Theoretical and Empirical Essays on Personalized Advertising

Alicia Baik
St. Paul, Minnesota

M.A. Economics, University of Virginia, 2012
M.S. Foreign Service, Georgetown University, 2008
B.A. Political Science, University of Minnesota, 2005

A Dissertation 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
May, 2017

Abstract

Personalized advertising is becoming one of the most defining features of the twenty-first century marketplace. The two essays in this dissertation analyze two important theoretical and empirical dimensions of personalized advertising.

The first essay develops a model of costly advertising and price competition among n quality-cost differentiated firms in which the individual consumer is the basic unit of analysis. Strategies involve mixing over both prices and whether to advertise. In equilibrium, only the top two firms advertise, earning "Bertrand-like" profits. Welfare losses initially rise then fall with the ad cost, with losses due to excessive advertising and sales by the "wrong" firm. Additionally, taking the limit of advertising costs to zero selects the equilibrium where the most efficient firm prices (with probability one) at the cost of its closest rival.

The second essay develops a model of personalized advertisements and their impact on customers' purchase paths in a context of limited consideration sets. Personalized advertisements are more likely to be considered relative to generic advertisements. The consideration set expands from the status quo shopping list to the product in the advertisement when it is considered. Consideration of products outside a customer's status quo purchase path has the greatest expected increase on her consideration set and purchase basket while products along her purchase path have limited impact. I empirically test these predictions and find that personalized campaigns increase sales in the promoted department and in the store overall. Campaigns for regularly purchased products have little impact on sales in the department or store, even though redemption rates are higher. Generic campaigns have the least impact on sales.

JEL Classifications: L21, M31, M37

Keywords: Targeted Advertising; Empirical Industrial Organization; Bounded Rationality; Purchase Path

Acknowledgements

The summation of the work in this dissertation reflects the encouraging investment from many individuals who have shaped me deeply over the course of my graduate work. It is difficult to capture the degree to which these individuals have sparked my passion for research and refined my thoughts, but I wish to thank them.

I wish to thank Federico Ciliberto for first believing in me when I was beginning to spread my wings as a researcher. Your motivation and inspiration gave me confidence to push hard in pursuit of innovative research. Additionally, your attention to detail helped ground me on my empirical precision. Thank you also for having the wisdom to know when to push and when to give space as I navigated personal challenges during my graduate work.

Also, thank you, Simon Anderson, for being a wonderful mentor and encouraging my passion for research. I will always cherish our afternoon teas and lunch discussions about the intricacies of personalized advertising. You are one of the most generous and creative people I know, and I am so thankful to have had the privilege of working so closely with you in my research.

Thank you, Nathan Larson, for taking an interest in personalized advertising from the beginning. My fledgling ideas are far more substantive because of your collaboration. You are a rare blend of intelligence and humility, which makes working with you such a pleasure.

I also want to thank Karim Chalak for having the patience to walk through my econometric model with me. You helped me think more carefully and precisely about my empirical research and the paper would not be what it is without your investment.

I have made some of my best friends during graduate school. I could not have survived the ups and downs of the past years without the support of Zhou Zhang and Kevin Curry, Carol and Ron Pantuck, Alicia and Steve Morales, Brian and Muriel

Russell, Keith and Lisa Goad, Stephanie Marvin, and Rachelle and Mike Keng. I also could not have finished writing the thesis without the support of my parents-in-law, Ho and Phuong-Loan Nguyen.

My parents have invested deeply into making me the person I am today. I want to thank them for pouring so much of themselves into me. Also, thank you, Halabogi, for believing in me since I was a child. I wish you could see this day after so many years of looking forward to it. Monica and Seung-Min, thank you for supporting me through so much. I am so thankful to have you as my siblings.

Finally, I would not have had the confidence nor the aspiration to pursue this research without the support and encouragement from my best friend and husband, Dao Nguyen. Thank you for first encouraging me to pursue research in this field, thank you for supporting me when I felt discouraged, and thank you for your constant confidence that I can accomplish big things. To Caleb and Tyler, you do not yet understand the contents of this dissertation, but I hope that one day it will make you proud.

Dedicated to Dao, Caleb, and Tyler.

Contents

Chapter 1

Introduction

Personalized advertising is quickly becoming one of the most defining features of the twenty-first century marketplace. As one marketer describes, true personalization is “perhaps the most difficult identity-driven marketing tactic to put into practice” because it requires “real-time responsiveness and relies on highly accurate and comprehensive customer data” (The Aisles Have Eyes, 2017). With rapid technological advancements in data collection and analysis, marketers and retailers have a greater capacity to move toward this ideal. Additionally, increasingly competitive retail marketing dynamics are pushing retailers even further toward this goal in order to differentiate themselves from competitors.

The two essays in this dissertation analyze two important theoretical and empirical dimensions of personalized advertising.

The first essay, “Personalized Pricing and Advertising: An Asymmetric Equilibrium Analysis,” coauthored with Simon Anderson and Nathan Larson, develops a model of advertising and price competition in which the individual consumer is the basic unit of analysis. The framework is motivated by firms’ ability to individually target prices based on rich purchase data merged across retailers and time. In this essay, customers are characterized by their preference profile for the products offered by different firms. While firms know that customers with these taste profiles exist, reaching them with personalized price advertisements is costly. Given a customer with a taste profile, firms simultaneously choose whether to advertise to her and if so, what individualized price to offer her. Because sending an offer is costly, equilibria involve mixed strategies over prices and the decision to advertise.

Our focus is on the asymmetric valuation case where customers value some products more than others. We find that the "best" firm always advertises and earns a rent equaling its social surplus advantage (valuation minus cost superiority) over its closest rival and that the second best firm advertises with positive probability below one, and earns zero expected profits. We also find that social efficiency falls then rises with advertisement costs, with losses due to wasteful advertisements and non-optimal purchases. These inefficiencies vanish when advertising costs go to zero or when they rise high enough to give the "best" firm a monopoly. Additionally, taking the limit of advertising costs to zero selects the equilibrium where the most efficient firm prices (with probability one) at the cost of its closest rival. Interestingly, the second-best firm makes an offer just often enough to keep the top firm from deviating to its monopoly price, providing fresh perspective on the long-standing selection problem of multiple equilibria in the classic model of Bertrand competition with asymmetric costs.

The second essay, "Personalized Advertising: A Theoretical and Empirical Analysis of the Customer Purchase Path," develops a model of personalized advertisements and their impact on customers' purchase paths. Motivated by consumers' inability to process all of the options available to them in the store and on advertisements, the model develops a framework of limited consideration sets with advertising. Personalized advertisements are more likely to be considered relative to generic advertisements. The consideration set expands from the status quo shopping list to the product in the advertisement when it is considered. When a customer considers a product outside of her status quo purchase path, she exposes herself to other localized marketing tools in the department of the advertised product, further expanding her consideration set. Products along her purchase path have little impact on the customer's consideration set apart from the product being advertised. Generic advertisements have little impact on a customer's consideration set because she is less likely to evaluate the product. The model predicts that personalized advertisements which promote products outside the customer's purchase path will have the greatest increase on customer sales.

I empirically evaluate the impact of personalized and generic advertisements on the overall purchase path of customers utilizing a unique dataset from a single unidentified national grocery retailer. The two targeting tools used by the retailer are reward and promotional coupon campaigns. Rewards offer discounts for a broad set of frequently purchased products. Promotion campaigns recommend and offer discounts for a specific set of products relevant to a targeted customer group.

