c/2 \geq pq_{TA}$ , which reduces to  $c \geq \widehat{c}$  where  $\widehat{c}$  is given by (19).

I interpret  $\widehat{c}$  as the lowest advertising price such that firms would not make more money by mass advertising.

Note that while  $D_{MA}(c)$  is not necessarily always greater than  $D_{TA}(c)$  for all values of  $c$ , I have that  $D_{MA}(\widehat{c}) > D_{TA}(\widehat{c})$  by (10). Therefore I have that the inverse demand  $D^{-1}$  for advertisements as a function of the number  $n$  of ads is given by (20).

$$D^{-1}(n) = \begin{cases} D_{TA}^{-1}(n) & \text{if } n \leq D_{TA}(\widehat{c}) & \text{(targeted advertising)} \\ \widehat{c} & \text{if } D_{TA}(\widehat{c}) \leq n \leq D_{MA}(\widehat{c}) & \text{(mixed advertising)} \\ D_{MA}^{-1}(n) & \text{if } n \geq D_{MA}(\widehat{c}) & \text{(mass advertising)} \end{cases} \quad (20)$$

Therefore I have Proposition 9.

**Proposition 9.** *For a low supply  $S(c)$  of ads, firms will target advertise. For an intermediate supply of ads, firms will mixed advertise. And for a high supply of ads, firms will mass advertising.*

### 7.3 Impact of Signal Accuracy on Welfare

In this section I consider how changing the signal accuracy  $\theta$  affects the equilibrium found in section 7.1. In particular I am concerned with finding the effect of  $\theta$  on consumer welfare  $CW$ , because I interpret stricter privacy laws as creating noisier signals about each consumer's taste. This section differs from the discussion in section 5 by the addition of a general supply  $S(p)$  of ads.

Similar to 5, I have that changing  $\theta$  does not change the mass advertising equilibrium, and therefore does not change  $D_{MA}(c)$ . Therefore in this section I focus on how  $\theta$  affects the mixed and the targeted advertising equilibria.

**Proposition 10.** *Under the mixed advertising equilibrium with endogenous ad prices, consumer welfare is strictly increasing in  $\theta$ , and ad price  $c = \widehat{c}$  is strictly decreasing in  $\theta$ .*

*Proof.* Under the mixed advertising equilibrium (for an intermediate supply of ads), I have that  $c = \widehat{c}$  or  $\theta = \hat{\theta}$ . Recall putting  $\theta = \hat{\theta}$  into (6), I have that the aggregate consumer surplus from sales under targeted and mass advertising are equivalent. Similarly setting  $c = \widehat{c}$ , I have that the aggregate consumer surplus from sales under targeted, mixed, and mass advertising are equivalent and equal to  $R - \mu + \mu[\ln(e^{b/\mu} + 1) + \ln(\frac{\mu}{2(F+\widehat{c})})]$ . By (19), I have that  $\widehat{c}$  is decreasing in  $\theta$ . Therefore under the mixed advertising equilibrium, I have that total consumer welfare is strictly increasing in  $\theta$ .  $\square$

Therefore, under the mixed advertising equilibrium with endogenous ad prices, a consumer welfare maximizing government should increase  $\theta$  by relaxing the privacy regulations.

Under the targeted advertising equilibrium (for a low supply of ads), I have that the advertising price  $c$  increases with  $\theta$  through the market clearing condition that the supply  $S(c)$  of ads equals the demand  $D_{TA}(c)$  of ads. Therefore by (6) and (18), I have (21).

$$CW_{TA}'(\theta) = CW_{TA}'(\theta|c = c^*) - \frac{\partial N_{TA}}{\partial \theta} \tau(c^*) [2(F + c^*/2) - A'(N_{TA}(\theta))] \quad (21)$$

$$\text{where } \frac{\partial N_{TA}}{\partial \theta} = \frac{\frac{1}{2}(e^{\frac{R+b-\mu}{\mu}} - e^{\frac{R-\mu}{\mu}})}{\left(\frac{1+\theta}{2}e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2}e^{\frac{R-\mu}{\mu}}\right)^2} > 0$$

$$\tau(c) \equiv \frac{\mu}{4(F + c/2)^2} \left[ S'(c) + \frac{\mu}{4(F + c/2)^2} \right]^{-1} > 0$$

Here  $CW_{TA}'(\theta|c = c^*)$  is the effect of  $\theta$  on consumer welfare found for the basic model evaluated at  $c = c^*$  (see (12)). From (21), I have Proposition 11.

**Proposition 11.** *Under targeted advertising, tougher privacy regulations (i.e. decreasing the signal accuracy  $\theta$ ) decreases the price  $c$  for an ad, which induces firms to enter the*

*market. This increases consumer welfare by giving each consumer a larger and better targeted selection of products to buy from. And this decreases consumer welfare by giving each consumer a larger number of annoying ads.*

*Proof.* A lower signal accuracy induces a lower ad cost  $c$ , which induces more firms to enter the market, which increases the product selection benefit by  $\frac{\partial N_{TA}}{\partial \theta} \tau(c^*) * 2(F + c^*/2) > 0$  and the ad annoyance cost by  $\frac{\partial N_{TA}}{\partial \theta} \tau(c^*) * A'(N_{TA}(\theta))$ .  $\square$

Note that if the supply  $S$  of ads is perfectly elastic, then I get the same results that I found in (12). And if  $S$  is perfectly inelastic, then there is no change in ad annoyance<sup>62</sup> and I get a similar results to the effect found in Bergemann and Bonatti (2011), who showed that in addition to a positive product selection benefit, an increasing signal accuracy would create a negative ad price effect on consumer welfare.

## 8 Ad Retention

Consumers don't always see or remember every ad sent to them. In this section I consider the case where consumers see, remember, or retain ads with at an i.i.d. probability of  $\alpha$  (the ad retention probability). I will extend this in two ways: through allowing consumers to choose their ad retention through paying for ad avoidance (section 8.4), and through allowing firms to choose ad retention by sending multiple messages to each consumer (section 8.5).

Obviously if  $\alpha$  is very small, then there would be a small number of firms competing for each consumer, and therefore the assumption of monopolistic competition should not hold. Therefore I assume that  $\alpha$  is "big enough" for the assumption monopolistic competition to hold. Also if  $\alpha = 1$ , then consumers see all ads and I have the basic model presented in =3. Therefore in this section I will examine the case where  $\alpha \in (0, 1)$ .

---

<sup>62</sup> This is because the total change in ad annoyance is  $\frac{\partial N_{TA}}{\partial \theta} (1 - \tau(c^*)) * A'(N_{TA}(\theta))$  and  $\tau(c^*) = 1$ .

## 8.1 Mass Advertising Equilibrium

Step 1. Firm Demand: Similar to 4.1, if firm  $j$  of type  $k \in \{0, 1\}$  sends ads to all consumers, then half of the consumers it sends ads to will be of type  $k$  and half of the consumers it sends ads to will be of type  $1 - k$ . Because each of these consumers will retain ads with at a probability  $\alpha$  (the ad retention probability), only  $\alpha/2$  of each type of consumer will retain firm  $j$ 's ad. In addition, if all firms mass advertise, then half of the firms that send ads to a consumer will be of his type, and half of the firms that send ads to a consumer will not be of his type. Because each of these consumers will retain ads with at a probability  $\alpha$  (the ad retention probability), each consumer sees  $\alpha/2$  ads from each type of firm. Firm  $j$ 's quantity  $Q_j$  sold as a function of its price  $p_j$  is given by (22), where  $K_{MA} \equiv 1 + \alpha N [e^{\frac{R+b-p}{\mu}} + e^{\frac{R-p}{\mu}}]$ .

$$Q_j = \alpha \frac{\frac{1}{2} [e^{\frac{R+b-p_j}{\mu}} + e^{\frac{R-p_j}{\mu}}]}{K_{MA}} \quad (22)$$

Step 2. Product Price: By optimizing firm  $j$ 's profit over its price  $p_j$  and using symmetry, I find that each advertiser sets a price of  $p = \mu$  and sells the same quantity  $Q$  of their good.

Step 3. Number of Entrants: Because firm  $j$  would advertise to all of the consumers, it would pay an advertising cost  $c$ . By free entry, I have the zero profit condition  $F + c = pQ$ , which solving for  $N$  becomes (23).

$$N = \frac{\mu}{2(F+c)} - \frac{1}{\alpha(e^{\frac{R+b-\mu}{\mu}} + e^{\frac{R-\mu}{\mu}})} \equiv N_{MA} \quad (23)$$

Here  $\frac{\mu}{2(F+c)}$  is the number of firms that would enter each sub-market if there were no outside option. It is also the effect of the entry cost  $F$  and the advertising cost  $c$  on the entry. And  $[\alpha(e^{\frac{R+b-\mu}{\mu}} + e^{\frac{R-\mu}{\mu}})]^{-1}$  is the number of firms discouraged from entering each sub-market due to the outside option.

Step 4. Consumer Welfare: The total consumer welfare is the sum of the

aggregate consumer surplus from sales minus the aggregate ad annoyance cost. From Anderson et al. (1992, p. 60-61), I have that the consumer surplus for this type of model (a monopolistic competitive logit model) is  $CS = \mu \ln K$ . Therefore the consumer welfare  $CW$  is given by (24).

$$CW = R - \mu + \mu \left[ \ln(e^{b/\mu} + 1) + \ln \alpha + \ln\left(\frac{\mu}{2(F+c)}\right) \right] - A(2\alpha N_{MA}) \equiv CW_{MA} \quad (24)$$

Here  $R - \mu + \mu[\ln(e^{b/\mu} + 1) + \ln \alpha + \ln(\frac{\mu}{2(F+c)})]$  is the aggregate consumer surplus gained from sales, while  $A(2\alpha N_{MA})$  is the aggregate ad annoyance cost. Similar to 4.1, Higher entry costs  $F$  or advertising costs  $c$  would hurt consumers through a worse product selection by way of the term  $\mu \ln(\frac{\mu}{2(F+c)})$  and would benefit consumers through less ad annoyance  $A(2\alpha N_{MA})$  by way of less entry. Also higher ad retention  $\alpha$  will benefit consumers through a better product selection by way of the term  $\mu \ln \alpha$  and would hurt consumers through more ad annoyance  $A(2\alpha N_{MA})$  by way of less entry.

## 8.2 Targeted Advertising Equilibrium

Step 1. Firm Demand: Similar to 4.2, If firm  $j$  of type  $k \in \{0,1\}$  sends ads to only those consumers with a signal of  $k$ , then firm  $j$  will send ads to half of the consumers:  $\frac{1+\theta}{2}$  of whom will be of type  $k$  and  $\frac{1-\theta}{2}$  of whom will be of type  $1-k$ . Because each of these consumers will retain ads with at a probability  $\alpha$  (the ad retention probability),  $\alpha \frac{1+\theta}{2}$  type  $k$  consumers and  $\alpha \frac{1-\theta}{2}$  type  $1-k$  consumers will retain the ad. In addition, if all firms target advertise, then  $\frac{1+\theta}{2}$  of the firms that send ads to a consumer will be of his type, and  $\frac{1-\theta}{2}$  of the firms that send ads to a consumer will be not of his type. Because each of these consumers will retain ads with at a probability  $\alpha$  (the ad retention probability), each consumer sees  $\alpha \frac{1+\theta}{2}$  of ads of its type and  $\alpha \frac{1-\theta}{2}$  of ads not of its type. I have that firm  $j$ 's quantity  $Q_j$  sold as a function of its price  $p_j$  is given by (25), where  $K_{TA} \equiv 1 + \alpha N \left[ \frac{1+\theta}{2} e^{\frac{R+b-p_j}{\mu}} + \frac{1-\theta}{2} e^{\frac{R-p_j}{\mu}} \right]$ .

$$Q_j = \alpha \frac{\frac{1}{2} \left[ \frac{1+\theta}{2} e^{\frac{R+b-p_j}{\mu}} + \frac{1-\theta}{2} e^{\frac{R-p_j}{\mu}} \right]}{K_{TA}} \quad (25)$$

Step 2. Product Price: By optimizing firm  $j$ 's profit over its price  $p_j$  and using symmetry, I find that each advertiser sets a price of  $p = \mu$  and sells the same quantity  $Q$  of their good.

Step 3. Number of Entrants: Because firm  $j$  would advertise to all of the consumers, it would pay an advertising cost  $c$ . By free entry, I have the zero profit condition  $F + c/2 = pQ$ , which solving for  $N$  becomes (26).

$$N = \frac{\mu}{2(F + c/2)} - \frac{1}{\alpha \left( \frac{1+\theta}{2} e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2} e^{\frac{R-\mu}{\mu}} \right)} \equiv N_{TA}(\theta) \quad (26)$$

Here  $\frac{\mu}{2(F+c/2)}$  is the number of firms that would enter each sub-market if there were no outside option. It is also the effect of the entry cost  $F$  and the advertising cost  $c/2$  on the entry. And  $[\alpha(\frac{1+\theta}{2} e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2} e^{\frac{R-\mu}{\mu}})]^{-1}$  is the number of firms discouraged from entering each sub-market due to the outside option. Note that, unlike (23), this term depends on the signal accuracy  $\theta$ , because here firms are using their signals to decide which consumers to show their ads.

Step 4. Consumer Welfare: The total consumer welfare is the sum of the aggregate consumer surplus from sales minus the aggregate ad annoyance cost. From Anderson et al. (1992, p. 60-61), I have that the consumer surplus for this type of model (a monopolistic competitive logit model) is  $CS = \mu \ln K$ . Therefore the consumer welfare  $CW$  is given by (27).

$$CW = R - \mu + \mu \left[ \ln \left( \frac{1+\theta}{2} e^{b/\mu} + \frac{1-\theta}{2} \right) + \ln \alpha + \ln \left( \frac{\mu}{2(F+c/2)} \right) \right] - A(\alpha N_{TA}) \equiv CW_{TA}(\theta) \quad (27)$$

Here  $R - \mu + \mu \left[ \ln \left( \frac{1+\theta}{2} e^{b/\mu} + \frac{1-\theta}{2} \right) + \ln \alpha + \ln \left( \frac{\mu}{2(F+c/2)} \right) \right]$  is the aggregate consumer surplus gained from sales, while  $A(\alpha N_{TA})$  is the aggregate ad annoyance cost.



Similar to 4.2, the signal accuracy  $\theta$  does matter, because firms only advertise to those consumers with signals that match their product characteristic. Higher signal accuracy  $\theta$  would benefit consumers through better product selection by way of the term  $\mu \ln(\frac{1+\theta}{2}e^{b/\mu} + \frac{1-\theta}{2})$  and would hurt consumers through higher ad annoyance  $A(N_{TA})$  by way of more entry.

**Proposition 12.** *Ad retention decreases consumer welfare by giving each consumer a smaller and less targeted selection of products to buy from. And increases consumer welfare by giving each consumer a smaller number of annoying ads.*

*Proof.* In both the targeted and mass advertising equilibria, the product selection benefit was decreased by  $-\ln \alpha > 0$ . Under the mass advertising equilibrium, consumer see  $\frac{(1-\alpha)\mu}{F+c}$  fewer ads. Under the targeted advertising equilibrium, consumer see  $\frac{(1-\alpha)\mu}{2(F+c/2)}$  fewer ads. Therefore in both equilibria, the ad annoyance cost was decreased.  $\square$

### 8.3 Equilibrium Determination Condition and the Impact of Signal Accuracy on Welfare

For most signal accuracies  $\theta$  there is only one possible equilibrium; this condition is identical to the condition found in 4.3, because  $\alpha$  cancels itself out. Fewer of a firm's ads are seen, but fewer of a firm's competitors ads are seen. In equilibrium, these forces cancel each other out when a firm decides whether to target or mass advertise. Similarly the consumer welfare at  $\widehat{\theta}$  is also the same for targeted and mass advertising.

**Proposition 13.** *Ad retention affects the product selection benefit independently of signal accuracy  $\theta$ .*

*Proof.* This follows directly from (24) and (27).  $\square$

Proposition 13 is due to logit demand. For different demand functions the effect of ad retention on the product selection benefit could be positive or negative.

Differentiating the consumer welfare under the targeted advertising equilibrium (given by (27)) by the signal accuracy  $\theta$ , I have (28).

$$CW_{TA}'(\theta) = \lambda(\theta) \left[ \mu \left( \frac{1+\theta}{2} e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2} e^{\frac{R-\mu}{\mu}} \right) - A'(\alpha N_{TA}(\theta)) \right] \quad (28)$$

where  $\lambda(\theta) \equiv \frac{\frac{1}{2}(e^{\frac{R+b-\mu}{\mu}} - e^{\frac{R-\mu}{\mu}})}{\left( \frac{1+\theta}{2} e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2} e^{\frac{R-\mu}{\mu}} \right)^2} = \alpha N_{TA}'(\theta) > 0$

Identical to (12), I have  $\lambda(\theta)\mu\left(\frac{1+\theta}{2}e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2}e^{\frac{R-\mu}{\mu}}\right) > 0$  is the aggregate consumer surplus gained from increasing  $\theta$  through more goods being offered to consumers (from a higher number of firms) and through an increase in the chance of products matching consumers' tastes (from more accurate signals). Similar to (12), I have  $\lambda(\theta)A'(\alpha N_{TA}(\theta)) > 0$  is the aggregate ad annoyance gained from increasing  $\theta$  through a higher number of advertisers. These two forces, the product selection effect and the ad annoyance effect, can make increasing signal accuracy  $\theta$  either increase or decrease total consumer welfare.

**Proposition 14.** *Under targeted advertising, tougher privacy regulations (i.e. decreasing the signal accuracy  $\theta$ ) decreases the ad annoyance cost. Compared to the basic model with  $\alpha = 1$ , an ad retention of  $\alpha < 1$  will*

(a) *diminish this decrease in annoyance cost if ad annoyance is convex.*

(b) *magnify this decrease in annoyance cost if ad annoyance is concave.*

*Proof.* This follows from the fact that ad retention lowers the number of ads seen by a consumer. Under the basic model ( $\alpha = 1$ ), consumers see  $\frac{\mu}{2(F+c/2)} - \frac{1}{\frac{1+\theta}{2}e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2}e^{\frac{R-\mu}{\mu}}}$  ads. With  $\alpha < 1$ , consumers see  $\frac{\alpha\mu}{2(F+c/2)} - \frac{1}{\frac{1+\theta}{2}e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2}e^{\frac{R-\mu}{\mu}}}$  ads.  $\square$

Note that with Proposition 13, the only effect ad retention has on how privacy regulations affect welfare is through Proposition 14.