As the theoretical model predicts, I find that promotional campaigns are more effective than reward campaigns at increasing customer sales. I also find that they increase sales in the departments of the promoted products, confirming the theoretical prediction that customers increase their consideration set when they leave their status quo purchase path to evaluate a new product. Reward campaigns have a higher redemption rate, but when customers do not redeem products, customers on average spend less at the store. Additionally, redemptions in the department of the promoted products increase sales by less than promoted products. Generic mailers and displays have little impact on store sales. I control for endogeneity of the targeted advertisements and selection of store visits with control functions within a sample selection model.

This dissertation is structured as follows: the two dissertation essays summarized above are presented in Chapters 2 and 3.

Chapter 2

Personalized Pricing and Advertising: An Asymmetric Equilibrium Analysis

2.1 Introduction

"*Recent advances in information technology have... made possible the instantaneous delivery of customized pricing offers to individual consumers.*" (Pricing with Precision and Impact, Boston Consulting Group 2002)

Mass marketing made possible through TV, newspapers, and billboards is increasingly evolving into individualized marketing. Firms previously limited to sending messages to heterogeneous groups of consumers (on network TV say) are now able to purchase information on *relatively* homogeneous sets of individuals from intermediary information brokers.[1] With finer levels of categorization, marketing precision is moving to the individual level. Comprehensive purchase history from various retailers can now be merged with demographic and web-site visit data to render very specific individual information on tastes, and firms can deliver individually-tailored price offers based on such information. This means that firms have the potential to

[1]We abstract away from the strategic role that such brokers may exercise in pricing information.

compete at the level of the individual consumer. As technological capacity develops and the cost of personalized pricing decreases, the potential for individualized price competition will only increase.

Motivated by these observations, in this paper we develop a model of advertising and price competition in which the individual consumer is the basic unit of analysis. A consumer is characterized by her profile of valuations for the products offered by different firms. While each firm is aware that consumers with this particular taste profile exist, reaching them with personalized-price advertisements is costly. This cost might reflect payment to a data broker to deliver the name and contact information of a consumer with this taste profile. It also reflects the "postage and handling" costs of preparing and delivering an individualized offer. (One could think of these offers as going out by text message, email, or personalized coupons in the mail, as opposed to *en masse* marketing.) Meanwhile, as in Butters (1977), Grossman and Shapiro (1984), and Stahl (1994), a consumer does not know that a product is available unless she receives an advertised offer from the firm selling it. Among the offers she receives, she chooses the one that yields her the greatest consumer surplus.[2]

Given a potential consumer with a particular taste profile, firms simultaneously choose whether to advertise to her and if so, what individualized price to offer her. We call the joint price and advertising decisions the Personalized Pricing and Advertising Model, henceforth PPAM. Because sending an offer is costly, equilibria involve mixed strategies over both prices and the decision to advertise.

Though we do not focus on it, the model accommodates the possibility that firms simultaneously compete for a broad range of consumers with different tastes. How-

[2]The possibility of consumer search is introduced in a later section of Butters (1977) and is an integral part of the model of Robert and Stahl (1993). Shaffer and Zhang (1995, 2002) and Bester and Petrakis (1995, 1996) have considered targeting by location and have included the cost of sending offers to customers. They assume offers are coarse, such as a common discount to a heterogeneous consumer group.

ever, under our assumptions about the precision of targeting, a firm's strategy with respect to one consumer is completely separable from its strategy with respect to another consumer, so competition for an entire consumer population may be treated as a collection of independent instances of our model.[3]

Our main focus is on the asymmetric valuation case in which a consumer values some products more highly than others. We find that with n firms, the $n-2$ "worst" ones sit out and do not advertise at all. The second "best" one advertises with positive probability below one, and earns zero expected profits; while the best one always advertises and earns a rent equaling its social surplus advantage (valuation minus cost superiority) over its closest rival. We also find that social efficiency falls then rises with advertisement costs, with losses due to wasteful advertisements and non-optimal purchases. These inefficiencies vanish when advertising costs go to zero or when they rise high enough to give the "best" firm a monopoly.

The pattern of our equilibrium results has some precedent in other asymmetric games with discontinuous pay-offs and (non-degenerate) mixed strategy equilibria. One point of resemblance is with the All-Pay-Auction treated in Hillman and Riley (1989) where different bidders have different values from winning. Baye, Kovenock, and de Vries (1996) present a broader set of symmetric and asymmetric combinations to this game by allowing ties in payoffs. A second prominent example is Varian's (1980) Model of Sales, extended to allow for heterogeneous numbers of "loyal"consumers across firms by Narasimham (1988) for duopoly and by Kocas and Kiyak (2006) for oligopoly.[4] In both games, there is a winner-take-all prize for the

[3]This separability is reasonable if (as we assume) a firm faces constant marginal costs to produce its product and to reach an additional consumer with a targeted ad. The latter is consistent with firms buying data about blocks of consumers (rather than individual by individual) as long as pricing is on a per-consumer basis and blocks of individuals with the same tastes are on offer.

[4]Baye, Kovenock, and de Vries (1992) find all the equilibria for the Model of Sales when all firms have the same number of loyal consumers (as in the original). In addition to the symmetric equilibrium analyzed by Varian (1980), there are also asymmetric ones. In these, at least two firms must be active: when there are only two firms in the market the symmetric equilibrium is the unique

fiercest competitor, but competing incurs costs that "losers" do not recover. In the all-pay auction, the interpretation of the prize and costs is straightforward. In the Model of Sales, the "prize" is sales to the set of informed consumers, while the cost of competing for these consumers by offering a discounted price is the foregone profit on a firm's loyal consumers.[5] In both games, only the two players with the highest win value contend the prize, and all other players choose not to (by bidding zero or not discounting, respectively). While the results in these two models and ours share a "family resemblance," the models themselves have significant differences such that no pair is formally equivalent (even when reduced to their symmetric versions). Hence, our results cannot be derived from existing ones in the literature.

By taking advertising costs to zero, we can provide a fresh perspective on the long-standing selection problem of multiple equilibria in the classic model of Bertrand competition with asymmetric costs. (That is to say, homogeneous goods, no advertising, and different (constant) marginal costs across firms.) We select the equilibrium where the most efficient firm prices (with probability one) at the cost of its closest rival. Interestingly, the second-best firm makes an offer just often enough to keep the top firm from deviating to its monopoly price.

Analysis of equilibrium price distributions in the literature frequently assumes that firms are symmetric and focuses on a symmetric mixed strategy equilibrium. We argue that the symmetric equilibrium, when it exists in our model, may be seriously misleading. First, we show the striking comparative static prediction that when a consumer views products as homogeneous, the symmetric equilibrium has

one, but not otherwise. Of particular interest for what follows in our paper is their result (Example 2, p.500) that with $n > 2$ there are equilibria with $k \geq 2$ firms symmetrically randomizing their prices and the others just charge the consumer reservation price.

[5] As clarified by Janssen and Moraga (2004), the Model of Sales is also at the heart of the literature on firm pricing and consumer search following Stahl (1989). In these search models, "informed" consumers (or "shoppers") know all prices, while others face a search cost and in equilibrium stop at the first firm sampled, and hence play the role of the "loyal" consumers. Baye and Morgan (2001) successfully expand the basic MoS framework to a two-sided market setting.

consumer surplus and welfare decreasing in the number of competing firms. This strong result stems from the indifference condition required to elicit advertising by all n firms. However, this equilibrium is not robust: with any heterogeneity in the consumer's valuations, the set of advertisers collapses down to two firms. Thus the perverse comparative static properties of the symmetric equilibrium may be seen as a symptom of this equilibrium's instability. However, with homogeneous products, the model also has many asymmetric equilibria. When we consider the limit case of heterogeneous firms as they approach homogeneity, then we select the particular asymmetric equilibrium in which only two firms are active (regardless of n) and one always advertises while the other does so with probability strictly less than one. In this limit equilibrium, the number of firms has no impact on welfare, and welfare is weakly higher (strictly for $n > 2$) than under the symmetric equilibrium.

The PPAM model of this paper can also be interpreted (by simply relabeling the ad cost as an entry cost) as a Bertrand model of pricing and (simultaneous) entry. Previous work in this vein by Sharkey and Sibley (1993) considers symmetric firms and the symmetric equilibrium (we extend their model in section 6.1). Their main result is that an increase in the number of potential firms stochastically raises prices. Stahl (1994) moreover shows that seller entry can decrease social surplus, using a model of price advertising that effectively bridges the Butters analysis to the Sharkey and Sibley one, and for which the limit case when advertising costs are linear in reach corresponds to our model. By contrast, we would select an asymmetric equilibrium, in which case equilibrium price distributions are unchanged when there are more potential firms.

In the following section, we describe the basic set-up of the PPAM and discuss the two key strategic variables: individualized price distributions and advertising. In Section 3 we characterize the equilibrium in terms of the offered surpluses, rolling the decision to advertise into these surplus distributions, and highlight the Bertrand limit.

In Section 4 we analyze two sources of competition-induced inefficiency: wasteful advertisements and non-optimal purchases. When firms are symmetric, the model has both symmetric and asymmetric equilibria. We evaluate both in Section 5 and show that the symmetric equilibrium has perverse comparative statics and is not robust to firm heterogeneity, indicating that the asymmetric outcome may be a more reasonable prediction. In Section 6 we argue that a number of our key results still apply if consumers have downward-sloping demand, if targeted advertising costs vary across firms, or if firms' information about consumer tastes is noisy. Section 7 concludes with a discussion of fruitful directions for future work.

2.2 Model

Each firm's problem will be separable across consumers, so we shall treat competition for an individual consumer as the basic unit of analysis. There are n single-product firms competing for the business of a single consumer who wishes to buy at most one unit from one of them. Each consumer considers the set of price offers she receives and purchases from the firm whose advertised offer gives her the greatest surplus. If she receives no ads or if none of the advertisements offer her weakly positive consumer surplus, she does not make a purchase. We assume that the consumer purchases whenever indifferent and randomizes if she is indifferent among several firms. Aside from this choice, the consumer has no strategic role in the game.

Let r_i be the consumer's individual reservation price for the product offered by firm i. (We assume that r_i is measured relative to some outside option which is normalized to zero.) Let p_i be the price offered to this consumer, and so $\sigma_i = r_i - p_i$ represents the consumer surplus offered by Firm i. This variable will allow us to conflate the advertising and pricing decisions into a single statistic, and we will show that in equilibrium all active firms have the same support for the consumer surplus they deliver.

As in the classic Butters (1977) model, a consumer is unaware of the availability of Firm i's product unless she receives an advertisement with a price offer from Firm i. Advertising is costly: each firm decides whether to inform the consumer about an individualized price at cost A; alternatively, a firm can choose not to advertise. In anticipation of mixed strategies, let a firm's cumulative price distribution conditional on advertising be $F_i(p)$. Thus, a strategy for Firm i is a pair $\{a_i, F_i\}$ where a_i is the probability that Firm i advertises. Firms choose these strategies simultaneously.

A firm that does not advertise earns zero profit, while if Firm i advertises price p_i, its expected profit is given by

$$\pi_i(p_i) = (p_i - c_i)\Pr\big(i \text{ sells} \mid p_i\big) - A$$

where c_i is the marginal cost of product i. Firms seek to maximize expected profit. As this is a static model of complete information, the solution concept is simply Nash equilibrium.

The social surplus from a purchase at Firm i is the difference between the consumer's reservation value and the cost of production, $s_i = r_i - c_i$. Throughout most of the paper, we assume that different firms offer different social surpluses, with no ties. A discussion of equilibria when some or all of the products offer the same surplus is reserved for Section 5. Given this assumption, we choose to label firms in decreasing order of social surplus: $s_1 > s_2 > ... > s_n$. Define the value advantage of Firm i over Firm j to be the difference $\Delta_{ij} = s_i - s_j$, which is strictly positive whenever $i < j$. If $A > s_2$ then there will either be no advertising and no sale in equilibrium (if we also have $A > s_1$), or else Firm 1 will hold a monopoly over the consumer. Thus, the interesting case, which we henceforth consider, is that $A \leq s_2$. Thus, at least two firms would want to advertise if they could earn monopoly profits by doing so.

2.3 Characterization of Equilibrium

2.3.1 Participation and Profits

We claim that any equilibrium has the features that the top firm advertises with probability one, the next best firm advertises with positive probability less than one, and no other firm advertises. Furthermore, the top firm earns expected profit equal to Δ_{12}, its surplus advantage over its closest rival, while the second-ranked firm earns 0. These two firms price in mixed strategies; their price supports are such that the consumer faces the same range of possible surplus offers at either firm. The highest price offered by each firm leaves the consumer with zero surplus, while the lowest price ever offered by each firm leaves the consumer with $s_2 - A$, the full social surplus from a sale (net of the ad cost) at Firm 2. The top firm advertises its monopoly price with positive probability; that is to say, its price distribution has an atom at its upper bound, the consumer's reservation value. The distribution of the second-best firm has no atoms, and (with the exception above) both firms' prices follow Pareto distributions.

We proceed through a series of lemmas to establish these results. We show first that if any firm is advertising, then all higher ranked firms advertise as well. Next we show that at most one firm makes strictly positive profits in equilibrium. Third, the profits of all active firms are strictly ranked in the natural order. Fourth, using these results, we establish that at most the top two firms are active. Then (fifth and sixth), we show that the second-ranked firm does advertise, with probability less than one, while the top firm advertises with certainty. These results imply (Lemma 7) that equilibrium profits are Δ_{12} for the top firm and zero for all others. These facts make a full characterization of equilibrium strategies relatively straightforward; this characterization is given in Proposition 1. In referring to (candidate) equilibrium profits for a firm i, for brevity we will often simply write π_i (without reference to the particular price offer) rather than $\pi_i(p_i)$ since Firm i will typically be indifferent over a range of optimal prices.