## 8.4 Ad Avoidance

In this section, I explore the consumers choice to block or avoid ads. Consumers decide this pre-game; they decide how many ads to block before learning their valuations of the products (which is revealed to them after they see the ads) and before they learn their valuation of the outside option.<sup>63</sup> Consumers choose ad avoidance by chooses to ignore  $n$  ads for an ad blocking cost  $B(n)$ , where  $B(0) = 0$ ,  $B'(n) > 0$  for all  $n$ , and  $A'(N) > B'(0)$  in equilibrium.<sup>64</sup> Which ads are blocked is chosen at random from the ads shown to the consumer.<sup>65</sup> Consumers benefit from not being annoyed by the avoided ads; ad annoyance  $A$  is now a function of the ads a consumer does not avoid. Consumers suffer from not being able to buy the products of avoided ads.

Johnson (2013) also explores ad blocking in targeted advertising by allowing consumers to choose a zero-one decision: whether to block all advertisements or block no advertisements. In this section, I extend this by allowing consumers to choose to block any quantity of advertisements. Also I am extending his analysis by exploring the costs of ad blocking, like the cost of buying ad blocking software or the cost of changing the television to another channel. In Johnson (2013), the only cost of ad blocking is the opportunity cost of not being able to buy the product that is advertised.<sup>66</sup>

Consumer welfare and equilibrium determination is determined in the same way as 8.1 - 8.4, because firms hold  $\alpha^*$ , the consumers' choice in ad retention, as a given constant. Each consumer chooses the  $\alpha^*$  that matches his or her welfare, which is the same as aggregate consumer welfare. The one difference is now con-

---

<sup>63</sup> Otherwise avoidance would be conditional on the value of the outside option, as in Johnson (2013).

<sup>64</sup> If this last assumption were not true, then the cost of blocking the first ad would be more than the cost of see it. And no ads would be blocked in equilibrium.

<sup>65</sup> Otherwise the consumer would need to spend effort learning about the product advertised to him to determine whether to block it.

<sup>66</sup> With no possibility of choosing to block some ads, additional ad blocking costs are the same as a lower value for the outside option.

sumers pay an additional ad blocking cost of  $B()$ . Therefore under mass advertising consumers choose  $\alpha^*$  according to (29), and under targeted advertising consumers choose  $\alpha^*$  according to (30).

$$A'(2\alpha^*N) - B'(2(1 - \alpha^*)N) = \frac{F + c}{\alpha^*} \quad (29)$$

$$A'(\alpha^*N) - B'((1 - \alpha^*)N) = \frac{F + c/2}{\alpha^*} \quad (30)$$

**Proposition 15.** *Under targeted advertising, tougher privacy regulations (i.e. decreasing the signal accuracy  $\theta$ ) do not affect consumer welfare through ad avoidance.*

*Proof.* Because consumers are optimizing welfare by choosing  $\alpha$ . According to the envelope theorem  $CW_{TA}'(\theta) = \frac{\partial}{\partial \theta} CW_{TA} \big|_{\alpha=\alpha^*}$ . Therefore, consumer welfare is unaffected by the changes in  $\alpha^*$ , because consumers are optimizing  $\alpha^*$  so that  $\frac{d}{d\alpha} CW_{TA} = 0$ .  $\square$

## 8.5 Shouting to Be Heard

Because consumers are blocking some of the ads they receive, firms have an incentive to send multiple ads to increase the probability of their ad being retained by the consumer. This effect is what Anderson and de Palma (2010) calls “shouting to be heard.” In this section, I allow firms to send multiple messages at the constant cost of  $c$  per message. Consumers, however, pay additional ad annoyance cost for each additional retained message; A consumer suffers  $A(N)$  from seeing  $N$  total ads.<sup>67</sup> Thereby consumers could be much worse off under ad retention.

Suppose a firm sent a consumer  $m$  ads, each with an i.i.d. probability of  $\beta$  of being seen (the individual ad retention). Therefore a consumer will see at least one of the firm’s ads with a probability of  $\alpha$  (the composite ad retention) given by equation (31).

---

<sup>67</sup> Future research could allow differences in the costs of seeing an already retained ad and seeing a new ad.

$$\alpha = 1 - (1 - \beta)^m \quad (31)$$

This effects the equilibrium result, because firms are now paying for more than one ad per consumer. The simpler case is when firms target advertise. It is discussed in 8.5.1. When a firm mass advertises, it chooses to send more ads toward consumers with its signal type. I will call this “targeted shouting.” It is discussed in 8.5.2.

### 8.5.1 Targeted Advertising under Shouting

Step 1. Firm Demand: Sending multiple ads to consumers is realized in three places: (1) through a higher  $\alpha$ , (2) through higher ad annoyance costs, and (3) through higher firm advertising costs. Therefore, the demand function (25) still holds for the  $\alpha$  found in (31).

Step 2. Product Price: By free entry, I have the zero profit condition  $F + mc/2 = pQ$ , which I solve to produce a similar result to (26):

$$N = \frac{\mu}{2(F + mc/2)} - \frac{1}{\alpha \left( \frac{1+\theta}{2} e^{\frac{R+b-\mu}{\mu}} + \frac{1-\theta}{2} e^{\frac{R-\mu}{\mu}} \right)} \quad (32)$$

New Step. Shouting First Order Condition: Firms optimize their choice of  $m$ . The marginal cost of sending an additional message per consumer is  $m/2$ . And the marginal revenue of sending an additional message per consumer is  $-\mu \ln(1 - \beta)^{\frac{(1-\alpha)}{2}} * [(\frac{1+\theta}{2})e^{\frac{R+b-\mu}{\mu}} + (\frac{1-\theta}{2})e^{\frac{R-\mu}{\mu}}]/K_{TA}$ . Substituting in (32), I find that all firms choose  $m$  such that:

$$\frac{c}{2} = -\ln(1 - \beta) \frac{1 - \alpha(m^*)}{\alpha(m^*)} (F + \frac{m^*c}{2}) \quad (33)$$

Note that  $m^*$  does not depend on  $\theta$ . Therefore changing the signal accuracy does not change the number of messages per consumer sent from an advertiser.

Therefore I have:

**Proposition 16.** *Under targeted advertising, there is no shouting effect on how signal accuracy  $\theta$  effects consumer welfare.*

### 8.5.2 Targeted Shouting when Mass Advertising

Step 1. Firm Demand: Suppose firm  $j$  of type  $k \in \{0, 1\}$  sends  $m_k$  ads to each consumer of its signal type and  $m_{1-k}$  ads to each consumer not of its signal type. Then half of the consumers (those of its signal type) will see its ad with a probability  $\alpha_k = 1 - (1 - \beta)^{m_k}$ , and half of the consumers (those not of its signal type) will see its ad with a probability  $\alpha_{1-k} = 1 - (1 - \beta)^{m_{1-k}}$ .  $\frac{1+\theta}{2}$  of signal type  $k$  will be type  $k$  and  $\frac{1-\theta}{2}$  of signal type  $k$  will be type  $1 - k$ . Likewise  $\frac{1-\theta}{2}$  of signal type  $1 - k$  will be type  $k$  and  $\frac{1+\theta}{2}$  of signal type  $1 - k$  will be type  $1 - k$ .

It is straight-forward to show that if  $2N$  firms enter the market, each firm sends  $m_k$  ads to each consumer of its signal type and  $m_{1-k}$  ads to each consumer not of its signal type, and if all firms sets a market price of  $p$ , firm  $j$  will sell to a share  $e^{\frac{R+b-p_j}{\mu}} / K_{sh}$  of the type  $k$  consumers that see its ad and a share  $e^{\frac{R-p_j}{\mu}} / K_{sh}$  of the type  $1 - k$  consumers that see its ad, where  $K_{sh} \equiv 1 + \Delta N$  and  $\Delta \equiv \alpha_k [\frac{1+\theta}{2} e^{\frac{R+b-p}{\mu}} + \frac{1-\theta}{2} e^{\frac{R-p}{\mu}}] + \alpha_{1-k} [\frac{1-\theta}{2} e^{\frac{R+b-p}{\mu}} + \frac{1+\theta}{2} e^{\frac{R-p}{\mu}}]$ . Therefore firm  $j$ 's quantity  $Q_j$  sold as a function of its price  $p_j$  is given by (34).

$$Q_j = \frac{\Delta/2}{K_{sh}} \quad (34)$$

Step 2. Product Price: Similar to section 4.1, by optimizing firm  $j$ 's profit over its price  $p_j$  and using symmetry, I find that each advertiser sets a price of  $p = \mu$  and sells the same quantity  $Q$  of their good.

Step 3. Number of Entrants: Because firm  $j$  is sending  $m_k$  ads to half of the consumers and  $m_{1-k}$  ads to the other half of the consumers, it would pay an average advertising cost of  $\bar{m}c$  for each of the consumers, where  $\bar{m} \equiv \frac{m_k + m_{1-k}}{2}$ . By free entry, I have the zero profit condition  $F + \bar{m}c = pQ$ , which solving for  $N$  becomes (35).

$$N = \frac{\mu}{2(F + \bar{m}c)} - \frac{1}{\Delta} \equiv N_{sh} \quad (35)$$

New Step. Shouting First Order Conditions:

Firm  $j$  would optimize its choice of the number of messages sent to each consumer of each signal type. It would send messages so that the marginal cost of sending an additional ad equals the marginal revenue. These first-order conditions become:

$$\text{FOC } m_k : \frac{c}{2} = -\ln(1 - \beta)(1 - \alpha_k^*) \left( \frac{1 + \theta}{2} e^{\frac{R+b-\mu}{\mu}} + \frac{1 - \theta}{2} e^{\frac{R-\mu}{\mu}} \right) \quad (36)$$

$$\text{FOC } m_{1-k} : \frac{c}{2} = -\ln(1 - \beta)(1 - \alpha_{1-k}^*) \left( \frac{1 - \theta}{2} e^{\frac{R+b-\mu}{\mu}} + \frac{1 + \theta}{2} e^{\frac{R-\mu}{\mu}} \right) \quad (37)$$

Note that the left-hand-side of (36) and (37) is the marginal costs of sending an additional ad to each consumer of a signal type and the right-hand-side is the marginal revenues.

**Proposition 17.** *Under targeted shouting (i.e. mass advertising under shouting to be heard), tougher privacy regulations (i.e. decreasing the signal accuracy  $\theta$ ):*

- (a) *decreases the number of messages a firm sends to each consumer of its own signal type, and*
- (b) *increases the number of messages a firm sends to each consumer not of its own signal type.*

*Proof.* By (36) and (37), I have that  $\frac{d\alpha_k^*}{d\theta} > 0$  and  $\frac{d\alpha_{1-k}^*}{d\theta} > 0$ . □

Therefore the equilibrium under shouting to be heard is: When  $\theta = 0$  (i.e. a meaningless signal), advertisers send the same number of messages to each consumer. Increasing  $\theta$ , increases the number of messages a firm sends to each consumer of its own signal type and decreases the number of messages each firms

sends to each consumer not of its signal type, until increasing  $\theta$  causes firms to switch to targeted advertising.<sup>68</sup>

It is straight forward to show that for low values of  $\theta$  that the product selection benefit  $\mu \ln K_{sh}$  and the ad annoyance cost  $A(2\bar{m}N_s h)$  are increasing in  $\theta$ .<sup>69</sup> Therefore depending on the shape of the ad annoyance function  $A()$ , it is possible to have an intermediate consumer welfare optimizing  $\theta$ .

## 9 Common Signals

Information sharing plays a role in the privacy discussion. Firms may receive the same information about a consumer. Firms would receive dependant signals on each consumers' taste, instead of an independent signal (as I have assumed in previous sections). In this section, I explore the other extreme, where firms have the same information or signal about a consumer. This shows how this information sharing would affect my consumer welfare and my equilibrium. This section would be particularly useful for considering the affects of regulating information sharing between firms.

In this section, for each consumer, all firms receive the same, independent (across consumers) signal about that consumer's type, which is true with a probability of  $\frac{1+\theta}{2}$  and false with a probability of  $\frac{1-\theta}{2}$ , where  $\theta \in [0, 1]$ . Here I interpret  $\theta$  as the accuracy of the common signal about each consumer.

For a low signal accuracy (when  $\theta < \widehat{\theta}$ ), the mass advertising equilibrium found in section 4.1 still holds. Firms are ignoring their signals, so information sharing doesn't matter.

For an intermediate signal accuracy, the mixed advertising equilibrium I will

---

<sup>68</sup> This occurs when either  $m_{1-k}^* \leq 0$  (because firms cannot send a negative number of messages to a consumer) or if  $\theta \geq \widehat{\theta}$  (because the firm must be making a positive profit on consumers not of its signal type, see 4.3).

<sup>69</sup> This is not necessarily true for large values of  $\theta$ . For larger values of  $\theta$ , decreases in the average number of messages per consumer  $\bar{m}$  can reduce ad annoyance costs. And firms can even be induced to leave the market, instead of entering.



present in 9.2 holds. If all other firms mass advertised, then the signal is accurate enough so that a firm could make a profit from targeted advertising. Therefore enough firms target advertise, so that firms make zero profit by targeted advertising. If all other firms targeted advertise, then a firm could make a profit mass advertising. By mass advertising, a firm would make a profit selling to those consumers who gave firms a false signal. Therefore enough firms mass advertise, so that firms make zero profit by mass advertising.

For a high signal accuracy, it is still not profitable to sell to consumer not of your signal type, even though none of the other firms of your type are selling to them. Therefore the targeted advertising equilibrium I will present in section 9.1 holds.

## 9.1 Targeted Advertising Equilibrium

Step 1. Firm Demand: If  $N_{TA}$  firms enter each sub-market and all firms target advertise, then each consumer with a true signal will see ads from  $N_{TA}$  firms of his type, and each consumer with a false signal will see ads from  $N_{TA}$  firms not of his type. If firm  $j$  of type  $k \in \{0, 1\}$  target advertises, then firm  $j$  will only advertise to half of the consumers:  $\frac{1+\theta}{2}$  of whom will be of type  $k$ , and  $\frac{1-\theta}{2}$  of whom will be of type  $1 - k$ . Therefore if  $N_{TA}$  firms enter each sub-market and if all firms sets a market price of  $p$ , firm  $j$  will sell to a share  $e^{\frac{R+b-p_j}{\mu}} / K_{TA}^t$  of the type  $k$  consumers and a share  $e^{\frac{R-p_j}{\mu}} / K_{TA}^f$  of the type  $1 - k$  consumers, where  $K_{TA}^t \equiv 1 + N_{TA} e^{\frac{R+b-p}{\mu}}$  and  $K_{TA}^f \equiv 1 + N_{TA} e^{\frac{R-p}{\mu}}$ .<sup>70</sup> Therefore firm  $j$ 's quantity  $Q_j$  sold as a function of its price  $p_j$  is given by (38).

$$Q_j = \frac{1}{2} \left[ \frac{1 + \theta e^{\frac{R+b-p_j}{\mu}}}{2 K_{TA}^t} + \frac{1 - \theta e^{\frac{R-p_j}{\mu}}}{2 K_{TA}^f} \right] \quad (38)$$

Note the difference between (38) and (4). In (4), every firm receives an independent signal about a consumer, so ever consumer receives the same number of false-signals and true-signals. Therefore every consumer sees the same number of

<sup>70</sup>These market shares are found in Anderson et al. (1992, p. 39-40) for a more general framework.

ads from firms of his type and the same number of ads from firms not of his type. In (38), firms receive the same signal about a consumer. Therefore all the products available to a consumer with a true signal are of his type and all the products available to a consumer with a false signal are not of his type.

Step 2. Product Price: Similar to section 4.1, by optimizing firm  $j$ 's profit over its price  $p_j$  and using symmetry, I find that each advertiser sets a price of  $p = \mu$  and sells the same quantity  $Q$  of their good.

Step 3. Number of Entrants: Firms would enter the market until there is no profit from entering the market. Therefore I have  $F + c/2 = pQ$ . This determines the number  $N_{TA}$  of entrants. The big difference between this and (5), is that this equation is not solvable for  $N_{TA}$ . Differentiating this equality with respect to signal accuracy  $\theta$ , I find (39).

$$N_{TA}'(\theta) = \frac{\Lambda - \Upsilon}{\frac{1+\theta}{2}\Lambda^2 + \frac{1-\theta}{2}\Upsilon^2} > 0 \quad (39)$$

$$\text{where } \Lambda \equiv \frac{e^{\frac{R+b-\mu}{\mu}}}{K_{TA}^t} > \Upsilon \equiv \frac{e^{\frac{R-\mu}{\mu}}}{K_{TA}^f} > 0$$

(39) shows that the number  $N_{TA}$  of ads a consumer receives and the number of firms that enter each sub-market, unambiguously increases in signal accuracy. This means that as firms get a better, shared signal about consumers, then more firms enter the market and target advertise.

Step 4. Consumer Welfare: The total consumer welfare is the sum of the aggregate consumer surplus from sales minus the aggregate ad annoyance cost. From Anderson et al. (1992, p. 60-61), I have that the consumer surplus for this type of model (a monopolistic competitive logit model) is  $CS = \mu \ln K$ . Therefore the consumer welfare  $CW$  is given by (40).

$$CW_{TA}(\theta) = \frac{1+\theta}{2} \ln K_{TA}^t + \frac{1-\theta}{2} \ln K_{TA}^f - A(N_{TA}) \quad (40)$$

Here  $\ln K_{TA}^t$  is the product selection benefit to a consumer with a true signal, and  $\ln K_{TA}^f$  is the product selection benefit to a consumer with a false signal. Because  $\frac{1+\theta}{2}$  of consumers have a true signal, the aggregate product selection benefit is  $\frac{1+\theta}{2} \ln K_{TA}^t$  from consumers with a true signal. Likewise the aggregate product selection benefit is  $\frac{1-\theta}{2} \ln K_{TA}^f$  from consumers with a false signal. In addition, all consumers face an ad annoyance from  $N_{TA}$  firms.

To test how signal accuracy affects total consumer welfare under targeted advertising, I differentiate (40) by signal accuracy  $\theta$ , given by (41).

$$CW'_{TA}(\theta) = \frac{1}{2}[\ln K_{TA}^t - \ln K_{TA}^f] + N_{TA}'(\theta) \left[ \left( \frac{1+\theta}{2} \Lambda + \frac{1-\theta}{2} \Upsilon \right) - A'(N_{TA}) \right] \quad (41)$$

Here  $\frac{1}{2}[\ln K_{TA}^t - \ln K_{TA}^f] > 0$  is the change in the product selection benefit of consumers as they switch from having false signals to having true signals. I refer to this as the *signal switching benefit*.  $N_{TA}'(\theta)\Lambda > 0$  is the change in product selection benefit of consumers with true signals and  $N_{TA}'(\theta)\Upsilon > 0$  is the change in product selection benefit of consumers with false signals as more firms enter the market. I refer to these as the *infra-marginal product selection benefits* of consumers with true and false signals. The aggregate  $N_{TA}'(\theta)(\frac{1+\theta}{2}\Lambda + \frac{1-\theta}{2}\Upsilon)$  would be the total infra-marginal product selection benefit. And  $N_{TA}'(\theta)A'(N_{TA})$  is the additional ad annoyance cost faced by each consumer from more firms entering the market. Note that like (12), signal accuracy increases the product selection benefit and the ad annoyance cost. Therefore signal accuracy still has an ambiguous affect on consumer welfare.