Lemma 1 *In any equilibrium, if $a_i > 0$, and $j < i$, then $a_j > 0$.*

Proof. Suppose toward a contradiction that there is an equilibrium with $a_j = 0$, $a_i > 0$, and $j < i$. Let $\hat{p}_i$ be the lowest price that Firm i ever advertises. (To be careful, we should have $\hat{p}_i$ be the infimum of Firm i's prices support, which may be degenerate.) Let $\hat{\pi}_i = \hat{q}_i(\hat{p}_i - c_i) - A$ be Firm i's expected profit when offering $\hat{p}_i$ and $\hat{q}_i > 0$ its probability of making a sale. (Again, for extra care, the limiting profit and sale probability as $p_i \to \hat{p}_i$.) Note that $\hat{\pi}_i \geq 0$, otherwise Firm i would not be active. Let $\hat{p}_j = \hat{p}_i + \left(r_j - r_i\right)$ be the price from Firm j that would make the consumer equally well off as price $\hat{p}_i$ at Firm i. If Firm j were to advertise price $\hat{p}_j - \varepsilon$, its probability of making a sale would be no less than $\hat{q}_i$, say $\hat{q}_j \geq \hat{q}_i$, and so it would earn profit

$$\begin{aligned}
\hat{\pi}_j^{\varepsilon} &= \left(\hat{p}_j - c_j - \varepsilon\right)\hat{q}_j - A \\
&= (\hat{p}_i - c_i)\,\hat{q}_j + \left(\Delta_{ji} - \varepsilon\right)\hat{q}_j - A \\
&\geq \hat{\pi}_i + \left(\Delta_{ji} - \varepsilon\right)\hat{q}_j
\end{aligned}$$

But then because $\Delta_{ji} > 0$, for ε small enough, $\hat{\pi}_j^{\varepsilon} > \hat{\pi}_i \geq 0$, so Firm j could earn strictly positive expected profit by deviating to advertising price $\hat{p}_j - \varepsilon$. ■

Lemma 2 *In equilibrium, at most one firm makes a strictly positive expected profit.*

Proof. Suppose to the contrary that there is an equilibrium with $\pi_i > 0$ and $\pi_j > 0$ for some firms i and j, with $j < i$. Then neither firm is indifferent between advertising and not advertising (as the latter earns zero profit), so both firms must be advertising with probability one. Let $\hat{p}_i$ be the supremum over all prices ever offered by i, with

$\hat{p}_j$ the supremum over prices offered by j. We must have $\hat{p}_i = \hat{p}_j + \left(r_i - r_j\right)$, so that the consumer is indifferent between prices $\hat{p}_i$ and $\hat{p}_j$. (Firm i will never advertise any $p_i > \hat{p}_j + \left(r_i - r_j\right)$, as this price would lose the sale for sure, earning profit $-A$, and similarly for Firm j.) Furthermore, Firm j's strategy must place an atom at $\hat{p}_j$. (If not, then Firm i's chance of winning the sale would tend to zero for p_i sufficiently close to $\hat{p}_i$, making it unprofitable to pay to advertise such prices.) Similarly, Firm i's strategy must place an atom at $\hat{p}_i$. The firms' profit margins $\hat{p}_i - c_i$ and $\hat{p}_j - c_j$ at these upper bound prices must be strictly positive, since otherwise they could not cover the advertising cost and earn positive profits. But then because Firm i ties Firm j's atom when offering $\hat{p}_i$, it could earn a strictly higher profit by deviating to an undercutting price, contradicting the optimality of including $\hat{p}_i$ in its support. (And similarly for Firm j.) ■

Lemma 3 *If Firm i, $(i > 1)$, advertises in equilibrium, then $\pi_i < \pi_j$ for all $j < i$.*

Proof. The argument follows essentially the same lines as Lemma 1. Suppose $\hat{p}_i$ is the lowest price that Firm i ever offers in equilibrium, with profit margin $\hat{p}_i - c_i$. Firm j earns a strictly larger profit margin, $\hat{p}_j - c_j = \hat{p}_i - c_i + \Delta_{ji}$ on the price $\hat{p}_j = \hat{p}_i + \left(r_j - r_i\right)$ that would leave the consumer equally well off as buying from i at $\hat{p}_i$. By offering slightly less than $\hat{p}_j$, Firm j could sell at least as often as Firm i does at price $\hat{p}_i$, thereby earning a profit strictly greater than π_i. Firm j's equilibrium profit must be at least this good; thus $\pi_j > \pi_i$. ■

Lemma 4 *No firm other than the top two advertises in equilibrium: $a_i = 0$ for $i \geq 3$.*

Proof. If Firm $i \geq 3$ were to advertise, then Lemma 1 implies that Firms 1 and 2 would do so as well, and then Lemmas 2 and 3 imply that π_2 must be zero. But then, another application of Lemma 3 would imply that $\pi_i < 0$, so advertising with positive probability cannot actually be a best response for Firm i after all. ■

Lemma 5 *Firm 2 advertises with positive probability less than one:* $a_2 \in (0, 1)$.

Proof. If Firm 2 did not advertise at all, then Firm 1's best response would be to advertise its monopoly price, $p_1 = r_1$, with probability one, leaving the consumer with zero surplus. Firm 2 could offer the consumer the same surplus at price $p_2 = r_2$, with profit margin $r_2 - c_2 = s_2$. Thus, by advertising a price that slightly undercuts Firm 1 by ε, Firm 2 could win the sale with probability one and earn profit $s_2 - A - \varepsilon$. Because $A < s_2$ by assumption, this deviation would be profitable for sufficiently small ε; thus $a_2 = 0$ is impossible. On the other hand, if $a_2 = 1$, then π_1 is strictly positive by Lemma 3, and so Firm 1 must also advertise with probability one. But then by arguments similar to Lemma 2, Firm 2's profit margin at the highest price it ever offers must be weakly negative. But then, Firm 2 does not cover its ad cost, and so $\pi_2 < 0$, contradicting the optimality of advertising with probability one. ■

Lemma 6 *Firm 1 advertises with probability one. That is,* $a_1 = 1$.

Proof. Lemmas 3 and 5 imply that $\pi_1 > 0$. But this means that Firm 1 cannot be indifferent to not advertising (and thereby earning zero profit), so $a_1 = 1$. ■

Lemma 7 *Equilibrium profits are* $\pi_1 = \Delta_{12}$ *for Firm 1 and* $\pi_i = 0$ *for all* $i > 1$.

Proof. As noted just above, Lemmas 3 and 5 imply that $\pi_1 > 0$. The fact that $\pi_i = 0$ for all $i > 1$ follows from Lemma 2. To pin down π_1, let $\underline{p}_1$ and $\underline{p}_2$ be the lower bounds on the supports of the price distributions used by Firms 1 and 2 respectively. These lower bounds must give the consumer equal surplus – that is, $\underline{p}_1 = \underline{p}_2 + (r_1 - r_2)$ – as if they did not, the firm offering the consumer the better deal could raise its price slightly without affecting its chance of making the sale. Next, we claim that $\underline{p}_2 - c_2 = A$. Clearly we cannot have $\underline{p}_2 - c_2 < A$, as in this case Firm 2 could not recover its ad cost by offering $\underline{p}_2$. On the other hand, if $\underline{p}_2 - c_2 > A$, then either (i)