## 9.2 Mixed Advertising Equilibrium

Step 1. Firm Demand: If  $N$  firms enter each sub-market and  $N_{MA}$  of those firms mass advertise, then each consumer with a true signal will see ads from  $N$  firms of his type and  $N_{MA}$  firms not of his type, and each consumer with a false signal

will see ads from  $N_{MA}$  firms of his type and  $N$  firms not of his type. If firm  $j$  of type  $k \in \{0,1\}$  of target advertises, then firm  $j$  will only advertise to half of the consumers:  $\frac{1+\theta}{2}$  of whom will be of type  $k$ , and  $\frac{1-\theta}{2}$  of whom will be of type  $1-k$ . Similar to section 9.1, firm  $j$ 's quantity  $Q_j$  sold as a function of its price  $p_j$  would given by (42), where  $K_{Mixed}^t \equiv 1 + Ne^{\frac{R+b-p}{\mu}} + N_{MA}e^{\frac{R-p}{\mu}}$  and  $K_{Mixed}^f \equiv 1 + N_{MA}e^{\frac{R+b-p}{\mu}} + Ne^{\frac{R-p}{\mu}}$ . If firm  $j$  of mass advertises, then firm  $j$  will advertise to all of the consumers: half of whom will be of each type.  $\frac{1+\theta}{2}$  of each type will have a true signal, and  $\frac{1-\theta}{2}$  of each type will have a false signal. Therefore firm  $j$ 's quantity  $Q_j$  sold as a function of its price  $p_j$  would given by (43).

$$Q_{TA}^j = \frac{1}{2} \left[ \frac{1+\theta}{2} \frac{e^{\frac{R+b-p_j}{\mu}}}{K_{Mixed}^t} + \frac{1-\theta}{2} \frac{e^{\frac{R-p_j}{\mu}}}{K_{Mixed}^f} \right] \quad (42)$$

$$Q_{MA}^j = \frac{1}{2} \left[ \frac{1+\theta}{2} \frac{1}{K_{Mixed}^t} + \frac{1-\theta}{2} \frac{1}{K_{Mixed}^f} \right] (e^{\frac{R+b-p_j}{\mu}} + e^{\frac{R-p_j}{\mu}}) \quad (43)$$

Step 2. Product Price: Similar to section 4.1, by optimizing firm  $j$ 's profit over its price  $p_j$  and using symmetry, I find that each advertiser sets a price of  $p = \mu$  and sells the same quantity  $Q$  of their good.

Step 3. Number of Entrants: From the zero profit conditions  $F + c/2 = pQ_{TA}$  and  $F + c = pQ_{MA}$ , I can solve for the number  $N$  of firms that enter each sub-market and the number  $N_{MA}$  of those firms that mass advertise. Yet that is algebraically messy and complicated, so instead I differentiate these zero profit conditions to find  $\frac{\partial K_{Mixed}^t}{\partial \theta} = \frac{K_{Mixed}^t}{1+\theta} > 0$  and  $\frac{\partial K_{Mixed}^f}{\partial \theta} = -\frac{K_{Mixed}^f}{1-\theta} < 0$ . This gives me that the number  $N$  of firms that enter each sub-market is increasing in  $\theta$ , and the number  $N_{MA}$  of those firms that mass advertise is decreasing in  $\theta$ . In addition, it gives me that the number  $N + N_{MA}$  of firms advertising to each consumer is decreasing at a constant rate in  $\theta$ , given be (44).<sup>71</sup> Therefore I conclude that under mixed advertising, increasing signal accuracy unambiguously decreases the ad annoyance faced by

<sup>71</sup> Note that (44) is negative and constant in  $\theta$  because: when  $\theta = \widehat{\theta}$  I have  $N = N_{MA}$ , and  $N''(\theta) + N_{MA}''(\theta) = 0$ .

each consumer.

$$N'(\theta) + N_{MA}'(\theta) = \frac{e^{b/\mu} - 1}{e^{2b/\mu} - 1} \frac{\frac{K_{Mixed}^t}{1+\theta} - \frac{K_{Mixed}^f}{1-\theta}}{e^{\frac{R-\mu}{\mu}}} < 0 \quad (44)$$

Step 4. Consumer Welfare: The total consumer welfare is the sum of the aggregate consumer surplus from sales minus the aggregate ad annoyance cost. From Anderson et al. (1992, p. 60-61), I have that the consumer surplus for this type of model (a monopolistic competitive logit model) is  $CS = \mu \ln K$ . Therefore the consumer welfare CW is given by (45).

$$CW_{mixed}(\theta) = \frac{1+\theta}{2} \ln K_{Mixed}^t + \frac{1-\theta}{2} \ln K_{Mixed}^f - A(N + N_{MA}) \quad (45)$$

Here  $\ln K_{TA}^t$  is the product selection benefit to a consumer with a true signal, and  $\ln K_{TA}^f$  is the product selection benefit to a consumer with a false signal. Therefore the aggregate product selection benefit is  $\frac{1+\theta}{2} \ln K_{Mixed}^t + \frac{1-\theta}{2} \ln K_{Mixed}^f$ . In addition, all consumers face an ad annoyance from  $N + N_{MA}$  firms.

To test how signal accuracy affects total consumer welfare under mixed advertising, I differentiate (45) by signal accuracy  $\theta$ , given by (46).

$$CW_{mixed}'(\theta) = \frac{1}{2} [\ln K_{Mixed}^t - \ln K_{Mixed}^f] - (N'(\theta) + N_{MA}'(\theta))A'(N + N_{MA}) > 0 \quad (46)$$

Here  $\frac{1}{2} [\ln K_{TA}^t - \ln K_{TA}^f] > 0$  is the signal switching benefit, or equivalently the change in the product selection benefit of consumers as they switch from having false signals to having true signals. Also I find that the total infra-marginal production selection benefit from consumers with true signals is  $1/2$ , and the total infra-marginal production selection benefit from consumers with false signals is  $-1/2$ . These two affects cancel each other. And  $(N'(\theta) + N_{MA}'(\theta))A'(N + N_{MA}) < 0$  is the loss in ad annoyance cost as fewer firms advertise to each consumer.

Because consumers benefit from both an increased product selection and a decreased ad annoyance, under the mixed equilibrium, consumer welfare unambiguously increases in signal accuracy. This result is similar to the mixed equilibrium in 7.2, where I found, under endogenous prices and a mixed advertising equilibrium, consumer welfare increases in signal accuracy because as firms switch from mass advertising to targeted advertising, consumers receive fewer ads.

## 10 Conclusion

In conclusion, in order to understand how privacy regulations affect consumer welfare, we need to understand the shape of the ad annoyance function. Are consumers more annoyed by the first ad or the second? This empirical question needs to be answered for us to understand whether we want to have an intermediate-level signal accuracy, or an extreme. And it needs to be answered for us to understand whether privacy regulations should limit the information firms can gather or make it more expensive for firms to gather additional information.

Ad annoyance is a possible explanation for why we value privacy, why we don't want firms to know too much information about us. While we may never be able to test empirically why people value privacy, we shouldn't claim that it has to be purely an intermediate good or purely from the fear of the criminal use of our information.

There is always a consumer welfare preferable targeted advertising equilibrium to the mass advertising equilibrium, under some level of information noise. Therefore we should not make our privacy laws so strict that firms choose not to use the information they are able to get to target their advertisements.

Future work would benefit from extending the framework in this model. The monopolistic competitive framework that I presented in this paper makes the model tractable and applicable to many advertising markets.

## References

- Agarwal, Nikhil, Susan Athey, and David Yang**, "Skewed Bidding in Pay-per-Action Auctions for Online Advertising," *American Economic Review*, 2009, 99 (2), 441–47.
- Agirgas, Cagdas**, "What Drives Media Bias? A Panel Study of Newspaper Archives: 1990-2009," *Job Market Paper*, 2011.
- Anand, Bharat and Ron Shachar**, "Targeted advertising as a signal," *Quantitative Marketing and Economics*, 2009, 7, 237–266.
- Anderson, Simon P. and André de Palma**, "Shouting to be heard in advertising," *Working Paper*, 2010.
- **and Régis Renault**, "Pricing, Product Diversity, and Search Costs: A Bertrand-Chamberlin-Diamond Model," *The RAND Journal of Economics*, 1999, 30 (4), 719–735.
- **and —**, "Advertising Content," *The American Economic Review*, 2006, 96 (1), 93–113.
- **, André de Palma, and Jacques-Francois Thisse**, *Discrete Choice Theory of Product Differentiation*, MIT Press, 1992.
- Athey, Susan and Joshua S. Gans**, "The Impact of Targeting Technology on Advertising Markets and Media Competition," *American Economic Review*, 2010, 100 (2), 608–13.
- Baeza-Yates, R. and B. Ribeiro-Neto**, *Modern Information Retrieval*, New York: ACM Press, 1999.

- Bergemann, Dirk and Alessandro Bonatti**, "Targeting in advertising markets: implications for offline versus online media," *RAND J of Economics*, 2011, 42 (3), 417–443.
- Blei, David M. and John D. Lafferty**, "Dynamic Topic Models," in "ICML" 2006.
- and —, "Topic Models," in A. Srivastava and M. Sahami, eds., *Text Mining: Classification, Clustering, and Applications*, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, 2009.
- , **Andrew Ng, and Michael Jordan**, "Latent Dirichlet allocation," *JMLR*, 2003, 3, 993–1022.
- , **Thomas Griffiths, Michael Jordan, and Joshua Tenenbaum**, "Hierarchical topic models and the nested Chinese restaurant process," in "NIPS" 2003.
- Center for Media Design and the Sequent Partners**, "Video Consumer Mapping Study," *Center for Media Design at Ball State University*, 2009.
- Chandra, Ambarish**, "Targeted Advertising: The Role of Subscriber Characteristics in Media Markets," *The Journal of Industrial Economics*, 2009, 57 (1), 58–84.
- de Cornière, Alexandre**, "Search Advertising," 2011. Paris School of Economics Job Market Paper.
- Diamond, Peter A**, "A Model of Price Adjustment," *Journal of Economic Theory*, 1971, 3 (2), 156–68.
- Dickey, James M.**, "Multiple Hypergeometric Functions: Probabilistic Interpretations and Statistical Uses," *Journal of the American Statistical Association*, 1983, 78 (383), 628–637.
- Dixit, Avinash and Joseph Stiglitz**, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 1977, 67 (3), 297–308.



- Esteban, Lola, Agustín Gil, and José M. Hernández,** "Informative Advertising and Optimal Targeting in a Monopoly," *The Journal of Industrial Economics*, 2001, 49 (2), 161–180.
- , **José M. Hernández, and José Luis Moraga-González,** "Customer Directed Advertising and Product Quality," *Journal of Economics & Management Strategy*, 2006, 15 (4), 943 – 968.
- Esteves, Rosa Branca,** "Customer Poaching and Advertising," *The Journal of Industrial Economics*, 2009, 57 (1), 112–146.
- Galeotti, Andrea and José Luis Moraga-González,** "A Model of Strategic Targeted Advertising," May 2004. CESifo Working Paper No. 1196.
- Gentzkow, Matthew and Jesse M. Shapiro,** "Media Bias and Reputation," *Journal of Political Economy*, 2006, 114 (2), 280–316. Date revised - 2006-08-01; Language of summary - English; Pages - 280-316; ProQuest ID - 56516165; Corporate institution author - Gentzkow, Matthew; Shapiro, Jesse M; DOI - econlit-0859377; 0859377; 0022-3808.
- Ghose, Anindya and Sha Yang,** "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets," *Management Science*, 2009, 55 (10), 1605–1622.
- Goldfarb, A. and C. Tucker,** "Search Engine Advertising: Channel Substitution When Pricing Ads to Context," *Management Science*, 2011.
- Griffiths, Thomas L. and Mark Steyvers,** "Finding Scientific Topics," *PNAS*, 2004, 101 (suppl. 1), 5228–5235.
- Hofmann, Thomas,** "Probilistic latent semantic analysis," in "UAI" 1999.
- Iyer, Ganesh, David Soberman, and J. Miguel Villas-Boas,** "The Targeting of Advertising," *Marketing Science*, 2005, 24 (3), 461 – 476.

**Johnson, Justin P.**, "Targeted Advertising and Advertising Avoidance," 2013. forthcoming RAND J of Economics.

**Larcinese, Valentino, Riccardo Puglisi, and Jr Snyder James M.**, "Partisan Bias in Economic News: Evidence on the Agenda-Setting Behavior of U.S. Newspapers," *Journal of Public Economics*, 2011, 95 (9-10), 1178–1189. Date revised - 2011-09-01; Language of summary - English; Pages - 1178-1189; ProQuest ID - 896012840; Corporate institution author - Larcinese, Valentino; Puglisi, Riccardo; Snyder, James M, Jr; DOI - econlit-1255076; 1255076; 10.1016/j.jpubeco.2011.04.006; 0047-2727.

**Lewis, Randall A. and David H. Reiley**, "Does Retail Advertising Work? Measuring the Effects of Advertising on Sales via a Controlled Experiment on Yahoo!," *Working Paper*, 2011.

— **and —**, "Advertising Effectively Influences Older Users: How Field Experiments Can Improve Measurement and Targeting," *Working Paper*, 2012.

**Lowy, Lisa Sharkis**, "The Existence and Characteristics of Dayparts on the Internet," *The OPA White Papers*, 2003, 1 (3).

**Minka, Thomas and John Lafferty**, "Expectation-Propagation for the Generative Aspect Model," *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, 2002, pp. 352–359.

**PricewaterhouseCoopers**, "IAB Internet Advertising Revenue Report: 2009 Full-Year Results," *Interactive Advertising Bureau (IAB)*, April 2010.

**Rosen-Zvi, Michal, Tom Griffiths, Mark Steyvers, and Padhraic Smyth**, "The Author-Topic Model for Authors and Documents," in "UAI" 2004.

**Shiman, Daniel R.**, "The Impact of Firms' Increased Information about Consumers on the Volume and Targeting of Direct Marketing," 1997. SSRN Working Paper.

- Steyvers, Mark and Tom Griffiths**, “Probabilistic Topic Models,” in T. Landauer, D. Mcnamara, S. Dennis, and W. Kintsch, eds., *Latent Semantic Analysis: A Road to Meaning*, Laurence Erlbaum, 2006.
- Teh, Yee-Whye, David Newman, and Max Welling**, “A Collapsed Variational Bayesian Inference Algorithm for Latent Dirichlet Allocation,” in “NIPS” 2006.
- Teh, Yee Whye, Michael I. Jordan, Matthew J. Beal, and David M. Blei**, “Hierarchical Dirichlet Processes,” *JASA*, 2006, 101.
- Villas-Boas, J. M.**, “Dynamic competition with customer recognition,” *RAND J. Econom*, 1999, 30, 604631.
- , “Price cycles in markets with customer recognition,” *RAND J. Econom*, 2004, 35, 486501.
- Wei, Xing and Bruce Croft**, “LDA-based document models for ad-hoc retrieval,” in “SIGIR” 2006.
- Willmore, Christopher**, “Targeted persuasive advertising.” PhD dissertation, The University of British Columbia 2008.
- Wolinsky, Asher**, “True Monopolistic Competition as a Result of Imperfect Information,” *The Quarterly Journal of Economics*, 1986, 101 (3), 493–512.

## Appendix

### A Comparative Statics on the Marketing Cost Threshold

By (13) and (14), when  $\theta^* \leq 1$ :

$$\frac{\partial \tilde{\theta}}{\partial f} = 2\rho \left[ \frac{1}{1 - v'(\theta)} + (c/2 - M(\theta^*)) \right] > 0 \quad (47)$$

$$\frac{\partial \tilde{\theta}}{\partial c} = \rho \left[ \frac{1}{1 - v'(\theta)} - \frac{f + 2M(\theta^*)}{f + c} \right] \quad (48)$$

$$\frac{\partial \tilde{\theta}}{\partial b} = \rho \omega(\theta^*) \left[ \frac{1}{1 - v'(\theta)} - \frac{f/2 + M(\theta^*)}{M'(\theta^*)} \right] \quad (49)$$

$$\frac{\partial \tilde{\theta}}{\partial \mu} = -\frac{b}{\mu} \frac{\partial \tilde{\theta}}{\partial b} \quad (50)$$

$$\text{where } \rho \equiv \frac{1}{f + c} \frac{e^{b/\mu} + 1}{e^{b/\mu} - 1} > 0$$

$$\omega(\theta) \equiv \frac{2M'(\theta)e^{b/\mu}}{\mu(e^{b/\mu} - 1)^2} > 0$$

## B Proof of Proposition 4

*Proof.* From (13), I have:

$$\begin{aligned} \frac{d_2 \theta^*}{dm df} &= -\frac{M''(\theta^*)[1 - v'(\theta^*)] - M'(\theta^*)v''(\theta^*)}{[M'(\theta^*)(1 - v'(\theta^*))]^2} \frac{d\theta^*}{dm} \\ &= -\frac{(\sigma - 1)^2 m \theta^{*\sigma} - (1/\sigma + 1 - \sigma)(f + c/2)}{\theta^{*2} [M'(\theta^*)(1 - v'(\theta^*))]^2} \frac{d\theta^*}{dm} \end{aligned}$$

$$\frac{d_2 \theta^*}{dm dc} = \frac{1}{2} \frac{d_2 \theta^*}{dm df}$$

$$\text{let } \bar{\sigma} \equiv \frac{1 + \sqrt{5}}{2}$$

$$\text{and } \bar{\theta} \equiv \left( \frac{[1/\sigma + 1 - \sigma][f + c/2]}{m[\sigma - 1]^2} \right)^{1/\sigma}$$

□

## Chapter III

---

# Showing Ads to the Wrong Consumers: Strategic Ad Platform Inefficiency in Online Targeted Advertising

### 1 Introduction

Whenever consumers search, communicate, buy, and surf online, they are bombarded with pop-up ads, banner ads, and sponsored-link ads. Online advertising has become an integral part of consumers' cyber-lives and in many cases an integral part of their real lives. Computing has displaced the radio as our second most time-consuming media outlet (CMD, et al., 2009).<sup>72</sup> At the same time, online advertising has become the third largest advertising market in the United States (PricewaterhouseCoopers, 2010). Yet, online advertising is different from traditional advertising through a much higher prevalence of "targeted advertising" and a high prevalence of "pay-per-click" (PPC) pricing, which is a method of charging for advertisements that is unique to online advertising. In this paper I present a model where an online ad platform "targets" advertisements to maximize its profit under pay-per-click pricing. I find that an online ad platform would not necessarily advertise in the same way that the merchant would, especially when consumers need to be induced to click on the advertisement.