Firm 1 has no atom at price $\underline{p}_1$, in which case Firm 2 wins for sure by advertising $\underline{p}_2$, thereby making strictly positive profit $\underline{p}_2 - c_2 > A$, or (ii) Firm 1 has an atom at $\underline{p}_1$, in which case Firm 2 could win for sure and make a strictly positive profit by deviating slightly below $\underline{p}_2$. As both cases are incompatible with zero profit for Firm 2 in equilibrium, we have $\underline{p}_2 - c_2 = A$. But this implies that $\underline{p}_1 - c_1 = A + \Delta_{12}$. Furthermore, Firm 2 cannot have an atom at $\underline{p}_2$ either (or else Firm 1 could do strictly better by deviating below $\underline{p}_1$), so Firm 1 wins with probability one when it offers $\underline{p}_1$, earning profit $\underline{p}_1 - c_1 - A = \Delta_{12}$. Since any other price in the support of Firm 1's price distribution must do equally well, we have $\pi_1 = \Delta_{12}$. ■

2.3.2 Mixed Strategy Offer Distributions

Notice that when Firm i advertises a price p_i, this is equivalent to offering the consumer a surplus of $\sigma_i = r_i - p_i$, so firms' strategies may be characterized either in terms of the distributions of prices they demand or the distributions of surpluses they offer. It is convenient to roll the decision to advertise into these surplus distributions by regarding a decision not to advertise as an offer of zero surplus. That is, let $G_i(\sigma) = \Pr(\sigma_i \leq \sigma)$ be the probability that the consumer's offer from Firm i is no better than σ, with the event that Firm i does not advertise recorded as $\sigma_i = 0$. Given the probability a_i that Firm i advertises, its price distribution conditional on placing an ad may be recovered from the identity

$$G_i(\sigma) = 1 - a_i + a_i \Pr\left(p_i \geq r_i - \sigma \mid \text{Firm } i \text{ advertises}\right)$$

That is, an offer weakly worse than σ means that Firm i either did not advertise, or advertised a price weakly higher than $r_i - \sigma$.

Proposition 1 *In equilibrium, the top firm advertises with probability one and makes expected profit equal to Δ_{12}, its surplus advantage over the second-ranked firm. The* *second-ranked firm advertises with probability* $a_2 = \frac{s_2 - A}{s_1} \in (0,1)$ *and earns zero expected profit. No other firm advertises. The surplus distributions offered to the consumer by Firms 1 and 2 are* $G_1(\sigma) = \frac{A}{s_2 - \sigma}$ *and* $G_2(\sigma) = \frac{A + \Delta_{12}}{s_1 - \sigma}$ *respectively, with common support* $\sigma \in [0, s_2 - A]$.

Proof. Lemmas 1 through 7 establish that $a_1 = 1$, $a_2 \in (0,1)$, and $a_3, ..., a_n = 0$. Let $\bar{\sigma}_i$ and $\underline{\sigma}_i$ be the upper and lower supports on the surplus distribution offered by Firm i, $i \in \{1,2\}$. Since Firm 2 does not always advertise, we have $\underline{\sigma}_2 = 0$. By standard arguments, these supports are common (with $\bar{\sigma}_1 = \bar{\sigma}_2 = \bar{\sigma}$ and $\underline{\sigma}_1 = \underline{\sigma}_2 = 0$), have no gaps, and have no atoms on $(0, \bar{\sigma}]$. If $\bar{\sigma}_1 > \bar{\sigma}_2$, then Firm 1 could be strictly less generous than $\bar{\sigma}_1$ and still sell with probability one, and *vice versa*, so $\bar{\sigma}_1 = \bar{\sigma}_2$. If $0 = \underline{\sigma}_2 < \underline{\sigma}_1$, then (i) if Firm 2 makes any offers in the interval $(0, \underline{\sigma}_1)$, they never succeed and thus lose money, or (ii) if Firm 2 makes no offers in $(0, \underline{\sigma}_1)$, then Firm 1 could make a less generous offer than $\underline{\sigma}_1$, sell no less often, and make more money. So $\underline{\sigma}_1 = \underline{\sigma}_2 = 0$. The argument for gaps is completely standard. For atoms, first note that $\bar{\sigma} \leq s_2 - A$ (as Firm 2 would lose money by advertising more generous offers). Thus the gross profit margin (before ad costs) on any offer is at least $s_2 - \bar{\sigma} \geq A > 0$ for Firm 2, and greater for Firm 1. Then standard undercutting arguments apply – by shifting its offer from slightly below to slightly above a rival's atom, a firm would enjoy a jump in its sales at (essentially) the same, strictly positive gross profit margin. Finally, note that Firm 2 sells with probability one when advertising $\sigma_2 = \bar{\sigma}$, thus earning net profit $(s_2 - \bar{\sigma}) - A$. But $\pi_2 = 0$ by Lemma 7, so $\bar{\sigma} = s_2 - A$.)

Note we have not ruled out atoms at $\sigma = 0$. Firm 2 must have such an atom, because it does not always advertise, while Firm 1 will turn out to have such an atom because it will advertise $p_1 = r_1$ with positive probability. We must be a bit careful in handling these, as advertised offers of $\sigma = 0$ incur ad cost A, while unadvertised offers do not.

When Firm 2 offers surplus $\sigma_2 \in (0, s_2 - A]$, it sells with probability $G_1(\sigma_2)$ and earns profit $(s_2 - \sigma_2) G_1(\sigma_2) - A$. Then, as $\pi_2 = 0$ and Firm 2 must be indifferent over its support, we have $G_1(\sigma) = \frac{A}{s_2 - \sigma}$ for $\sigma \in (0, s_2 - A]$. Similarly, when Firm 1 offers $\sigma_1 \in (0, s_2 - A]$, it sells with probability $G_2(\sigma_1)$ and earns profit $(s_1 - \sigma_1) G_2(\sigma_1) - A = \pi_1 = \Delta_{12}$; thus we have $G_2(\sigma) = \frac{A + \Delta_{12}}{s_1 - \sigma}$ for $\sigma \in (0, s_2 - A]$. Notice that Firm 1 advertises $\sigma_1 = 0$ with positive probability $G_1(0) = \frac{A}{s_2}$. Given this, Firm 2 cannot find it optimal to *advertise* $\sigma_2 = 0$ itself – doing so would tie Firm 1's atom, while undercutting with a slightly better offer would win twice as often. Thus any probability mass on $\sigma_2 = 0$ reflects Firm 2's failure to advertise. Since $G_2(0) = \frac{A + \Delta_{12}}{s_1}$, we have $1 - a_2 = \frac{A + \Delta_{12}}{s_1}$ and so $a_2 = \frac{s_2 - A}{s_1}$. ■

The short-cut intuition for some of the key values in the Proposition is as follows. First, because Firm 2 earns zero profit in equilibrium then its lowest price (at which it wins for sure) is A above its unit production cost. This is analogous to the lowest price in Butters' (1977) model: anything lower would not cover the cost of sending the ad. Thus the highest consumer surplus value of $s_2 - A$ is attained when buying at that price. When Firm 1 matches this surplus level, its corresponding price is $r_1 - (s_2 - A)$ and it wins for sure. Subtracting its unit production cost, then 1's gross revenue is $\Delta_{12} + A$. Subtracting from this amount the cost A of sending the ad gives 1's equilibrium profit level as the value of its advantage, Δ_{12}. Firm 1 gets the same profit when it delivers zero surplus to the consumer, pricing at r_1 and earning a gross profit of s_1 when it wins. Firm 1 only wins at this highest price when its rival does not advertise, which happens with probability $(1 - a_2)$, and costs A. This profit indifference property $s_1(1 - a_2) - A = \Delta_{12}$ ties down the *rival's* advertising probability as $a_2 = 1 - \frac{\Delta_{12} + A}{s_1} = \frac{s_2 - A}{s_1}$. Notice here the inherent asymmetry between ad levels, which remain distinctly different even as social surpluses get arbitrarily close. Even for small social surplus differences the dominant firm always advertises while the weaker one rarely contests it if A is a significant fraction of s_2. We return to this asymmetry below.

Now consider the probability $G_1(0)$ that Firm 1 charges its top price, r_1, delivering zero consumer surplus. In the mixed strategy equilibrium, this probability must make Firm 2 indifferent between advertising and not. If Firm 2 sets its price just below r_2, it wins the consumer with probability $G_1(0)$ for a gross profit of s_2 at a cost of A. Thus $G_1(0)\, s_2 = A$, and so the probability that Firm 1 sets the top price is thus A/s_2. Notice that this probability goes to 1 as A rises to s_2, so that Firm 1 sets its monopoly price more frequently as the cost of advertising rises. Indeed, for $A \geq s_2$ (but $A < s_1$) Firm 1 is an uncontested monopolist and always prices at r_1.

2.3.3 Bertrand Limit as $A \to 0$

In the usual version of asymmetric Bertrand competition when the consumer is notified about firms' price offers automatically and costlessly, the standard pure strategy equilibrium has the second-ranked firm pricing at cost, $p_2 = c_2$, while the top firm offers $p_1 = c_2 + (r_1 - r_2)$, the highest price at which its product is weakly preferred over Firm 2's, sells with probability one, and earns profit Δ_{12}. However, somewhat *ad hoc* arguments must be made to dispense with technical complications before reaching this intuitive conclusion.[6] Our model delivers this outcome naturally, as the unique limit of equilibria as the advertising cost vanishes. Corollary 1 follows directly from Proposition 1 by taking $A \to 0$.

Corollary 1 *In the limit as $A \to 0$, the top firm advertises with probability one and makes expected profit equal to the social surplus difference Δ_{12}. The second-ranked*

[6]Tirole (1988, p.234) notes two problems: the open-set problem of ϵ-undercutting, and the possibility of an equilibrium price between the two cheaper firms' costs. The former problem is typically solved by invoking an efficient allocation rule to allocate customers to the socially preferable firm when faced with price ties (see, e.g., Lederer and Hurter, 1986). The latter problem can be resolved by eliminating weakly dominant strategies (e.g. Tirole, 1988, p.234, fn 37.), although such recourse would also eliminate the second-best firm pricing at its cost. An alternative solution is to consider a fine grid of prices and again eliminate weakly dominated strategies (see Mas-Collel, Whinston, and Green, 1995, p.430).

firm advertises with probability s_2/s_1 and earns zero expected profit. No other firm advertises.

Furthermore, it is straightforward to verify that consumer surplus tends to s_2 – the full surplus from the second-best option – and total social surplus tends toward its first-best level s_1 (implying the consumer buys from Firm 1 with probability tending to one). These points match the standard Bertrand result as well. While the firms never stop mixing over the full support of prices, as $A \to 0$ the weight Firm 1 places on its most competitive offer goes to one. In contrast, as $A \to 0$, Firm 2 advertises just often enough at every surplus level $\sigma \in [0, s_2]$ so that Firm 1 is not tempted to make any offer less generous than its most competitive one.

2.3.4 Mixed Strategy Prices

We can now determine the price distributions for the top two firms, $F_1(p)$ and $F_2(p)$ respectively, conditional on their advertising. These price distributions follow directly from the identity linking prices, advertising, and surplus using $p = r_1 - \sigma$. For Firm 1, we have $G_1(r_1 - p) = \Pr(p_1 \geq p) = 1 - F_1(p) + \Pr(p_1 = p)$. This yields:

$$F_1(p) = \begin{cases} 1 - \frac{A}{(p-c_1)-\Delta_{12}} & \text{if } p \in [c_1 + \Delta_{12} + A, r_1) \\ 1 & \text{if } p \geq r_1 \end{cases}$$

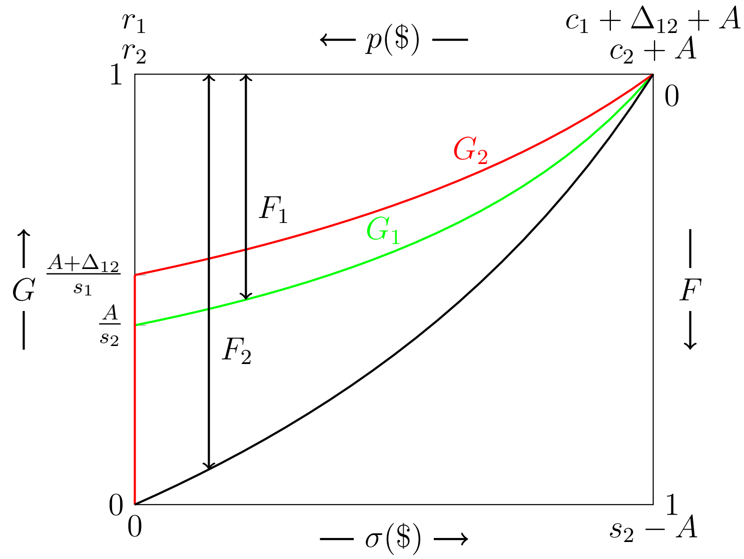
where the atom at zero surplus translates into an atom at the consumer's reservation price because Firm 1 is advertising with probability one. Because Firm 2 does not advertise with positive probability, we have $G_2(r_2 - p) = 1 - a_2 + a_2 \Pr(p_2 \geq p)$, or (using $G_2(0) = 1 - a_2$), $\Pr(p_2 < p) = \frac{1-G_2(r_2-p)}{1-G_2(0)}$. As this distribution is atomless, we may substitute a weak inequality and plug in to get:

$$F_2(p) = \frac{s_1}{s_2 - A}\left(1 - \frac{\Delta_{12} + A}{\Delta_{12} + (p - c_2)}\right) \quad \text{if } p \in [c_2 + A, r_2)$$

As is often the case with price competition in mixed strategies, both firms' price distributions are in the generalized Pareto family with tail exponent 1. Empirical evidence suggests that pricing strategies generally follow a Pareto distribution. A number of well-known papers derive Pareto distributions from their mixed strategy analysis including Butters (1977), Varian (1980), Baye and Morgan (2001), and Stahl (1989).

The construction of the price distributions, $F_1(p)$ and $F_2(p)$ from the surplus distributions $G_1(\sigma)$ and $G_2(\sigma)$ is shown in Figure 2.1 below. Henceforth we will return to using the surplus distributions in our analysis because of the convenient structure.

Figure 2.1: Equilibrium Price and Offer Distributions



The lower horizontal axis measures σ for distributions G_1 and G_2. $G_1(0)$ is the mass point for a zero surplus offer by Firm 1; this equals mass point for Firm 1 setting price r_1, $1 - F_1(r_1)$. The upper horizontal axis measures price from the right for price distributions, F_1 and F_2.

2.4 Consumer Surplus, Social Surplus, Advertising Costs

2.4.1 Consumer Surplus

A number of facts about equilibrium consumer welfare emerge rather directly from inspection of the surplus distributions G_1 and G_2. To begin with, we can determine which of the two active firms tends to give the consumer better offers.