---

<sup>72</sup>Television remains the most time-consuming media outlet.

*Targeted advertising* is when different ads are shown to different consumers based on tastes, locations, or demographics. By advertising for Geno's Cheesesteaks in The Philadelphia Inquirer or for Meow Mix in Cat Fancy, advertisers hope to increase the effectiveness of their ads. Yet targeted advertising is more prevalent and more precise on the internet, because Google, Amazon, Facebook, and other online advertising platforms are better able to personalize ads to fit consumer characteristics and therefore better able to induce consumers to click on their ads through more information about consumers and more precise computing technology (Bergemann and Bonatti, 2011). Google tracks what content we view and what searches we perform through temporary internet files and uses it to show us personalized ads.<sup>73</sup> Amazon uses our shopping histories to recommend products and services to us.<sup>74</sup> Facebook uses our profiles to personalize ads by age, gender, keywords, education, workplace, relationship status, relationship interests, and languages.<sup>75</sup>

Although initially online ad platforms charged a merchant per thousand viewers shown its ad, today most online ad platforms use *pay-per-click* (PPC) pricing.<sup>76</sup> Here a merchant pays for his ad based on the number of consumers who click on his ad or visit his site instead of based on the number of consumers that see his ad or buy his product. This pricing system is practically nonexistent in other forms of advertising like newsprint or television, because there is no easy, verifiable way to determine which ad should get credit for a consumer who visits the retailer. Instead newsprint and television advertisers pay based on the expected sales the ad will generate or the number of people who see the ad.

Some previous literature has examined targeted advertising (see for example: Athey and Gans, 2010; Bergemann and Bonatti, 2011), pay-per-click advertising (Agarwal et al., 2009) and the role of advertising platforms (Ghose and Yang, 2009).

<sup>73</sup>See [http://www.google.com/intl/en\\_us/ads/ads\\_1.html/](http://www.google.com/intl/en_us/ads/ads_1.html/) for details.

<sup>74</sup>See [http://www.amazon.com/gp/seller-account/mm-summary-page.html/ref=gw\\_m\\\_b\\\_awus?ie=UTF8&ld=AZAdvertiseMakeM&topic=200260730](http://www.amazon.com/gp/seller-account/mm-summary-page.html/ref=gw_m\_b\_awus?ie=UTF8&ld=AZAdvertiseMakeM&topic=200260730) for details.

<sup>75</sup>See <http://www.facebook.com/advertising/> for details

<sup>76</sup>This is sometimes called cost-per-click (CPC) pricing.

Yet only this paper and the parallel research of de Cornière (2011) analyze targeted pay-per-click advertising in a market with an ad platform. He primarily explores the case where the merchants choose how to target their advertisements. A merchant chooses which consumers to show its ad based on consumer characteristics. I explore in this paper and he briefly considers in an extension the case where the ad platform chooses how to target advertisements. The ad platform chooses which consumers it shows an ad based on consumer characteristics. I get different results from his extension, because I allow the ad platform to have more control over which consumers see an ad. His targeting technology only allows the choice of a minimum reservation price. Consumers with lower valuations for the product would not see the ad, while consumers with higher valuations for the product would see the ad. In this paper, I show that the ad platform would not necessarily want to target ads in this way.

To do this I adapt a classic costly search model. Here, each consumer does not know how much he will value the product (their reservation price) until he clicks on an ad. A consumer will only click on the ad, if his expected benefit from doing so is more than the search cost, which is the opportunity cost or travel cost.<sup>77</sup>

In addition I let advertising be informative. Here, a consumer will not be able to buy the product or even click on the ad, unless the consumer is shown the ad. The ad platform can induce consumers to click by only showing the ad to a subset of consumers. The firm and the consumers know that if a consumer is shown the ad, they belong to this subset of consumers. This influences the firm's pricing decision and can increase the expected benefit from clicking on the ad.

Because ad platforms benefit by increasing the clicking on the ad, while merchants benefit by increasing its profits, their incentives generally are not aligned. I start by assuming that the pay-per-click price of an ad is exogenous. This assumption is common in the targeted advertising literature (see for example Anand

---

<sup>77</sup>This assumes there is no added utility from shopping.

and Shachar, 2009; Iyer et al., 2005; Johnson, 2013). Although special cases exist where the pay-per-click price is fixed,<sup>78</sup> in general how an ad platform targets its advertisements should influence what merchants are willing to pay for advertising. Therefore I then relax this assumption by allowing the ad platform to make a take-it-or-leave-it offer to the merchant when choosing how to target the ad.

In the first period of my game, each consumer draws his reservation price  $r$  for the merchant's product. This reservation price is not observed by the consumer, because the consumer does not know the merchant's product's characteristics, and it is not observed by the merchant, because the merchant does not know the consumer's characteristics. Only the advertising platform knows both, so only the advertising platform learns the consumer's reservation price  $r$ . In the second period of the game, the advertising platform chooses how the ad is targeted by choosing the proportion of consumers with each reservation price that is shown the ad; this proportion is observed by all agents. In the third period of the game, simultaneously the firm chooses its single price  $p$ , and each consumer who is shown the ad decides whether or not to click on it. Then all the consumers who clicked on the ad observe their reservation price  $r$  and the product price  $p$ . Then in the last period of the game, each consumer who clicked on the ad decides whether or not to purchase the product. Then the game ends with all agents receiving their payoffs.

The consumers who did not click on the ad get a payoff of zero, whether or not they were shown the ad. The consumers who clicked on the ad pay an ad annoyance cost  $b$ . I interpret the search cost as the forgone opportunities and energy that the consumer gives up to review the product details.<sup>79</sup> If a consumer

---

<sup>78</sup>For example, many online advertising platforms, including Google, provide a minimum pay-per-click price. A fixed pay-per-click price would occur when only one advertiser is bidding on a keyword. See <https://adwords.google.com/select/KeywordToolExternal> for details.

<sup>79</sup>The consumer was not already looking for a product when the ad popped up. I am not modeling the decision to surf the net. I am modeling the decision to click on an ad. The decision to surf is exogenous.



buys the product, then he gets payoff  $r - p$  in addition to paying  $b$ . The firm gets a payoff equal to its price,  $p$ , for every consumer who buys the product while paying a fixed pay-per-click price  $c$  for every consumer who clicks on his ad. The payoff to the advertising platform is the advertising revenue from the firm.

When an online advertising platform's choice of to whom it shows an ad influences the number of clicks and not the pay-per-click price of an ad, then the platform will not show the ad to *some* consumers that it would rationally expect to buy the product in order to change the shape of the demand curve, inducing lower prices. This increases the expected benefit from clicking, leading to more clicking. For example, Google might show an ad for a Gershwin album to a rap-music-loving teenager and not to a music professor. And Amazon might show an ad for a book on game theory to a garage mechanic and not show the same ad to me. This shows the importance that pay-per-click pricing plays on how an ad platform targets its advertisements.

## 2 Literature Review

My model is similar to the Bertrand-Chamberlin-Diamond search models with a one-sided market found in Wolinsky (1986) and Anderson and Renault (1999), and most similar to de Cornière (2011) and Anderson and Renault (2006).

In de Cornière (2011), Anderson and Renault (2006), and this paper if consumers knew their reservation prices before clicking on the ad then we would face the classic Diamond (1971)'s Paradox. Yet limited information about a product in the ad, through ad content (Anderson and Renault, 2006) or through the consumer's knowledge that the ad was targeted toward them (de Cornière, 2011, and my model) induces clicking while avoiding the Diamond (1971) Paradox.

In Anderson and Renault (2006), a merchant encourages some consumers to click on its ad or travel to its store through the content of its advertising. Their advertising content acts similarly to targeted advertising in my model, signaling a

consumer whether or not to click on an ad. The key difference between our models is that I include an advertising platform and pay-per-click pricing.

Like de Cornière (2011), I find that an advertising platform would not target-advertise in the same way that a merchant would. His targeting technology only allows the choice of a minimum reservation price. Consumers with lower valuations for the product would not see the ad, while consumers with higher valuations for the product would see the ad. He finds that the ad platform would not set the same minimum reservation price that the merchant would, sometimes higher and sometimes lower. I relax this targeting technology by allowing the ad platform to choose (with some reasonable constraints) the proportion of consumers that it shows the ad to for each reservation price. I show that for high enough search costs the ad platform would show the ad to some consumers who it rationally expects not to buy the product and not showing the same ad to other consumers who it would rationally expect to buy the product.

In addition, we have two key differences in our models: 1) de Cornière (2011) is a repeated search game while my model is not repeated, and 2) he assumes that consumers are already induced to click on the first ad. This causes him to find in the case where the ad platform chooses how to target the advertisements that targeted advertising induces higher product prices, while I find that they lead to lower product prices. In his paper a consumer chooses when to stop looking at ads (through a repeated costly search game) while in this paper a consumer chooses whether or not to look at the first ad (and pay the first search cost). For example, it would be more reasonable to use his model when consumers are searching for a product through a list of ads on Amazon. While it would be more reasonable to use my model when consumers are deciding whether to click on a banner ad on a webpage.

## 2.1 The Effect of Targeted Advertising on Product Price

The prevailing reasoning in the informative targeted advertising literature is targeted advertising induces higher product prices by reducing price competition (see for example: Iyer et al., 2005).<sup>80</sup> In this paper, I show that an ad platform will targeted advertise in a way that lowers prices to discourage ad avoidance. De Cornière (2011) argues that an platform would strategically target the ad to induce higher product prices further so the ad platform can charge more for advertising. I get a different result than de Cornière (2011), because in de Cornière (2011) a consumer chooses when to stop looking at ads while in this paper a consumer chooses whether or not to look at the first ad.

Under a product monopoly, Esteban et al. (2001) argues that targeted advertising sometimes induces higher product prices because advertising may effectively become a fixed cost instead of a marginal cost. I get a different result from Esteban et al. (2001), because I show that under pay-per-click targeted advertising the merchant would still treat the cost of advertising like a fixed cost. Under a product monopoly with endogenous product quality, Esteban et al. (2006) argues that targeted advertising sometimes induces higher product prices and sometimes induces lower product prices depending on which consumers are more willing to pay for additional quality: those with higher valuations for the product or those with lower valuations for the product. My results are not comparable with his because I do not endogenize product quality.

When costs of finding a consumer willing to buy the product in two different market segments are asymmetric, Galeotti and Moraga-González (2004) argues that targeted advertising induces higher product prices in the more expensive segment and lower prices in the less expensive segment. In my model, in a sense, the ad

---

<sup>80</sup>Willmore (2008) also argues that this was the case in persuasive targeted advertising. When targeted advertising is treated as a signal of heterogeneous product characteristics, Anand and Shachar (2009) argues that targeted advertising can induce lower product prices because a false signaling merchant can use lower prices to compensate consumers.

platform is making the market into a more expensive segment. Fewer consumers who see the ad will be willing to buy the product. Yet the ad platform does so in a way that induces lower product prices, not higher. I get a different result than Galeotti and Moraga-González (2004), because in my model, the ad platform is targeting the advertisement in a specific way to change the shape of the product's demand curve.

## 2.2 Ad Avoidance

In my model, I allow consumers to choose whether or not to click on an ad, therefore *avoiding* paying the search cost to learn about their reservation price for the product, learn the price of the product, and get a chance to buy the product. This choice is a choice of *advertisement avoidance*. Johnson (2013) studied the effect of targeted advertising with advertisement avoidance. In Johnson (2013), an individual merchant ignores the effect of its targeting strategy on consumer ad avoidance strategies, and the product price is exogenous. In my model, the ad platform is strategically playing off the rationally expected product price and the ad avoidance decision of the consumers to maximize its own profit.

## 3 The Model

Here I present a basic costly search model of a two-sided market for online targeted advertising. In one side of the market, an online *merchant* buys the opportunity to offer its product to consumers from a single, monopolist online *advertisement platform* through informative advertising. In the other side of the market, there is a unit mass of consumers whom the ad platform could choose to show the ad. Those consumers who are shown the ad choose whether to suffer ad annoyance in order to have the opportunity of buying the merchant's product. The advertising platform chooses which consumers see the ad in order to maximize the rational expectation of its ad revenue.

I present the model in the order of the timing of the game in order to avoid confusion about who knows what when. I will discuss each agent's objectives when I discuss their choices.

**Phase 0/Setup**, the *reservation price allocation phase*: The merchant's product is summarized by a characteristic  $x$ , which is a point drawn from a uniform distribution on the unit circle. The value of  $x$  is observed by the merchant and the platform, but not by the consumers. This assumption is motivated by the idea that only those agents who are familiar with the product know the value of  $x$ .<sup>81</sup>

Each consumer's taste is summarized by a characteristic  $y$ , which is a point drawn from a uniform distribution on the unit circle. The value of  $y$  is observed by the consumer and the platform, but not by the merchant and the other consumers. This assumption is motivated by the idea that only those agents who are familiar with the consumer's tastes know the value of  $y$ .<sup>82</sup> This consumer has a reservation price  $r \equiv 2|x - y|$  for the product. Because all consumers equidistant from  $x$  would have identical tastes for the product, I index each consumer by his reservation price  $r$ . Note only the ad platform knows both  $x$  and  $y$ , therefore only the ad platform knows  $r$  for each consumer. At this point, the merchant and the consumers only have the distributional knowledge that  $r \sim i.i.d.U[0, 1]$ .

**Phase 1**, the *targeting decision phase*: the advertising platform informative advertises the merchant's product at a constant pay-per-click price  $c$  for a constant marginal cost normalized to zero. The advertising platform chooses the probability that a consumer with reservation price  $r$  sees the ad to maximize the advertising revenue (i.e. its own profit)  $A \equiv cQ_c$ . This choice is represented by the function  $f : [0, 1] \rightarrow [0, 1]$ : a consumer with a reservation price  $r$  has  $f(r)$  chance of seeing the ad and  $1 - f(r)$  chance of not seeing the ad. A consumer who sees no ad does not know about the product, and therefore cannot purchase the product. A consumer

---

<sup>81</sup>The ad platform could have observed some of the merchant's past sales, or the ad platform could have examined the merchant's website itself.

<sup>82</sup>The ad platform could have observed some of the consumer's past purchases, or the consumer could have told the ad platform its tastes through online surveys.

understands that he has been chosen to receive an ad conveys information about his match quality. This is in fact the only information content of the ad. The ad platform chooses the value of  $f$  for every price  $p \in [0, 1]$  such that  $f$  is *nice*. A function  $f$  is *nice* if it satisfies functional form conditions C1-5 in Appendix B; these conditions amount to assuming that  $f$  is not too discontinuous.

In order for targeted advertising to exist, I assume that the advertising platform's choice of targeting strategy  $f$  is publicly observed. This is to disallow a credibility issue, where the advertising platform tells the merchant and consumers it is using one targeting strategy while it is really using another.<sup>83</sup>

To simplify my notation I define the variable  $\omega^i$  as one if consumer  $i$  is shown the ad and zero if consumer  $i$  is not shown the ad.<sup>84</sup>

Also to keep my model simple, I assume that the ad reveals no information about the product to the consumer. This is similar to most of the advertising literature.

**Phase 2, the clicking and pricing decision phase:** Here the merchant chooses its single price  $p$ , and each consumer who is shown the ad chooses whether or not to click, simultaneously and independently.<sup>85</sup>

The merchant sells a single good over the Internet and faces a constant marginal cost normalized to zero. It chooses the price  $p$  of its product to maximize its expectation of its profit  $\Pi \equiv pQ - cQ_c$ , where  $Q$  is the quantity of consumers

---

<sup>83</sup>If the choice of  $f$  were not observed, any equilibrium would involve showing the ad to all consumers. You can see this if you suppose that the ad platform assumes that consumers would believe the strategy it announced. Then it would be optimal for the ad platform to choose to show the ad to all consumers and announce another strategy to try to fool the consumers. The consumers would rationally expect the ad platform to lie and show the ad to all consumers, so no equilibrium with  $f(r) < 1$  for some  $r \in [0, 1]$  could exist.

<sup>84</sup>Note the ad platform is choosing  $f$  and not  $\omega$ . Therefore all consumers with the same reservation price have the same chance of seeing the ad. This rules out equilibria where different consumers with the same  $r$  have different expectations from clicking on the same ad.

<sup>85</sup>The assumption that price and clicking is chosen after targeting advertising is motivated by the idea of a behemoth ad platform catering to many different markets. In addition it is motivated by the theoretical concept that you need to give the ad platform incentive not to show the ad to every consumer, so consumers would have to choose to click after the ad platform commits and announces its targeting strategy.

that buy its product,  $c$  is the pay-per-click cost of its advertisement, and  $Q_c$  is the quantity of consumers that click on its ad.

Each consumer  $r$  decides whether to click on the ad to maximize his expected utility  $E[u(r)] \equiv (E[(r - p)\omega_b^r] - b)\omega_c^r$ , where  $\omega_b^r$  is one if consumer  $r$  buys the product (and otherwise is zero),  $b$  is the ad annoyance cost (or search cost) of clicking on the ad, and  $\omega_c^r$  is one if consumer  $r$  clicks on the ad (and otherwise is zero). Note that consumer  $r$  is choosing  $\omega_c^r$ . Further note that as consumer  $r$  has not learned the price of the product and his reservation price  $r$  so his future choice of  $\omega_b^r$  may be at this point unknown.

I define the quantity of the quantity of consumers that click on the ad as  $Q_c \equiv \int_0^1 \omega_c^r dr$ .

**Phase 3**, the *price reassurance and reservation price revelation phase*: Each consumer  $r$  who clicked on the ad now sees the product price  $p$  and his reservation price  $r$ .<sup>86</sup>

**Phase 4**, the *sales phase*: Here the consumers who clicked on the ad choose whether or not to buy the product. Each consumer  $r$  who clicked on the ad chooses whether to purchase the product to maximize his utility  $u(r) \equiv (r - p)\omega_b^r - b$ . Note that consumer  $r$  is now choosing  $\omega_b^r$ , so consumer  $r$  will buy the product (i.e.  $\omega_b^r = 1$ ) if  $r \geq p$ . I assume that consumers buy when indifferent. If a consumer does not click on the ad, then he cannot buy the product, so his utility  $u(r)$  is automatically zero.

I define the quantity of the product sold as  $Q \equiv \int_0^1 \omega_b^r dr$ . If  $Q = 1$ , then all consumers buy the product, and if  $Q = 0$ , then no consumers buy the product.

**END, Profits**: Here all the agents get their payoffs.

## 4 Each Consumer's Clicking Decision

In this section, I will start to recursively solve for a subgame perfect Nash Equilibrium. In the last section, I found that a consumer who has clicked on the ad

---

<sup>86</sup>In equilibrium, the product price that the consumers rationally expect will be the actual price. Hence, the revelation of the price can be seen as a reassurance of the price.

will buy the product if his reservation price  $r$  is greater than or equal to the price  $p$ . Here, I solve for a consumer's decision whether to click on the ad once he has been shown the ad. In the next section, I will analyze the merchant's pricing decision. From these two sections I will be able to aggregate all the decisions made in Phase 2, the clicking and pricing decision phase. Then in Section 6, I will solve for the equilibrium by solving for the ad platform's decision of which consumers to advertise to. All lemmas and proofs are given in Appendix A.

To calculate his expected payoff, a consumer has two observed pieces of information: 1) the fact that he is shown the ad and 2) the advertising platform's targeting strategy. From this information he develops rational beliefs about the price  $p^*$ , which I will refer to as the rationally expected price, and a distribution of his possible reservation prices  $f/\bar{F}$ , where  $\bar{F} \equiv \int_0^1 f(p)dp$ . Weighting  $f$  by  $1/\bar{F}$  makes a probability density function—the probability density function of a random reservation price from those consumers shown the ad.

Therefore if a consumer is shown the ad, then his expected benefit  $b_c(p^*, f)$  from clicking on the ad is given by equation (1), where  $p^*$  is the consumer's rational expectations or beliefs of the products price.

$$b_c \equiv E[\max\{r - p^*, 0\} | \omega = 1] = \int_{p^*}^1 (p - p^*) f(p) dp / \bar{F} \quad (1)$$

A consumer's expected payoff from clicking is his expected benefit  $b_c$  minus his search cost  $b$ . Thus I have the *consumer clicking condition* (CCC) given below.

**Consumer Clicking Condition (CCC):** A consumer who is shown the ad will click on the ad if and only if  $b_c \geq b$ .

This is an identical decision for all consumers that are shown the ad. Either all consumers who are shown the ad click or none click. Note that this means that if  $f$  satisfies C1-5 then the density function of consumers clicking  $f_c$  satisfies C1-5.



## 5 The Merchant's Pricing Decision

In this section, I solve for the merchant's pricing decision. This occurs simultaneously with the consumer's clicking decisions. In the following section, I will use the conditions found in both sections to solve for the advertising platform's targeting decision and solve the equilibrium.

The merchant chooses its price  $p^*$  to maximize its profit  $(p - c)Q - cQ_c$ . Because the merchant takes the advertising cost  $cQ_c$  as a constant sunk cost, the merchant chooses its price  $p^*$  to maximize its sales revenue  $pQ$ .

The merchant takes the density function  $f$  of consumers shown the ad and infers its demand function  $Q(p)$ . Suppose that the merchant believes that, conditional on being shown the ad, a consumer will click through with probability  $\theta$ , regardless of  $r$ . Then he will anticipate demand of  $Q(p^*) = \theta \int_{p^*}^1 f(p)dp$ . I have already established in section 4 that either all consumers who are shown the ad click or none click. Therefore the merchant believes that either  $\theta = 0$  or  $1$ . Therefore, the merchant sets prices according to the *merchant's profit maximizing condition* (MPMC) given below.<sup>87</sup>

**Merchant's Profit Maximizing Condition (MPMC):** The merchant will choose its price  $p^*$  such that  $p^* = \arg \max_p p' \int_{p'}^1 f(p)dp$ .

If  $f$  is continuous at  $p^*$ , then the merchant's profit maximizing condition is equivalent to the first-order-condition given in equation (2).

$$p^* f(p^*) = \int_{p^*}^1 f(p)dp \quad (2)$$

This equation is the standard Bertrand profit maximization first-order conditions for the demand curve  $Q(p) = \int_p^1 f(r)dr$ .

---

<sup>87</sup>If  $\theta = 0$ , any price would maximize profit. Therefore I assume that the merchant will arbitrarily chose the price given by MPMC to rule out the uninteresting degenerate equilibrium where the merchant sets its price really high and none of the consumers click on the ad.

## 6 The Ad Platform's Targeting Decision

In this section, I solve for the advertising platform's choice of the proportion  $f(p)$  of consumers with each reservation price  $r = p$  to whom the ad is shown, given the consumer and merchant strategies discussed above. Because the pay-per-click price  $c$  has already been determined before the game, the advertising platform chooses its targeting strategy to maximize the quantity  $Q_c$  of consumers clicking. The advertising platform knows the consumer clicking condition (CCC) and the merchant's profit maximizing condition (MPMC), so the advertising platform is able to rationally expect who will click under any choice of  $f$ .

### 6.1 A Benchmark Result: Advertise to All Consumers

In this first subsection, I analyze when the advertising platform would show the ad to all consumers. Here, the function  $f(p) = 1\{p \in [0, 1]\}$ , so by the profit maximizing condition, the merchant would set the standard monopoly price  $p^m = 1/2$ .

Anticipating this price outcome, consumers would click on the ad if the expected benefit  $b_c$  from clicking on the ad were greater than the search cost  $b$ . This happens when the consumer clicking condition (CCC) is met and simplifies to the condition given in equation (3).

$$b_1 \equiv \int_{p^m}^1 (p - p^m) dp = 1/8 \geq b \quad (3)$$

If this condition is met, then the advertising platform would want to advertise to everyone because doing so gets all consumers to click on the ad, yielding the maximum possible mass of consumers clicking,  $Q_c = 1$ .<sup>88</sup>

In addition, it would also be optimal for the ad platform to advertise to all consumers if the search cost  $b$  were so high that no consumer would ever click on the ad under any targeting strategy. In this case any targeting strategy is optimal.

---

<sup>88</sup>This is formalized in Lemma 1.

## 6.2 Advertise to Some Consumers

Next, I develop some preliminary results to examine the case when the advertising platform would show the ad to some but not all consumers. When  $f(p) = 1\{p \in [0, 1]\}$  would not induce consumers to click, while some other targeting strategy  $\tilde{f}$  would then the advertising platform would optimally show the ad to some but not all consumers.<sup>89</sup>

In this section I describe a functional form of  $f$  that the advertising platform would choose. I do so by finding a functional form of  $f$  that the advertising platform would weakly prefer. I show that given any targeting strategy  $\tilde{f}$  satisfying C1-5, there exists a targeting strategy  $f$  satisfying C1-5 and the functional form shown in Proposition 18.<sup>90</sup>

**Proposition 18.** *Given any targeting strategy  $\tilde{f}$  satisfying C1-5, there exists a targeting strategy  $f$  satisfying C1-5 and the following functional form condition that produces (weakly) more clicking:*

$$f(p) = \begin{cases} 0 & \text{if } p < \underline{p} \\ 1 & \text{if } p \in [\underline{p}, p^*] \\ (p^*/p)^2 & \text{if } p \in (p^*, \widehat{p}] \\ 1 & \text{if } \widehat{p} \neq 1 \text{ and } p \in (\widehat{p}, 1] \end{cases}$$

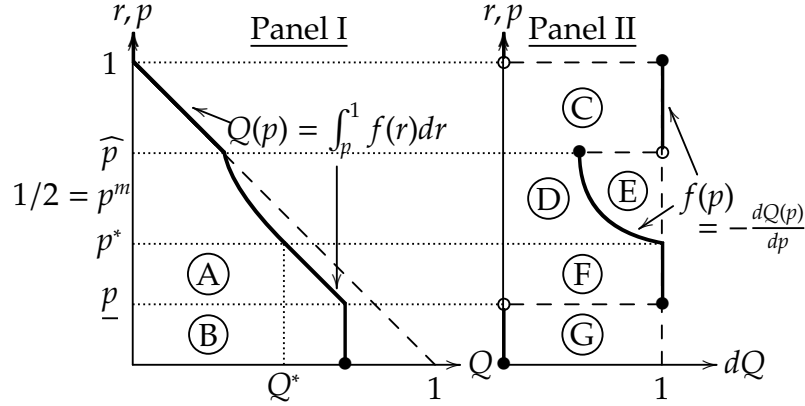
where  $0 \leq \underline{p} \leq p^* \leq \widehat{p} \leq 1$  and  $p^*$  satisfies (MPMC) for  $f$ .

Figure 1 illustrates the targeting strategies shown in Proposition 18. The thick lines are various forms of the targeting strategy  $f$  in different spaces: I) the demand curve  $Q(p) = \int_p^1 f(r)dr$  and II) the density function  $f(p) = -\frac{dQ(p)}{dp}$ . In panel I, I compare the demand curve  $Q(p)$  to the function  $1 - p$  (i.e. the dashed line), the demand curve that the merchant would face if all consumers were shown and

<sup>89</sup>This is formalized in Lemma 2.

<sup>90</sup>I find this functional form piecemeal in lemmas 3, 4 and 5.

Figure 1: A Weakly Preferable Targeting Strategy



subsequently clicked on the ad. In panel II, I compare the density function  $f$  to the function  $1\{p \in [0, 1]\}$  to compare the density of consumers shown the ad to the density of all consumers at every reservation price.

The advertising platform uses the merchant's profit-maximizing condition to pick the merchant's rationally expected price  $p^*$ . Panel I illustrates this by having the merchant pick its price  $p^*$  to maximize  $pQ(p)$ , given by  $\textcircled{A} + \textcircled{B}$ . The merchant's first-order condition from equation (2) indicates how it would maximize its profit in Panel II. The merchant would pick  $p^*$  to equate the marginal loss  $pf(p)$  from raising its price (i.e.  $\textcircled{F} + \textcircled{G}$ ) with the marginal gain  $\int_p^1 f(r)dr$  from raising its price (i.e.  $\textcircled{C} + \textcircled{D}$ ).

The targeting strategy given in Figure 1 gives the merchant equal profit for setting any price between  $p^*$  and  $\widehat{p}$ . In Panel I, the box  $\textcircled{A} + \textcircled{B}$  of the merchant's before-advertising expense profit  $pQ(p)$  is equal and maximized for any price  $p \in [p^*, \widehat{p}]$ . Thus the demand curve between prices  $p^*$  and  $\widehat{p}$  is a demand curve of constant profit  $Q(p) = (p^*/p)^2$ . In Panel II, the box  $pf(p)$  of marginal profit lost from raising the price (i.e.  $\textcircled{F} + \textcircled{G}$ ) is equal to the region  $Q(p) = \int_p^1 f(r)dr$  of marginal profit gained (i.e.  $\textcircled{C} + \textcircled{D}$ ) for any price  $p \in [p^*, \widehat{p}]$ . This produces the curve between  $p^*$  and  $\widehat{p}$ .<sup>91</sup>

<sup>91</sup>Note that the condition that a merchant who is indifference between a set of prices would always pick the lowest price is playing a role here by making the merchant choose the price  $p^*$ .

In the strategy depicted in Figure 1, the advertising platform is showing the ad to consumers  $Q_c = \textcircled{C} + \textcircled{D} + \textcircled{F}$  and not to consumers  $\textcircled{E} + \textcircled{G}$ . This induces a rationally expected (weakly) lower product price, which makes the consumers more willing to click on the ad. In the next section I will show when it is optimal for an ad platform to show the ad to consumers who would not buy the product  $\textcircled{F}$  and not to consumers who would buy the product  $\textcircled{E}$ .

### 6.3 Optimal Advertising

I now build upon the previous two subsections to find the model's equilibrium and the ad platform's optimal targeting strategy  $f$  for any search cost  $b$  and any pay-per-click price  $c$ . In Section 7 I will examine how the payoffs depend on  $b$  and  $c$ .

When  $f(p) = 1\{p \in [0, 1]\}$  would not induce consumers to click, while some other targeting strategy  $\tilde{f}$  would then the advertising platform would optimally show the ad to some but not all consumers.<sup>92</sup> The ad platform would choose to show the ad to just enough consumers so the search cost  $b$  equals the consumer's expected benefit from clicking on the ad  $b_c$ . If  $b_c < b$ , then the consumers would not be induced to click on the ad. If  $b_c > b$ , then the ad platform could increase its profits by increasing the amount of consumers it shows the ad.<sup>93</sup> Using this and Proposition 18, I find Proposition 19.

**Proposition 19.** *An optimal click-maximizing targeting strategy satisfies C1-5 and the following:*

- (a) *when  $b \leq b_1$ : the advertising platform shows the ad to all consumers*

---

Without this condition, the ad platform could choose a strategy close to this strategy to guarantee the merchant choose the lowest price, by giving the merchant slightly more profit from choosing  $p^*$ . The problem with this is that for any close strategy, the ad platform can choose a strategy slightly closer to gain slightly more clicks.

<sup>92</sup>This is formalized in Lemma 2.

<sup>93</sup>This is formalized in Lemma 2.

(b) when  $b_1 < b \leq b_2$ :  $\widehat{p} \in (p^m, p_1]$  and

$$f(p) = \begin{cases} 1 & \text{if } p \in [0, p^*] \\ (p^*/p)^2 & \text{if } p \in (p^*, \widehat{p}] \\ 1 & \text{if } p \in (\widehat{p}, 1] \end{cases}$$

(c) when  $b_2 < b \leq b_3$ :

$$f(p) = \begin{cases} 0 & \text{if } p < \underline{p} \\ 1 & \text{if } p \in [\underline{p}, p^*] \\ (p^*/p)^2 & \text{if } p \in (p^*, p_1] \\ 1 & \text{if } p \in (p_1, 1] \end{cases}$$

(d) when  $b_3 < b \leq b_4$ :  $\widehat{p} \in (p_1, p_2]$  and

$$f(p) = \begin{cases} 0 & \text{if } p < p^* \\ (p^*/p)^2 & \text{if } p \in (p^*, \widehat{p}] \\ 1 & \text{if } p \in (\widehat{p}, 1] \end{cases}$$

(e) when  $b > b_4$ : no consumer would ever click on the ad so any  $f$  is optimal;

where  $0 < b_1 < b_2 < b_3 < b_4$ ,  $0 < p^* \leq p^m < p_1 < p_2 < 1$ , and  $p^*$  satisfies (MPMC) for  $f$ .

For low ad-annoyance costs, the platform focuses on excluding some potential inframarginal consumers by increasing  $\widehat{p}$ . While this does tend to make consumers more pessimistic about their  $r$ , conditional on seeing the ad, it also induces a lower  $p^*$  from the merchant. The latter effect is sufficiently strong so that not only do consumers become more willing to click, but they are more willing to click than if the ad platform had instead excluded consumers that would not be willing to buy the product by increasing  $\underline{p}$ . For somewhat higher ad-annoyance costs, the ad platform excludes additional potential consumers who it knows would not be willing to buy the product by increasing  $\underline{p}$ , because the additional potential inframarginal consumers that the ad platform would have to exclude would have higher reservation prices  $r$ . For even higher ad-annoyance costs, the platform would have to exclude even more potential inframarginal consumers by increasing  $\widehat{p}$ , because

Table 2: Comparative Static Results of the Equilibrium Payoffs

Search Cost	Consumers	Ad Platform	Merchant if $c > p^*$	Merchant if $c < p^*$
$b < b_1$	$\frac{\partial E(u^i)}{\partial b} = -1,$ $\frac{\partial E(u^i)}{\partial c} = 0$	$\frac{\partial A}{\partial b} = 0,$ $\frac{\partial A}{\partial c} > 0$	$\frac{\partial \Pi}{\partial b} = 0, \frac{\partial \Pi}{\partial c} < 0$	
$b_1 < b < b_2$	$\frac{\partial E(u^i)}{\partial b} = 0,$  $\frac{\partial E(u^i)}{\partial c} = 0$	$\frac{\partial A}{\partial b} < 0,$	$\frac{\partial \Pi}{\partial b} > 0,$	$\frac{\partial \Pi}{\partial b} < 0, \frac{\partial \Pi}{\partial c} < 0$
$b_2 < b < b_3$				$\frac{\partial \Pi}{\partial b} > 0, \frac{\partial \Pi}{\partial c} < 0$
$b_3 < b < b_4$		$\frac{\partial A}{\partial c} > 0$	$\frac{\partial \Pi}{\partial c} < 0$	$\frac{\partial \Pi}{\partial b} < 0, \frac{\partial \Pi}{\partial c} < 0$
$b > b_4$		$\frac{\partial A}{\partial b} = \frac{\partial A}{\partial c} = 0$	$\frac{\partial \Pi}{\partial b} = \frac{\partial \Pi}{\partial c} = 0$	

it would not be advertising to any potential consumers who it knows would not be willing to buy the product.

## 7 Examination of Equilibrium Payoffs

In this section I examine the payoffs of the equilibrium where: 1) the ad platform chooses the click-maximizing targeting strategy found in Proposition 19, 2) the merchant chooses its profit-maximizing price through equation (2) and 3) the consumers click when  $b \leq b_4$ . Table 2 summarizes my findings. I begin by examining a representative consumer's expected utility. Then I examine the ad platform's and the merchant's profits.

### 7.1 Consumer Payoffs

By Lemma 6, if the search cost  $b > b_1$ , then a consumer's expected benefit  $b_c$  from clicking on the ad would be equal to  $b$ . Therefore his expected payoffs  $E[u^i] \equiv b_c - b$  from clicking would be equal to zero. Therefore as long as  $b > b_1$ ,  $\partial E(u^i)/\partial b = 0$ . Yet if  $b < b_1$ , then all consumers would choose click on the ad for any search cost  $b$  or for any merchant's marginal cost  $c$  of production. Also changing the pay-per-click

price  $c$  does not change  $f$  or  $p^*$ , so  $\partial E(u^i)/\partial c = 0$  for any  $b$  and  $c$ .

## 7.2 Ad Platform Profits

If the consumer search cost  $b < b_1$ , then the ad platforms profit  $A$  (i.e. the total ad revenue) would be equal to the pay-per-click price  $c$  times the total mass of consumers clicking  $Q_c$ , which equals one. Because both are exogenous, increasing or decreasing the consumer search cost  $b$  would not change the ad platform's profit  $A$ . If  $b_1 < b < b_4$ , then the ad platform's profit  $A$  would be equal to the mass of consumers shown the ad  $\bar{F} \equiv \int_0^1 f(p)dp$  times the pay-per-click price  $c$ . Because  $\bar{F}$  is decreasing in search cost  $b$ , so is the ad platform's profit  $A$ . Because  $\bar{F}$  is unaffected by changing  $c$ ,  $\partial A/\partial c = Q_c = \bar{F} > 0$ . Yet if  $b > b_4$ , then the ad platform cannot induce any consumers to click on the ad, so changing the search cost  $b$  or the pay-per-click price  $c$  would not affect the ad platform's profit  $A$ .