Proposition 2 *The consumer's surplus offers from the top firm first order stochastically dominate her surplus offers from the second-ranked firm. (That is, $G_1(\sigma) \succsim_{FOSD} G_2(\sigma)$).*

Proof. Noting that $G_2(\sigma) = \frac{A+\Delta_{12}}{(s_2-\sigma)+\Delta_{12}}$ makes it clear that $G_2(\sigma)$ can be written as a convex combination of $G_1(\sigma) = \frac{A}{s_2-\sigma}$ and $\frac{\Delta_{12}}{\Delta_{12}} = 1$; thus $G_2(\sigma) \geq G_1(\sigma)$ (with $G_2(\sigma) < G_1(\sigma)$ on the interior of the support: $\sigma < s_2 - A$). ■

This contrasts with the familiar results for asymmetric Bertrand competition when firms' price offers are announced automatically and costlessly. In that case, Firm 2 prices at its cost, and Firm 1 prices at its cost plus Δ_{12}, the markup that makes the consumer indifferent between offers, and the consumer receives surplus s_2 from either firm. Intuitively, because of its higher profit margin, Firm 1 has a greater incentive than does Firm 2 to sweeten its surplus offer to be sure it wins. This logic applies with or without costly advertising; however without advertising, the amount by which Firm 1 needs to sweeten its offer relative to Firm 2 shrinks to zero since it can undercut Firm 2's pure strategy arbitrarily closely.

Realized consumer surplus is just $\sigma_{\max} = \max(\sigma_1, \sigma_2)$, since the consumer picks the best offer she gets. The cumulative distribution function for consumer surplus is then

$$\begin{aligned} G_{\max}(\sigma) &= G_1(\sigma) G_2(\sigma) \\ &= \left(\frac{A}{s_2-\sigma}\right)\left(\frac{A+\Delta_{12}}{s_1-\sigma}\right) \\ &= \left(\frac{A}{s_2-\sigma}\right)\left(\frac{A+s_1-s_2}{s_1-\sigma}\right) \end{aligned} \tag{2.1}$$

Using $G_{\max}(\sigma)$, we determine the impact on consumer surplus from changes in the competitive environment. Several of the highlights are summarized below.

Proposition 3 *The distribution of realized consumer surplus is increasing (in the sense of first order stochastic dominance) in s_2. It is decreasing in s_1 and in the ad cost A.*

Proof. These properties follow directly from (2.1). ■

Corollary 2 *Expected consumer surplus increases in s_2 and decreases in s_1 and A.*

It might be tempting to argue that consumer surplus must rise as A declines because a lower barrier to reaching the consumer must surely make the market more competitive. This is not necessarily wrong in the end, but it misses some subtlety. Because the firms' profits do not vary with A, consumer surplus moves in lockstep with total social surplus as A declines. There are two effects on social surplus to consider. First, allocative efficiency – namely, the chance that Firm 1 (the firm with the highest social surplus) gets the sale – generally appears to be U-shaped in A (as Firm 1 wins with probability tending to one for $A \approx s_2$ and $A \approx 0$, but loses with positive probability in between).[7] Thus when A is large, a reduction in ad costs increases the chance of a sale by the wrong firm, tending to reduce consumer surplus. However, the second effect is that the total cost of advertising, which ends up being borne by the consumer, unambiguously declines with a decline in A. In the end, the second effect dominates.[8]

The fact that a better second-ranked option helps the consumer to carve out more surplus is natural and would hold in the textbook Bertrand setting as well. It is less obvious that an improvement in her best option s_1 should hurt the consumer – after all it would have no effect at all in the textbook Bertrand setting. Here, as one can see from the $G_2(\sigma)$ term within $G_{\max}(\sigma)$ in (2.1), a stronger best choice s_1 induces the second-ranked firm to back off and compete less vigorously, thereby hurting the consumer.

Expected consumer surplus may be computed directly from $G_{\max}(\sigma)$ in (2.1):[9]

$$CS = E_{G_{\max}}(\sigma) = s_2 - A\left(1 + \frac{A+\Delta_{12}}{\Delta_{12}} \ln\left(\frac{s_2}{s_1}\frac{A+\Delta_{12}}{A}\right)\right)$$

[7] These points follow from $a_2 \to 0$ as $A \to s_2$ and social surplus tending to its first-best level s_1 (implying an efficient allocation with probability one) as $A \to 0$.

[8] This is not completely trivial, since total advertising volume $a_1 + a_2 = 1 + \frac{s_2 - A}{s_1}$ is decreasing in A. However total ad cost $A(a_1 + a_2)$ is increasing in A over $A \in [0, s_2]$.

[9] An interesting alternative form emerges by recasting the logarithmic expression in terms of profit margins. Let $\bar{\mu}_1 = \bar{p}_1 - c_1$ and $\underline{\mu}_1 = \underline{p}_1 - c_1$ be Firm 1's largest and smallest gross profit margins in

Defining

$$L(A,s_1,s_2) = A\left(1+\frac{A+\Delta_{12}}{\Delta_{12}}\ln\left(\frac{s_2}{s_1}\frac{A+\Delta_{12}}{A}\right)\right)$$

we have $CS = s_2 - L(A,s_1,s_2)$ – that is, the consumer earns her asymmetric Bertrand payoff s_2, minus a loss term that is increasing in the ad cost A. (We know that $L(A,s_1,s_2) \geq A$ because $\sigma_{\max}$ has upper support $s_2 - A$.) Furthermore, one can show that $\lim_{A\to 0} L(A,s_1,s_2) = 0$, so as advertising costs vanish, the consumer tends toward her asymmetric Bertrand payoff.

2.4.2 Advertising and Social Surplus

Denote expected social surplus as $SS = CS + \pi_1 + \pi_2$. At equilibrium profits,

$$SS = s_1 - L(A,s_1,s_2)$$

First-best social surplus in the absence of advertising costs would just be $SS_{eff} = s_1$, the surplus from allocating the consumer to Firm 1. Thus, $L(A,s_1,s_2)$ also may be interpreted as the shortfall of equilibrium social surplus below its first-best level. If the consumer is unaware of an unadvertised product, the reasonable benchmark is the second-best (constrained-efficient) social surplus that takes the necessity of advertising into account. This is $SS_{2bo} = s_1 - A$, where now the cost of apprising the consumer of her first-ranked option is included. Then we may write

$$\begin{aligned} SS &= SS_{2bo} - \big(L(A,s_1,s_2) - A\big) \\ &= SS_{2bo} - A\left(\frac{A+\Delta_{12}}{\Delta_{12}}\right)\ln\left(\frac{s_2}{s_1}\frac{A+\Delta_{12}}{A}\right) \end{aligned}$$

equilibrium (with $\bar{p}_1 = r_1$ and $\underline{p}_1 = c_1 + \Delta_{12} + A$), and define $\bar{\mu}_2$ and $\underline{\mu}_2$ similarly for Firm 2. Then,

$$CS = \bar{\mu}_2 - \underline{\mu}_2\left(1+\frac{\bar{\mu}_1}{\bar{\mu}_1 - \underline{\mu}_2}\ln\left(\frac{\bar{\mu}_2/\underline{\mu}_2}{\bar{\mu}_1/\underline{\mu}_1}\right)\right).$$

This expression delivers a simple statistic with which to compute consumer surplus under personalized price competition using only the highest and lowest profit margins for Firms 1 and 2.