## 7.3 Merchant Profits

If the consumer search cost  $b < b_1$ , then the merchant sets the monopolist price  $p^m$  and its ad is shown to all consumers. Small changes in the consumer search cost do not discourage or encourage more consumers from clicking on the ad or buying the product, so  $\partial \Pi/\partial b = 0$ .

If  $b_1 < b < b_4$ , then by equation (2) I have  $\partial \Pi/\partial \widehat{p} = 2(p^* - c) * \partial p^*/\partial \widehat{p}$ . Therefore if the merchant faces a high enough pay-per-click price  $c$ , then the effect of changing the search cost  $b$  on the merchant's profit  $\Pi$  would be overwhelmed by the effect on the ad revenue  $A$ , so  $\partial \Pi/\partial b > 0$ . Yet if the merchant faces a low enough pay-per-click price  $c$ , then the effect of changing the search cost  $b$  on the merchant's profit  $\Pi$  would be overwhelmed by its profit from sales  $pQ$ . If  $b_1 < b < b_2$  or  $b_3 < b < b_4$ , then a higher search cost  $b$  leads to a higher  $\widehat{p}$ , which leads to a lower price  $p^*$  and less profit  $\Pi$ . If  $b_2 < b < b_3$ , then a higher search cost  $b$  leads to a higher  $\underline{p}$ . The merchant's profit from sales  $pQ$  would not be affected and the merchant faces a lower advertisement cost  $A$ , so  $\partial \Pi/\partial b > 0$ .



If the consumer search cost  $b > b_4$ , then no consumer would ever click on the ad, so  $\partial\Pi/\partial b$ .

Also because  $c$  does not affect  $f$ ,  $\partial\Pi/\partial c = -\bar{F} = Q_c \geq 0$

## 8 Take-it-or-leave-it Offer Advertising

One criticism of my model is that I take the pay-per-click price of advertising as an exogenous constant  $c$ . In this section I explore when the effect of the targeting strategy influences the price of advertising. I analyze a simple adaptation to my model with an endogenous price of advertising.

### 8.1 The Take-it-or-leave-it Offer Model

Here I present my adaptation to my model to include endogenous pay-per-click pricing. Instead of a pay-per-click price  $c$ , the merchant chooses whether to accept or reject a take-it-or-leave-it offer from the advertisement platform that includes which consumers would be shown the ad and a fee  $\phi$  for advertising.

**Phase 0**, the *reservation price allocation phase*: same as in Section 3.

**Phase 1**, the *targeting decision phase and take-it-or-leave-it offer phase*: Here the ad platform chooses  $f$  and makes a take-it-or-leave-it advertising offer  $\phi$  to the merchant to maximize its expected profit  $A$ , which is  $\phi$  when the merchant accepts the take-it-or-leave-it advertising offer and 0 otherwise. The advertisement platform chooses  $f$  as described in Section 3.  $f$  still must satisfy C1-5. Unlike Phase 1 in Section 3, consumers are not shown the ad in this phase. The ad platform is only committing to show the ad to the consumers specified by  $f$  if the merchant accepts the take-it-or-leave-it advertising offer.

Here the merchant and the consumers learn both the take-it-or-leave-it advertising offer  $\phi$  and the targeting decision  $f$ . It is public knowledge.

**After Phase 1 and Before Phase 2**, the *offer acceptance phase*: Here the merchant chooses either to accept the take-it-or-leave-it offer (and set  $A = \phi$ ) or to reject the

take-it-or-leave-it offer (and set  $A = 0$ ) to maximize its expectation of its profit  $\Pi$  which is  $pQ - \phi$  when it accepts the take-it-or-leave-it offer and 0 otherwise.

If the merchant accepts this offer, then each consumer with a reservation price of  $r = p$  is shown the ad with probability  $f(p)$  and the merchant pays the ad platform  $\phi$  for advertising (instead of paying  $A = cQ_c$ ). If the merchant rejects the offer, then no consumers are shown the ad, and the ad platform and the merchant get a payoff of 0.

**Phases 3-END:** same as in Section 3, except  $A$  is  $\phi$  when the merchant accepts the take-it-or-leave-it advertising offer and 0 otherwise.

## 8.2 The Take-it-or-leave-it Equilibrium

The merchant would accept any offer satisfying  $\phi \leq p^* \int_{p^*}^1 f(p)dp$  and (CCC). The ad platform will make the largest take-it-or-leave-it offer that the merchant would be willing to accept, so the ad platform will set  $\phi = p^* \int_{p^*}^1 f(p)dp$  and thus extract all the surplus from the merchant. The ad platform will choose its targeting strategy  $f$  to maximize the offer  $\phi$  that the merchant would accept. Therefore the ad platform will choose a targeting strategy that maximizes the profit  $p^* \int_{p^*}^1 f(p)dp$  of the merchant. One such targeting strategy is given in Proposition 20.<sup>94</sup>

**Proposition 20.** *An optimal merchant-profit-maximizing targeting strategy satisfies C1-5 and the following:*

- (a) *when  $b \leq b_5$ :  $p^* = p^m$  and*
- $$f(p) = \begin{cases} 0 & \text{if } p < p^m \\ 1 & \text{if } p \in (p^m, 1] \end{cases}$$
- (b) *when  $b_5 < b \leq b_4$ :  $\widehat{p} \in (p^m, p_2]$  and*

---

<sup>94</sup>Lemma 7 in Appendix A is used to prove Proposition 20. See Appendix A for a formal proof of both.

$$f(p) = \begin{cases} 0 & \text{if } p < p^* \\ (p^*/p)^2 & \text{if } p \in (p^*, \widehat{p}] \\ 1 & \text{if } p \in (\widehat{p}, 1] \end{cases}$$

(c) when  $b > b_4$ : no consumer would ever click on the ad so any  $f$  is optimal;

where  $0 < b_5 < b_4$ ,  $0 < p^* \leq p^m < p_2 < 1$ , and  $p^*$  satisfies (MPMC) for  $f$ .

The merchant gets no profit from advertising to consumers that it would never sell to. Therefore the ad platform has no reason to advertise to these consumers when trying to maximize the value of a take-it-or-leave-it offer  $\phi$ , as seen in Proposition 20. Note that for low enough search costs  $b \leq b_3$ , the advertising platform would advertise to some consumers that would not buy the product to increase the number of clicks on the ad, as seen in Proposition 19. Therefore when  $b \leq b_3$ , take-it-or-leave-it offer advertising is more efficient than pay-per-click advertising with fixed pay-per-click prices.

Athey and Gans (2010) argued that when there is no private cost from advertising to consumers that would not buy the product, there is no need not to advertise to these consumers. When  $b < b_1$ , it is still optimal for the ad platform to show the ad to all consumers, because (CCC) would still be met and the merchant would get the same profit. Proposition 20 shows that when there is no private cost and no private gain from advertising to consumers that would not buy the product, there is no need to advertise to these consumers either.

Another interesting result of Proposition 20 is the ad platform's strategy when  $b_5 \leq b \leq b_4$ . Here the ad platform is not able to induce consumers to click on the ad with the targeting strategy  $f(p) = 1\{p \in [p^m, 1]\}$ , so the ad platform chooses a targeting strategy that will induce the merchant to charge a lower price. This is similar to the result in Anderson and Renault (2006). They explore endogenous advertising content in a costly search model. They find that when search cost are large enough, merchants will commit to lower prices in their advertisement content to induce consumers visit their store or click on their ad.

Also note that when search cost  $b$  satisfies  $b \geq b_3$ , then the targeting strategies chosen in Propositions 19 and 20 are identical. Therefore pay-per-click advertising with fixed pay-per-click prices is efficient when search costs  $b$  are large enough.

### 8.3 Which Model is More Realistic?

Exogenous pay-per-click pricing is not an unrealistic assumption. It means that the merchant has already committed to advertising at a pay-per-click price  $c$  when the ad platform chooses whom to show the ad. Also in industries with many merchants, it is reasonable to expect that merchants take the price of advertising as exogenous. Yet if there is little cost to drawing out complicated contracts between the ad platform and the merchant, then it is reasonable to assume the equilibrium reflects the take-it-or-leave-it offer equilibrium. Therefore either model may be more appropriate depending on the industry.

## 9 Concluding Remarks

I found that when the platform is maximizing the number of clicks, it will not show the ad to some consumers that it would rationally expect to buy the product. And it will show an ad to some consumers that it would rationally expect not to buy the product. Targeting in this way changes the shape of the demand curve, inducing online merchants to lower their prices  $p^*$ . This increases the expected benefit from clicking  $b_c$ , leading to more clicking.

Future research in targeted advertising should look at the incentives of online advertising platforms. My model shows that the ad platform could show ads strategically inefficiently and de Cornière (2011) shows that an ad platform can over advertise. In Section 8, I showed how take-it-or-leave-it offer pricing can induce the ad platform to target its ads efficiently. Yet in reality, online advertising platforms sell a limited quantity of ad space through auctions. Future research should look at how online auctions and an advertising capacity constraint influence how ad

platforms choose to target advertisements.

Also future research needs to test how competition between ad platforms and online merchants affects targeted advertising. Perhaps the kind of strategic inefficiency I found does not exist when multiple online ad platforms (say Yahoo and Google) compete to sell advertisement space to online merchants. Or perhaps the substitution between goods sold by online merchants induces ad platforms to differentiate the ads shown to different consumers.

## References

- Agarwal, Nikhil, Susan Athey, and David Yang**, "Skewed Bidding in Pay-per-Action Auctions for Online Advertising," *American Economic Review*, 2009, 99 (2), 441–47.
- Agirgas, Cagdas**, "What Drives Media Bias? A Panel Study of Newspaper Archives: 1990-2009," *Job Market Paper*, 2011.
- Anand, Bharat and Ron Shachar**, "Targeted advertising as a signal," *Quantitative Marketing and Economics*, 2009, 7, 237–266.
- Anderson, Simon P. and André de Palma**, "Shouting to be heard in advertising," *Working Paper*, 2010.
- **and Régis Renault**, "Pricing, Product Diversity, and Search Costs: A Bertrand-Chamberlin-Diamond Model," *The RAND Journal of Economics*, 1999, 30 (4), 719–735.
- **and —**, "Advertising Content," *The American Economic Review*, 2006, 96 (1), 93–113.
- **, André de Palma, and Jacques-Francois Thisse**, *Discrete Choice Theory of Product Differentiation*, MIT Press, 1992.

- Athey, Susan and Joshua S. Gans**, "The Impact of Targeting Technology on Advertising Markets and Media Competition," *American Economic Review*, 2010, 100 (2), 608–13.
- Baeza-Yates, R. and B. Ribeiro-Neto**, *Modern Information Retrieval*, New York: ACM Press, 1999.
- Bergemann, Dirk and Alessandro Bonatti**, "Targeting in advertising markets: implications for offline versus online media," *RAND J of Economics*, 2011, 42 (3), 417–443.
- Blei, David M. and John D. Lafferty**, "Dynamic Topic Models," in "ICML" 2006.
- and —, "Topic Models," in A. Srivastava and M. Sahami, eds., *Text Mining: Classification, Clustering, and Applications*, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, 2009.
- , **Andrew Ng, and Michael Jordan**, "Latent Dirichlet allocation," *JMLR*, 2003, 3, 993–1022.
- , **Thomas Griffiths, Michael Jordan, and Joshua Tenenbaum**, "Hierarchical topic models and the nested Chinese restaurant process," in "NIPS" 2003.
- Center for Media Design and the Sequent Partners**, "Video Consumer Mapping Study," *Center for Media Design at Ball State University*, 2009.
- Chandra, Ambarish**, "Targeted Advertising: The Role of Subscriber Characteristics in Media Markets," *The Journal of Industrial Economics*, 2009, 57 (1), 58–84.
- de Cornière, Alexandre**, "Search Advertising," 2011. Paris School of Economics Job Market Paper.
- Diamond, Peter A**, "A Model of Price Adjustment," *Journal of Economic Theory*, 1971, 3 (2), 156–68.

- Dickey, James M.**, "Multiple Hypergeometric Functions: Probabilistic Interpretations and Statistical Uses," *Journal of the American Statistical Association*, 1983, 78 (383), 628–637.
- Dixit, Avinash and Joseph Stiglitz**, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 1977, 67 (3), 297–308.
- Esteban, Lola, Agustín Gil, and José M. Hernández**, "Informative Advertising and Optimal Targeting in a Monopoly," *The Journal of Industrial Economics*, 2001, 49 (2), 161–180.
- , **José M. Hernández, and José Luis Moraga-González**, "Customer Directed Advertising and Product Quality," *Journal of Economics & Management Strategy*, 2006, 15 (4), 943 – 968.
- Esteves, Rosa Branca**, "Customer Poaching and Advertising," *The Journal of Industrial Economics*, 2009, 57 (1), 112–146.
- Galeotti, Andrea and José Luis Moraga-González**, "A Model of Strategic Targeted Advertising," May 2004. CESifo Working Paper No. 1196.
- Gentzkow, Matthew and Jesse M. Shapiro**, "Media Bias and Reputation," *Journal of Political Economy*, 2006, 114 (2), 280–316. Date revised - 2006-08-01; Language of summary - English; Pages - 280-316; ProQuest ID - 56516165; Corporate institution author - Gentzkow, Matthew; Shapiro, Jesse M; DOI - econlit-0859377; 0859377; 0022-3808.
- Ghose, Anindya and Sha Yang**, "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets," *Management Science*, 2009, 55 (10), 1605–1622.
- Goldfarb, A. and C. Tucker**, "Search Engine Advertising: Channel Substitution When Pricing Ads to Context," *Management Science*, 2011.

**Griffiths, Thomas L. and Mark Steyvers**, “Finding Scientific Topics,” *PNAS*, 2004, 101 (suppl. 1), 5228–5235.

**Hofmann, Thomas**, “Probilistic latent semantic analysis,” in “UAI” 1999.

**Iyer, Ganesh, David Soberman, and J. Miguel Villas-Boas**, “The Targeting of Advertising,” *Marketing Science*, 2005, 24 (3), 461 – 476.

**Johnson, Justin P.**, “Targeted Advertising and Advertising Avoidance,” 2013. forthcoming RAND J of Economics.

**Larcinese, Valentino, Riccardo Puglisi, and Jr Snyder James M.**, “Partisan Bias in Economic News: Evidence on the Agenda-Setting Behavior of U.S. Newspapers,” *Journal of Public Economics*, 2011, 95 (9-10), 1178–1189. Date revised - 2011-09-01; Language of summary - English; Pages - 1178-1189; ProQuest ID - 896012840; Corporate institution author - Larcinese, Valentino; Puglisi, Riccardo; Snyder, James M, Jr; DOI - econlit-1255076; 1255076; 10.1016/j.jpubeco.2011.04.006; 0047-2727.

**Lewis, Randall A. and David H. Reiley**, “Does Retail Advertising Work? Measuring the Effects of Advertising on Sales via a Controlled Experiment on Yahoo!,” *Working Paper*, 2011.

— **and —** , “Advertising Effectively Influences Older Users: How Field Experiments Can Improve Measurement and Targeting,” *Working Paper*, 2012.

**Lowy, Lisa Sharkis**, “The Existence and Characteristics of Dayparts on the Internet,” *The OPA White Papers*, 2003, 1 (3).

**Minka, Thomas and John Lafferty**, “Expectation-Propagation for the Generative Aspect Model,” *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, 2002, pp. 352–359.



- PricewaterhouseCoopers**, "IAB Internet Advertising Revenue Report: 2009 Full-Year Results," *Interactive Advertising Bureau (IAB)*, April 2010.
- Rosen-Zvi, Michal, Tom Griffiths, Mark Steyvers, and Padhraic Smyth**, "The Author-Topic Model for Authors and Documents," in "UAI" 2004.
- Shiman, Daniel R.**, "The Impact of Firms' Increased Information about Consumers on the Volume and Targeting of Direct Marketing," 1997. SSRN Working Paper.
- Steyvers, Mark and Tom Griffiths**, "Probabilistic Topic Models," in T. Landauer, D. Mcnamara, S. Dennis, and W. Kintsch, eds., *Latent Semantic Analysis: A Road to Meaning.*, Laurence Erlbaum, 2006.
- Teh, Yee-Whye, David Newman, and Max Welling**, "A Collapsed Variational Bayesian Inference Algorithm for Latent Dirichlet Allocation," in "NIPS" 2006.
- Teh, Yee Whye, Michael I. Jordan, Matthew J. Beal, and David M. Blei**, "Hierarchical Dirichlet Processes," *JASA*, 2006, 101.
- Villas-Boas, J. M.**, "Dynamic competition with customer recognition," *RAND J. Econom*, 1999, 30, 604-631.
- , "Price cycles in markets with customer recognition," *RAND J. Econom*, 2004, 35, 486-501.
- Wei, Xing and Bruce Croft**, "LDA-based document models for ad-hoc retrieval," in "SIGIR" 2006.
- Willmore, Christopher**, "Targeted persuasive advertising." PhD dissertation, The University of British Columbia 2008.
- Wolinsky, Asher**, "True Monopolistic Competition as a Result of Imperfect Information," *The Quarterly Journal of Economics*, 1986, 101 (3), 493–512.

## Appendix

### A Mathematical Proofs and Lemmas

**Lemma 1.** *If  $b_1 = 1/8 \geq b$ , then the advertising platform chooses to show the ad to all consumers.*

*Proof of Lemma 1.* If  $f(p) = 1\{p \in [0, 1]\}$  by (MPMC) the merchant would set the standard monopoly price  $p^m = 1/2$ . Therefore consumers would rationally expect the merchant to set the price  $p^* = p^m = 1/2$ . Therefore by (CCC), the consumers would click on the ad if  $1/8 \geq b$ . This gives the advertising platform its maximum possible ad revenue. C4 and C5 guarantee uniqueness by preventing removable discontinuities.  $\square$

**Lemma 2.** *If  $b_1 < b$  and there exists a targeting strategy  $\tilde{f}$  satisfying C1-5 that would induce consumers to click, then the advertising platform will not advertise to every consumer.*

*Proof of Lemma 2.* The advertising platform strictly prefers targeting strategy  $\tilde{f}$  to  $f(p) = 1\{p \in [0, 1]\}$ . Therefore  $f$  is not optimal.  $\square$

**Lemma 3.** *Given any targeting strategy  $\tilde{f}$  satisfying C1-5, there exists a targeting strategy  $f$  satisfying C1-5 and the following functional form condition that produces (weakly) more clicking:*

$$f(p) = \begin{cases} 0 & \text{if } p < \underline{p} \\ 1 & \text{if } \underline{p} \neq p^* \text{ and } p \in [\underline{p}, p^*] \end{cases}$$

where  $0 \leq \underline{p} \leq p^* \leq 1$  and  $p^*$  satisfies (MPMC) for  $f$ .

*Proof of Lemma 3.* Let  $\tilde{p}$  be the price set by the merchant under  $\tilde{f}$ . If  $\tilde{f}$  satisfies the functional form condition, then let  $f = \tilde{f}$  and  $p^* = \tilde{p}$ . Otherwise  $0 < \int_0^{\tilde{p}} \tilde{f}(p)dp < \tilde{p}$ , because  $\tilde{f}$  satisfies C1-3. If  $\int_0^{\tilde{p}} \tilde{f}(p)dp = 0$  or  $\tilde{p}$  then  $\tilde{f}$  would satisfy the functional form condition. Let  $\underline{p} \equiv \tilde{p} - \int_0^{\tilde{p}} \tilde{f}(p)dp$ . Let  $f(p) \equiv 1\{p \leq \underline{p}\} + \tilde{f} * 1\{p > \underline{p}\}$ . Because  $f(p') = \tilde{f}(p') \forall p' > \tilde{p}$ , we have that  $p' \int_{p'}^1 f(p)dp = p' \int_{p'}^1 \tilde{f}(p)dp \forall p' > \tilde{p}$ . Therefore by (MPMC), the merchant would prefer  $\tilde{p}$  to all prices  $p' > \tilde{p}$  under the targeting strategy  $\tilde{f}$ . If the merchant would sets price  $p^* = \tilde{p}$  under the targeting strategy  $f$ , then  $f$  satisfies the functional form condition. If the merchant would set its price  $p^* < \tilde{p}$  under the targeting strategy  $f$ , then  $f$  adds more slackness to the consumer clicking condition because:

$$\begin{aligned} \int_{p^*}^1 (p - p^*)f(p)dp &= \int_{\tilde{p}}^1 (p - \tilde{p})\tilde{f}(p)dp + (\tilde{p} - p^*) \int_{\tilde{p}}^1 \tilde{f}(p)dp + \int_{p^*}^{\tilde{p}} (p - p^*)dp \\ &> \int_{\tilde{p}}^1 (p - \tilde{p})\tilde{f}(p)dp \end{aligned} \quad (4)$$

Therefore if consumers would click on the ad under the targeting strategy  $\tilde{f}$ , then consumers would click on the ad under targeting strategy  $f$ .  $\square$

**Lemma 4.** *Given any targeting strategy  $\tilde{f}$  satisfying C1-5, there exists a targeting strategy  $f$  satisfying C1-5 and  $f(p^*) = 1$  that produces (weakly) more clicking, where  $p^*$  satisfies (MPMC) for  $f$ .*

*Proof of Lemma 4.* Let  $\tilde{p}$  be the price set by the merchant under  $\tilde{f}$ . If  $\tilde{f}(\tilde{p}) = 1$ , then let  $f = \tilde{f}$  and  $p^* = \tilde{p}$ . If  $\int_0^{\tilde{p}} \tilde{f}(p)dp > 0$ , then such an  $f$  exists by Lemma 3. Otherwise define  $f_\epsilon(p) \equiv \tilde{f}(p) * 1\{p > \tilde{p}\} + 1\{\tilde{p} - \epsilon \leq p \leq \tilde{p}\} \forall \epsilon > 0$ . Let  $p_\epsilon$  be the price set by the merchant under  $f_\epsilon$ . By (MPMC),  $\tilde{p}f(\tilde{p}) \geq \int_{\tilde{p}}^1 \tilde{f}(p)dp$ . Therefore  $\tilde{p} = \tilde{p}f_\epsilon(\tilde{p}) > \int_{\tilde{p}}^1 f_\epsilon(p)dp \forall \epsilon > 0$ . Hence  $p_\epsilon < \tilde{p}$  by (MPMC). Choose an arbitrary small  $\epsilon > 0$  such that  $\tilde{p} - \epsilon = (\tilde{p} - \epsilon)f_\epsilon(\tilde{p} - \epsilon) > \int_{\tilde{p}-\epsilon}^1 f_\epsilon(p)dp$ . Therefore  $p_\epsilon = \tilde{p} - \epsilon$ .

Let  $f = f_\epsilon$  and  $p^* = p_\epsilon$ . Note that  $f$  adds more slackness to the consumer clicking condition because of equation (4). Therefore if consumers would click on the ad under targeting strategy with the density function  $\tilde{f}$ , then consumers would click on the ad under targeting strategy with the density function  $f$ .  $\square$

**Lemma 5.** *Given any targeting strategy  $\tilde{f}$  satisfying C1-5, there exists a targeting strategy  $f$  satisfying C1-5 and the following functional form condition that produces (weakly) more clicking:*

$$f(p) = \begin{cases} (p^*/p)^2 & \text{if } \widehat{p} \neq p^* \text{ and } p \in (p^*, \widehat{p}] \\ 1 & \text{if } \widehat{p} \neq 1 \text{ and } p \in (\widehat{p}, 1] \end{cases}$$

where  $0 \leq p^* \leq \widehat{p} \leq 1$  and  $p^*$  satisfies (MPMC) for  $f$ .

*Proof of Lemma 5.* Let  $\widehat{p}$  be the price set by the merchant under  $\tilde{f}$ . If  $\tilde{f}$  satisfies the functional form condition, then let  $f = \tilde{f}$  and  $p^* = \widehat{p}$ . If no consumers would be induced to click under  $\tilde{f}$ , then any targeting strategy is weakly preferable to  $\tilde{f}$ , so choose  $f$  that satisfies C1-5 and the functional form condition.

Otherwise  $\int_{\widehat{p}}^1 \tilde{f}(p)dp < 1 - \widehat{p}$  and consumers are induced to click under  $\tilde{f}$ . By Lemma 4 choose a targeting strategy  $f_1$  that produces weakly more clicking than  $\tilde{f}$  satisfying C1-5 and  $f_1(p^*) = 1$ , where  $p^*$  is the rationally expected price under  $f_1$ . Note that  $(p^*/p)^2 \leq 1 \forall p \geq p^*$ .

Because  $f_1$  has no removable, infinite or essential discontinuities and only finite jump discontinuities, I split  $(p^*, 1]$  into a countable series of intervals such that: 1)  $f_1$  is either entirely weakly above or entirely weakly below the curve  $(p^*/p)^2$  for any given interval, 2) no two weakly above intervals border each other (I would join these two intervals together to make one interval), and 3) no two weakly below intervals border each other. Let  $\{p_k\}_{k=0,1,\dots}$  be the sequence of prices at the bounds of these intervals. Note  $p_0 = p^*$ .

Because  $g_1(p) \equiv (p^*/p)^2 1\{p \in (p^*, p_1]\} + f(p) 1\{p \notin (p^*, p_1]\}$  solves the differential equation for  $p^* \int_{p^*}^1 g_1(p) dp = p' \int_{p'}^1 g_1(p) dp \forall p' \in [p^*, p_1]$ , by (MPMC) the first such interval  $(p^*, p_1]$  is necessarily weakly above  $(p^*/p)^2$ . Therefore I choose  $\tilde{p}_1$  satisfying  $\int_{p^*}^{p_2} f_1(p) dp = \int_{p^*}^{\tilde{p}_1} (p^*/p)^2 dp + (p_2 - \tilde{p}_1)$ .

Define  $f_2(p) \equiv (p^*/p)^2 1\{p \in (p^*, \tilde{p}_1]\} + 1\{p \in (\tilde{p}_1, p_2]\} + f(p) 1\{p \notin (p^*, p_2]\}$ . Note that the merchant would set price  $p^*$  under  $f_2$ . For prices outside the interval  $(p^*, p_2]$ , the merchant's demand curve would not change. For prices inside the interval  $(p^*, p_2]$ , the merchant would have less incentive for raising its price above  $p^*$ . Also note that the expected value of clicking,  $b_c$ , would increase, so consumers would have more incentive to click. Further note that mass of consumers clicking,  $Q_c$ , would remain constant.

Now  $f_2$  on the interval  $(p^*, p_3]$  is necessarily weakly above  $(p^*/p)^2$ ,  $f_2$  on the interval  $(p_3, p_4]$  is necessarily weakly below  $(p^*/p)^2$ , and so on. I repeat the same process I used to define  $f_2$  from  $f_1$  to define  $f_3$  from  $f_2$ . I iterate over this to get a function that satisfies the functional form conditions. If this takes finite iterations then my terminating function satisfies the functional form conditions. I define this as the targeting strategy  $f$ . If this takes infinite iterations then I use the limit of the subsequence of the functions (which must exist because my choice of a function is bounded). Let me call this function  $\widehat{f}$ .  $\widehat{f}$  would not have any essential or infinite discontinuities because  $\widehat{f}$  is bounded and the slope of  $\widehat{f}$  is bounded from below. I eliminate all the removable discontinuities from  $\widehat{f}$  and revalue my jump discontinuities by C4 and C5. I call this function  $f$ , which satisfies the functional form conditions and C1-5. □

---

*Proof of Proposition 18.* This follows directly from Lemmas 3, 4, and 5. □

---

**Lemma 6.** *If  $b_1 < b$  and there exists a targeting strategy  $\tilde{f}$  satisfying C1-5 that would induce consumers to click, then the advertising platform would choose a targeting strategy such that  $b_c = b$ .*

*Proof of Lemma 6.* By way of contradiction suppose not. Suppose  $b > b_1$  and there exists a targeting strategy  $\tilde{f}$  that would induce consumers to click satisfying C1-5. Suppose there exists a click maximizing targeting strategy  $f$  satisfying C1-5 and  $b_c = \int_{p^*}^1 (p - p^*)f(p)dp \neq b$  where  $p^*$  be the price set by the merchant under  $f$ . If  $b_c < b$  then  $f$  would not induce any consumers to click on the ad, so  $b_c > b$ . By Lemma 2, the advertisement platform shows the ad to some but not all of the consumers so  $0 < \int_0^1 f(p)dp < 1$ . Because  $\int_0^1 f(p)dp < 1$  and  $f$  has only finite discontinuities, choose a price  $p' \in (0, 1)$  such that  $f(p') < 1$  and  $f$  is continuous at  $p'$ . Define  $f_\epsilon(p) \equiv f(p) * 1\{p \in [0, p' - \epsilon]\} + 1\{p \in (p' - \epsilon, p' + \epsilon]\} + f(p) * 1\{p \in (p' + \epsilon, 1]\}$   $\forall \epsilon \in (\max\{p', 1 - p, \}, 0)$ . Let  $p_\epsilon$  be the price set by the merchant under  $f_\epsilon$ . Because  $f$  is continuous at  $p'$ , choose an  $\epsilon > 0$  such that  $\int_{p_\epsilon}^1 (p - p_\epsilon)f_\epsilon(p)dp > b$ . Consumers would still be induced to click under  $f_\epsilon$  by (CCC) and  $\int_0^1 f_\epsilon(p)dp > \int_0^1 f(p)dp$ . Thus the advertisement platform strictly prefers the feasible targeting strategy  $f_\epsilon$  to  $f$ .  $\square$

*Proof of Proposition 19.* Proof by parts:

Part 1. When  $b \leq b_1$ :  $f(p) = 1\{p \in [0, 1]\}$ .

Define  $b_1 \equiv \int_{p^m}^1 (p - p^m)dp = 1/8$ , where  $p^m = 1/2$  is the price set by the merchant facing the demand curve  $(1 - p)1\{p \in [0, 1]\}$ . By Lemma 1, if  $b \leq b_1$ , then the unique optimal targeting strategy is  $f(p) = 1\{p \in [0, 1]\}$ .

Part 2. The search costs  $b > b_1$  such that there exists a targeting strategy  $\tilde{f}$  satisfying C1-5 that would induce consumers to click form a contiguous interval with an infimum of  $b_1$ .

Define  $b_s \equiv b_1/[1 - p^m]$ , where  $p^m = 1/2$  is the price set by the merchant facing the demand curve  $(1 - p)1\{p \in [0, 1]\}$ . By (CCC), if the search cost  $b$  were less than or equal to  $b_s$ , then the consumers would click under the targeting strategy

$f(p) = 1\{p \in [p^m, 1]\}$ . Because  $b_s > b_1$ , there exists search costs  $b > b_1$  for which the targeting strategy  $m = f$  does not satisfy (CCC) and at least one other targeting strategy  $\tilde{f}$  satisfies (CCC).

If the ad platform has a strategy that can induce some consumers to click for search cost  $b'$ , then it can use the same strategy to induce some consumers to click for search cost  $b'' < b'$ . Therefore, by Lemma 2, the search costs  $b > b_1$  such that there exists a targeting strategy  $\tilde{f}$  satisfying C1-5 that would induce consumers to click form a contiguous interval with an infimum of  $b_1$ .

*Part 3.* Given any  $\underline{p} \in (0, p^*(\underline{p}, \widehat{p}))$  and any  $\widehat{p} \in [p^*(\underline{p}, \widehat{p}), p_1)$  there exists a  $\underline{p}' \in (0, \underline{p})$  and a  $\widehat{p}' \in (\widehat{p}, p_1)$  satisfying  $\bar{F}(\underline{p}, \widehat{p}) = \bar{F}(\underline{p}', \widehat{p}')$  and  $b_c(\underline{p}', \widehat{p}') > b_c(\underline{p}, \widehat{p})$ .

Define  $f_{(\underline{p}, \widehat{p})}(p) \equiv 1\{p \in [\underline{p}, p^*(\underline{p}, \widehat{p})]\} + (p^*/p)^2 1\{p \in (p^*(\underline{p}, \widehat{p}), \widehat{p}]\} + 1\{p \in (\widehat{p}, 1]\}$ , where  $p^*(\underline{p}, \widehat{p})$  is the rationally expected price under  $f_{(\underline{p}, \widehat{p})}$ . Define  $\bar{F}(\underline{p}, \widehat{p}) \equiv \int_0^1 f_{(\underline{p}, \widehat{p})}(p) dp$ . Note that  $f_{(\underline{p}, \widehat{p})}$  satisfies C1-5 for any  $\underline{p} \in (0, p^*(\underline{p}, \widehat{p}))$  and any  $\widehat{p} \in [p^*(\underline{p}, \widehat{p}), p_1)$ .

I will begin by analyzing the case where  $\underline{p} < p^*(\underline{p}, \widehat{p})$ . By (MPMC) when  $\widehat{p} > 1/2 = p^m$ , I have  $p^* = \int_{p^*}^{\widehat{p}} [(p^*/p)^2] dp + 1 - \widehat{p}$ , therefore  $p^*(\underline{p}, \widehat{p}) = \sqrt{\widehat{p}(1 - \widehat{p})}$ . Note that changing  $\underline{p}$  does not change  $p^*(\underline{p}, \widehat{p})$ , because the merchant would not choose to sell to those consumers anyway, so I write  $p^*(\widehat{p}) = \sqrt{\widehat{p}(1 - \widehat{p})}$ . Also note that if  $\widehat{p} \leq 1/2$  then I would not have a targeting strategy in the form given by Proposition 18, because the merchant would set  $p^* = 1/2$ . Further note that when  $\underline{p} < p^*(\underline{p}, \widehat{p})$ , I have  $\bar{F}(\underline{p}, \widehat{p}) = 2p^*(\underline{p}, \widehat{p}) - \underline{p} = 2\sqrt{\widehat{p}(1 - \widehat{p})} - \underline{p}$ .

The advertisement platform can increase the expected benefit  $b_c$  of clicking on the ad in two ways: through increasing  $\underline{p}$  and through increasing  $\widehat{p}$ . Doing so decreases the mass  $\bar{F}(\underline{p}, \widehat{p})$  of consumers clicking. The decrease in the mass from increasing one choice variable can be offset by the increase in the mass from decreasing the other choice variable. The effect of changing the expected benefit  $b_c$  of clicking through increasing  $\underline{p}$  (and holding  $\widehat{p}$  constant) is given in equation (5) and through increasing  $\widehat{p}$  (and holding  $\underline{p}$  constant) is given in equation (6).

$$\left. \frac{\partial b_c}{\partial \bar{F}(\underline{p}, \underline{\widehat{p}})}(\underline{p}, \underline{\widehat{p}}) \right|_{\text{through } \underline{p}} = \frac{\partial b_c}{\partial \underline{p}}(\underline{p}, \underline{\widehat{p}}) \left/ \frac{\partial \bar{F}(\underline{p}, \underline{\widehat{p}})}{\partial \underline{p}}(\underline{p}, \underline{\widehat{p}}) \right. = -\frac{CS(\underline{p}, \underline{\widehat{p}})}{\bar{F}(\underline{p}, \underline{\widehat{p}})^2} \quad (5)$$

$$\text{where } CS(\underline{p}, \underline{\widehat{p}}) \equiv \int_{p^*(\widehat{p})}^1 (p - p^*(\widehat{p})) f_{(\underline{p}, \underline{\widehat{p}})}(p) dp = \frac{\widehat{p}(1-\widehat{p})}{2} \ln\left(\frac{\widehat{p}}{1-\widehat{p}}\right) + \frac{(1-\widehat{p})^2}{2}$$

$$\text{Note: } b_c = CS(\underline{p}, \underline{\widehat{p}}) / \bar{F}(\underline{p}, \underline{\widehat{p}})$$

$$\begin{aligned} \left. \frac{\partial b_c}{\partial \bar{F}(\underline{p}, \underline{\widehat{p}})}(\underline{p}, \underline{\widehat{p}}) \right|_{\text{through } \widehat{p}} &= \frac{\partial b_c}{\partial \widehat{p}}(\underline{p}, \underline{\widehat{p}}) \left/ \frac{\partial \bar{F}(\underline{p}, \underline{\widehat{p}})}{\partial \widehat{p}}(\underline{p}, \underline{\widehat{p}}) \right. \\ &= \frac{\partial CS(\underline{p}, \underline{\widehat{p}}) / \partial \widehat{p}}{\bar{F}(\underline{p}, \underline{\widehat{p}}) * \partial \bar{F}(\underline{p}, \underline{\widehat{p}}) / \partial \widehat{p}} - \frac{CS(\underline{p}, \underline{\widehat{p}})}{\bar{F}(\underline{p}, \underline{\widehat{p}})^2} \end{aligned} \quad (6)$$

Here,  $CS(\underline{p}, \underline{\widehat{p}}) / \bar{F}(\underline{p}, \underline{\widehat{p}})^2$  is the increase in  $b_c$  through decreasing the mass  $\bar{F}(\underline{p}, \underline{\widehat{p}})$  of consumers clicking by increasing  $\underline{p}$ . Because  $p^*(\widehat{p})$  and  $f_{(\underline{p}, \underline{\widehat{p}})}(p)$  above  $p^*(\widehat{p})$  does not depend on  $\underline{p}$ , changing  $\underline{p}$  does not affect the Consumer Surplus  $CS[\underline{p}, \underline{\widehat{p}}]$ , so changing  $\underline{p}$  only affects  $b_c$  through decreasing the mass  $\bar{F}(\underline{p}, \underline{\widehat{p}})$ . This affect is  $CS(\underline{p}, \underline{\widehat{p}}) / \bar{F}(\underline{p}, \underline{\widehat{p}})^2$ .

Yet, changing  $\widehat{p}$  changes  $p^*$ , so changing  $\widehat{p}$  changes  $b_c$  through changing  $\bar{F}(\underline{p}, \underline{\widehat{p}})$  and  $CS(\underline{p}, \underline{\widehat{p}})$ . Equation (7) decomposes  $\partial CS(\underline{p}, \underline{\widehat{p}}) / \partial \widehat{p}$ .

$$\frac{\partial CS(\underline{p}, \underline{\widehat{p}})}{\partial \widehat{p}} = PE(\widehat{p}) - Y(\widehat{p}) - Z(\widehat{p}) \quad (7)$$

$$\text{where } PE(\widehat{p}) \equiv -(1 - \widehat{p}) \frac{\partial p^*(\widehat{p})}{\partial \widehat{p}} = -\frac{(1-\widehat{p})(1-2\widehat{p})}{2\sqrt{\widehat{p}(1-\widehat{p})}} > 0$$

$$\begin{aligned} \text{where } Y(\widehat{p}) &\equiv - \int_{p^*(\widehat{p})}^{\widehat{p}} \frac{\partial}{\partial \widehat{p}} [(p - p^*(\widehat{p})) \frac{(p^*(\widehat{p}))^2}{p^2}] dp \\ &= -\frac{3}{2}(1 - 2\widehat{p}) \left( \frac{1}{3} \ln \frac{\widehat{p}}{1-\widehat{p}} + \sqrt{\frac{1-\widehat{p}}{\widehat{p}}} - 1 \right) \end{aligned}$$



$$\text{and where } Z(\widehat{p}) \equiv (\widehat{p} - p^*(\widehat{p}))(1 - \frac{(p^*(\widehat{p}))^2}{\widehat{p}^2}) = -(1 - 2\widehat{p})(1 - \sqrt{\frac{1-\widehat{p}}{\widehat{p}}}) > 0$$

Here  $PE(\widehat{p})$  is the *price effect* on all consumers with reservation prices above  $\widehat{p}$ . It reflects how much each consumer with a reservation price above  $\widehat{p}$  will benefit from the lowering of the price set by the merchant.  $Y(\widehat{p})$  is the *infra-marginal consumer loss effect*. By increasing  $\widehat{p}$ , the ad platform is inducing the merchant to choose a lower price  $p^*(\widehat{p})$ . By lowering  $p^*(\widehat{p})$ , the constant  $(p^*(\widehat{p}))^2$  is lower. This in turn lowers the ad platforms choice of  $f_{(\underline{p}, \widehat{p})}$  between  $p^*(\widehat{p})$  and  $\widehat{p}$ .  $Z(\widehat{p})$  is the *marginal consumer loss effect*. By increasing  $\widehat{p}$ , the advertisement platform is not advertising to some consumers with reservation prices  $r = \widehat{p}$ .  $Z(\widehat{p})$  captures the effect of the loss of the advertising to these consumers on the consumer surplus.

Note that this shows that  $\partial CS(\underline{p}, \widehat{p})/\partial \widehat{p}$  is independant of  $\underline{p}$ .

As  $\widehat{p}$  approaches  $p^m = 1/2$ : the price effect  $PE(\widehat{p})$ , the infra-marginal consumer loss effect  $Y(\widehat{p})$  and the marginal consumer loss effect  $Z(\widehat{p})$  converge to zero. And  $PE'(\widehat{p})$  converges to 1, while  $Y'(\widehat{p})$  and  $Z'(\widehat{p})$  converge to zero. Therefore for small values of  $\widehat{p}$ ,  $\partial CS(\underline{p}, \widehat{p})/\partial \widehat{p}$  is strictly greater than zero.

As  $\widehat{p}$  approaches 1: the price effect  $PE(\widehat{p})$  converges to zero, the infra-marginal consumer loss effect  $Y(\widehat{p})$  goes to infinity, and the marginal consumer loss effect  $Z(\widehat{p})$  goes to 1. Thus  $\partial CS(\underline{p}, \widehat{p})/\partial \widehat{p}$  goes to negative infinity.

Because  $PE(\widehat{p})$ ,  $Y(\widehat{p})$  and  $Z(\widehat{p})$  are continuous with respect to changes in  $\widehat{p}$ ,  $\partial CS(\underline{p}, \widehat{p})/\partial \widehat{p}$  is continuous with respect to changes in  $\widehat{p}$ . Therefore by the Intermediate Value Theorem, there exists at least one  $p_1 \in (1/2, 1)$  such that  $\partial CS(\underline{p}, p_1)/\partial \widehat{p} = 0$ . Setting (7) equal to zero, I have  $p_1 = e/(1 + e) \approx 0.731$ .

Therefore when  $\underline{p} < p^*(\underline{p}, \widehat{p})$  and  $\widehat{p}$  is less than  $p_1 = e/(1 + e)$ , the ad platform can increase  $b_c$ , while holding the mass  $\bar{F}$  of consumers clicking constant by increasing  $\widehat{p}$  while decreasing  $\underline{p}$ .

Part 4. Given any  $\underline{p} = p^*(\underline{p}, \widehat{p})$  and any  $\widehat{p} \in [p^*(\underline{p}, \widehat{p}), p_1)$  there exists a  $\underline{p}' \in (0, \underline{p})$  and a  $\widehat{p}' \in (\widehat{p}, p_1)$  satisfying  $\bar{F}(\underline{p}, \widehat{p}) = \bar{F}(\underline{p}', \widehat{p}')$  and  $b_c(\underline{p}', \widehat{p}') > b_c(\underline{p}, \widehat{p})$ .

Define  $f_{(\underline{p}, \widehat{p})}(\underline{p})$ ,  $p^*(\underline{p}, \widehat{p})$ , and  $\bar{F}(\underline{p}, \widehat{p})$  as in Part 3.

This uses the argument as Part 3, with a caveat: I need to show that the ad platform would never set  $\underline{p}$  above the price  $p^*(\underline{p}, \widehat{p})$  when the FOC given by (2) does not hold. Obviously if the FOC holds then it is true by taking the limit of the argument given in Part 3.

If  $p^* < \int_{\underline{p}^*}^1 f_{(\underline{p}, \widehat{p})}(p) dp$ , then the merchant would set a price  $p^* > \underline{p}$ . Therefore consider when  $\underline{p} = p^*(\underline{p}, \widehat{p})$  and  $\underline{p} > \int_{\underline{p}}^1 f_{(\underline{p}, \widehat{p})}(p) dp$ . Choose an  $\epsilon > 0$  small enough so  $\underline{p} - \epsilon = p^*(\underline{p} - \epsilon, \widehat{p})$  and  $(\underline{p} - \epsilon) > \int_{\underline{p} - \epsilon}^1 f(p) dp$ . Targeting strategy  $f_{(\underline{p} - \epsilon, \widehat{p})}$  would produce a higher  $b_c$  than  $f_{(\underline{p}, \widehat{p})}$  and  $\bar{F}(\underline{p} - \epsilon, \widehat{p}) > \bar{F}(\underline{p}, \widehat{p})$ .

Therefore when  $\underline{p} = p^*(\underline{p}, \widehat{p})$ , the ad platform can increase  $b_c$ , while holding the mass  $\bar{F}$  of consumers clicking constant by choosing a smaller  $\underline{p}$  and a larger  $\widehat{p}$  to compensate for the  $\bar{F}$  gained due to a smaller  $\underline{p}$ .

Part 5. When  $b_1 < b \leq b_2$ :  $f(p) = 1\{p \in [0, p^*]\} + (p^*/p)^2 1\{p \in (p^*, \widehat{p}]\} + 1\{p \in (\widehat{p}, 1]\}$  where  $\widehat{p} \in (p^m, p_1]$

Define  $f_{(\underline{p}, \widehat{p})}(p)$ ,  $p^*(\widehat{p})$ , and  $\bar{F}(\underline{p}, \widehat{p})$  as in Part 3.

$$\begin{aligned} \text{Define } b_2 &\equiv \int_{p^*(p_1)}^1 (p - p^*(p_1)) f_{(0, p_1)}(p) dp / \bar{F}(0, p_1) \\ &= \frac{1}{4\sqrt{e}} \approx 0.152 \end{aligned}$$

$$\text{Note: } p^*(p_1) = \frac{\sqrt{e}}{1+e}$$

By (CCC), if the search cost  $b \leq b_2$ , then the consumers would click under the targeting strategy  $f_{(0, p_1)}$ .

By Parts 3 and 4, if the search cost  $b$  satisfies  $b_1 < b_2 < b_2$ , then the click maximizing targeting strategy that gives just enough expected benefit  $b_c$  to get consumers to click would be for  $\underline{p}$  to be as low as possible; in this case,  $\underline{p} = 0$ .

Part 6. Given any  $\underline{p} \in (0, p^*(\underline{p}, \widehat{p}))$  and any  $\widehat{p} \in (p_1, p_2]$  there exists a  $\underline{p}' \in (\underline{p}, p^*(\underline{p}', \widehat{p}'))$  and a  $\widehat{p}' \in (p_1, \widehat{p})$  satisfying  $\bar{F}(\underline{p}, \widehat{p}) = \bar{F}(\underline{p}', \widehat{p}')$  and  $b_c(\underline{p}', \widehat{p}') > b_c(\underline{p}, \widehat{p})$ .

This follows directly from the argument given in Part 3. For  $\widehat{p} > p_1$ ,  $\partial CS(\underline{p}, \widehat{p}) / \partial \widehat{p} < 0$ .

Part 7. When  $b_2 < b \leq b_3$ :  $f(p) = 1\{p \in [\underline{p}, p^*]\} + (p^*/p)^2 1\{p \in (p^*, p_1]\} + 1\{p \in (p_1, 1]\}$  where  $\underline{p} \in (0, p^*]$ .

Define  $f_{(\underline{p}, \widehat{p})}(p)$ ,  $p^*(\widehat{p})$ , and  $\bar{F}(\underline{p}, \widehat{p})$  as in Part 3.

$$\begin{aligned} \text{Define } b_3 &\equiv \int_{p^*(p_1)}^1 (p - p^*(p_1)) f_{(0, p_1)}(p) dp / \bar{F}(0, p_1) \\ &= \frac{1}{2\sqrt{e}} \approx 0.303 \end{aligned}$$

$$\text{Note: } p^*(p_1) = \frac{\sqrt{e}}{1+e}$$

By (CCC), if the search cost  $b \leq b_3$ , then the consumers would click under the targeting strategy  $f_{(p^*(p_1), p_1)}$ .

By Part 6, if the search cost  $b$  satisfies  $b_2 \leq b \leq b_3$ , the click maximizing targeting strategy would have  $\widehat{p} = p_1$ . The price cutoff  $\underline{p}$  would be just low enough to give the expected benefit  $b_c$  equal to the search cost  $b$ .

Part 8. When  $b_3 < b \leq b_4$ :  $f(p) = (p^*/p)^2 1\{p \in [p^*, \widehat{p}]\} + 1\{p \in (\widehat{p}, 1]\}$  where  $\widehat{p} \in (p_1, p_2]$ .

Define  $f_{(\underline{p}, \widehat{p})}(p)$ ,  $p^*(\widehat{p})$ , and  $\bar{F}(\underline{p}, \widehat{p})$  as in Part 3.

After increasing  $\widehat{p}$  to  $p_1$  and increasing  $\underline{p}$  to  $p^*$ , then the expected benefit of clicking,  $b_c$ , still might be large enough to induce the consumer to click on the ad. As long as the equation (6) is negative, the ad platform can still increase the expected benefit from clicking on the ad by increasing  $\widehat{p}$ . Increasing  $\widehat{p}$  further still increases the expected benefit from clicking on the ad through decreasing the mass of consumers clicking on the ad. When  $|\partial CS(\underline{p}, \widehat{p}) / \partial \widehat{p}| > |\partial \bar{F}(\underline{p}, \widehat{p}) / \partial \widehat{p} * CS(\underline{p}, \widehat{p}) / \bar{F}(\underline{p}, \widehat{p})|$ , then the advertisement platform can still increase the expected benefit of clicking on the ad by increasing  $\widehat{p}$ .

I have established that  $\partial CS(\underline{p}, \widehat{p}) / \partial \widehat{p}$  is zero when  $\widehat{p} = p_1$  and negative infinity as

$\widehat{p} \rightarrow 1$ . Because of the continuity of  $\partial CS(\underline{p}, \widehat{p})/\partial \widehat{p}$  and  $\partial \bar{F}(\underline{p}, \widehat{p})/\partial \widehat{p} * CS(\underline{p}, \widehat{p})/\bar{F}(\underline{p}, \widehat{p})$ , there must exist a  $p_2 > p_1$  such that  $\partial CS(\underline{p}, p_2)/\partial \widehat{p} = \partial \bar{F}(\underline{p}, p_2)/\partial \widehat{p} * CS(\underline{p}, p_2)/\bar{F}(\underline{p}, p_2)$ . Setting  $\underline{p} = p^* = \sqrt{p(1-p)}$  and setting equation (6) equal to zero, I find that  $p_2$  solves  $(1 + p_2)/p_2 = \ln[p_2/(1 - p_2)]$ . Therefore  $p_2 \approx .893$ .

$$\begin{aligned} \text{Define } b_4 &\equiv \int_{p^*(p_2)}^1 (p - p^*(p_2)) f_{(p^*(p_2), p_2)}(p) dp \bar{F}(p^*(p_2), p_2) \\ &= \frac{\sqrt{p_2(1-p_2)}}{2} \left( \ln \frac{p_2}{1-p_2} + \frac{1-p_2}{p_2} \right) \\ &\approx 0.346 \end{aligned}$$

By (CCC), if the search cost  $b \leq b_4$ , then the consumers would click under the targeting strategy  $f_{(p^*(p_2), p_2)}$ .

By the argument in Part 3, if the search cost  $b$  satisfies  $b_3 \leq b \leq b_4$ , the click maximizing targeting strategy would want to lower  $\widehat{p}$  to increase the expected benefit  $b_c$  to  $b$ .

Part 9. When  $b > b_4$ : no consumer would ever click on the ad so any  $m$  is optimal.

If  $b > b_4$  then it is impossible for the advertising platform to induce the consumers to click on the ad. If there were a strategy  $f''$  then there would be a strategy of the form given in Proposition 18 that could induce consumers to click on the ad. But my comparative static results in Part 2 rule out this possibility.  $\square$

**Lemma 7.** When  $b \leq b_5 \equiv 1/4$ , then a merchant-profit-maximizing targeting strategy is  $f(p) = 1\{p \in [p^m, 1]\}$ .

*Proof of Lemma 7.* Suppose not. Let  $\widetilde{f}$  satisfying C1-5 maximize the merchant's profits. Let  $\widetilde{p}$  be the price set by the merchant under  $\widetilde{f}$ .

Under the demand function  $1 - p$ , the merchant would choose to set its price as

$p^m = 1/2$ . Therefore by the Weak Axiom of Profit Maximization:  $p^m(1-p^m) \geq \tilde{p}(1-\tilde{p})$ . I have that  $1 - \tilde{p} \geq \int_{\tilde{p}}^1 \tilde{f}(p)dp$ , by C1. Thus  $p^m(1 - p^m) \geq \tilde{p} \int_{\tilde{p}}^1 \tilde{f}(p)dp$ .

Note that consumers would click under the targeting strategy  $f(p) = 1\{p \in [p^m, 1]\}$  by (CCC). And that this gives the merchant the profit  $p^m(1 - p^m) = b_5 = 1/4$ .  $\square$

*Proof of Proposition 20.* Define  $b_5 \equiv \int_{p^m}^1 (p - p^m)dp / (1 - p^m) = 1/8$ , where  $p^m$  is the price set by the merchant facing the demand curve  $1 - p$ . And define  $b_4$  as in Part 8 of the proof of Proposition 19.

When  $b \leq b_5$ , consumers would click under the targeting strategy  $f_0(p) \equiv 1\{p \in [p^m, 1]\}$  by (CCC). This maximizes the merchants profit by Lemma 7.

When  $b_5 \leq b \leq b_4$ , consumers would not click under  $f_0$ , so the ad platform has to choose a targeting strategy that commits the merchant to a lower price to encourage them to click. By Lemma 6, the ad platform would do so to make  $b_c = b$ . By Proposition 18, the ad platform would do so with a targeting strategy of the form (a) or (b) of Proposition 18. Raising  $\underline{p}$  would not affect the merchant's profit as long as  $\underline{p} < p^*$ , so the ad platform would set  $\underline{p}$  equal to the rationally expected price  $p^* = \widehat{p}(1 - \widehat{p})$  of the merchant. But doing so would not be enough (because  $f_0$  does not induce consumers to click), so the ad platform would raise  $\widehat{p}$  above  $p^m$ . Because  $b \leq b_4$ , it is possible to induce consumers to click with a high enough  $\widehat{p}$ .

When  $b > b_4$ , by the argument in Part 9 of the proof of Proposition 19, it is impossible for the ad platform to induce consumers to click.  $\square$

## B Functional Form Conditions

**C1**  $f(p) \in [0, 1]$  for all  $p \in [0, 1]$ .

**C2**  $f$  is not discontinuous at an infinite number of points.

**C3** for all  $p \in [0, 1]$ , the right-hand and left-hand limits of  $f(p)$  exists and are finite.

**C4**  $\lim_{p \rightarrow \underline{p}^+} f(p) = f(\underline{p})$  where  $\underline{p}$  is the smallest  $p \in [0, 1]$  such that  $f(p) \neq 0$ .

**C5**  $\lim_{p \rightarrow p'^-} f(p) = f(p')$  for all  $p' \neq \underline{p}$ .

C1 prevents the advertising platform from showing the ad to more consumers than those who exist. C2 is necessary so we can split  $f$  into a finite number of continuous intervals. C3 ensures that  $f$  is integrable and that there are no essential discontinuities.<sup>95</sup> C4 and C5 prevent removable discontinuities and restrict the jump discontinuities to be of the most convenient direction for my mathematical proofs. Note that C4 and C5 are not that restrictive because they do not restrict the mass of consumers shown an ad along any interval.

---

<sup>95</sup>C2 and C3 are not the same as  $f$  being measurable. Consider the function  $f'$  which is one for all rational numbers and zero for all irrational numbers.  $f'$  is measurable, but violates both C2 and C3