Shortfalls below these two benchmarks arise from two sources: excessive ad costs and the wrong firm (Firm 2) winning the sale. These can be easily decomposed. Given advertising $a_1 = 1$ and $a_2 = \frac{s_2 - A}{s_1}$, the total social cost of advertising is $A\left(1 + \frac{s_2 - A}{s_1}\right)$. Of this, Firm 1's share A is necessary, in the constrained-efficient sense, while Firm 2's share $A\frac{s_2 - A}{s_1}$ is wasteful. Thus, the "Avoidable inefficiency," or $SS_{2bo} - SS = L(A, s_1, s_2) - A$, may then be broken down as follows. The social cost of misallocation is $L(A, s_1, s_2) - A\left(1 + \frac{s_2 - A}{s_1}\right)$, or

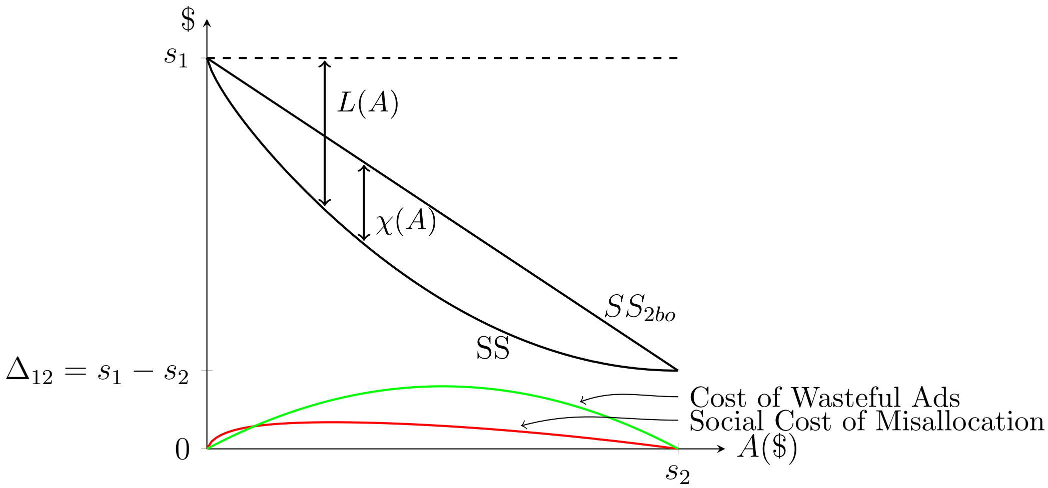
$$\begin{aligned}
\text{Cost of wasteful advertising } &= Aa_2 = A\frac{s_2 - A}{s_1} \\
\text{Social cost of misallocation } &= \Delta_{12} \Pr(\text{Firm 2 wins}) \\
&= \frac{A(A + \Delta_{12})}{\Delta_{12}} \ln\left(\frac{s_2}{s_1}\frac{A + \Delta_{12}}{A}\right) - A\frac{s_2 - A}{s_1}
\end{aligned}$$

We have already established that $L(A, s_1, s_2)$ is increasing in A, so the gap between equilibrium and first-best social surplus shrinks as ad costs decline. However, this decline is driven in large part by a mechanical effect: the declining cost of Firm 1's certain advertising, $Aa_1 = A$. If we view this cost as unavoidable, as the second-best benchmark does, then the relationship of equilibrium efficiency to ad costs is more nuanced.

Proposition 4 *Avoidable inefficiency $SS_{2bo} - SS$ vanishes at $A = 0$ and $A = s_2$. Furthermore, it is positive, strictly concave, and single-peaked in A over $A \in (0, s_2)$.*

Proof. Let $\chi(A) = SS_{2bo} - SS = \frac{A(A+\Delta_{12})}{\Delta_{12}} \ln\left(\frac{s_2}{s_1}\frac{A+\Delta_{12}}{A}\right)$. It is immediate that $\chi(s_2) = 0$, and $\lim_{A\to 0} \chi(A) = 0$ follows by taking the limit. Differentiation yields $\chi'(A) = \frac{1}{\Delta_{12}}(2A + \Delta_{12}) \ln\left(\frac{s_2}{s_1}\frac{A+\Delta_{12}}{A}\right) - 1$, so $\chi'(s_2) = -1$ and $\lim_{A\to 0} \chi'(A) = \infty$. Thus $\chi(A)$ is strictly positive near the endpoints of $(0, s_2)$. Differentiating again yields $\chi''(A) = \frac{2}{\Delta_{12}} \ln\left(\frac{s_2}{s_1}\right) + \frac{1}{\Delta_{12}}\left(2\ln\frac{A+\Delta_{12}}{A} + \frac{A}{A+\Delta_{12}} - \frac{A+\Delta_{12}}{A}\right)$. The first term is strictly negative because $s_2 < s_1$, so for concavity it will suffice to show the second term is negative as well. Write the second term as $\frac{1}{\Delta_{12}}\xi\left(\frac{A+\Delta_{12}}{A}\right)$ for $\xi(z) = 2\ln z + \frac{1}{z} - z$. We claim that $\xi(z) < 0$ for all $z > 1$ (and so $\xi\left(\frac{A+\Delta_{12}}{A}\right)$ because $\frac{A+\Delta_{12}}{A} > 1$). To show this claim, observe that $\xi(1) = 0$ and $\xi'(z) = -\left(1 - \frac{1}{z}\right)^2$. ■

Figure 2.2: Social Surplus and Avoidable Inefficiency



$SS_{2bo} = s_1 - A$ is the second-best social surplus. SS is the equilibrium social surplus and $\chi(A)$ is the avoidable inefficiency, which vanishes at $A = 0$ and $A = s_2$; $\chi(A)$ can be decomposed into the cost of wasteful ads and the social cost of misallocation. $L(A) = \chi(A) + A$ is the difference between equilibrium social surplus and the first-best social surplus.

The fact that the equilibrium is second-best optimal at $A = s_2$ is straightforward, because for $A \geq s_2$ the second-ranked firm cannot afford to enter the market and so the first-ranked firm has a monopoly. Social surplus increases as advertising costs fall below s_2, permitting the second-ranked firm to enter, though the effect is negligible at first. Social losses due to socially excessive advertising and sales by the wrong firm rise as ad costs decline as shown on Figure 2.2. In this sense, lower ad costs initially open the door to the second-ranked firm, giving it a chance to win sales (which it should not do, from the standpoint of efficiency), thereby creating an incentive for it to advertise (which it also should not do). Total advertising volume continues to rise as A falls, but eventually the cost of excessive advertising begins to decline as its ad cost tends to zero. Furthermore, as A falls, the chance of Firm 2 winning rises to a peak before declining to zero.

2.5 Equilibrium when Firms are Symmetric

This section considers a symmetric version of our model in which all firms are identical.[10] There is a symmetric equilibrium in which all of the firms advertise with positive probability less than one. However, there is also a natural asymmetric equilibrium obtained by taking limits in Proposition 1 as differences in surplus vanish; we call this the limiting asymmetric equilibrium.[11] We characterize both equilibria below and then discuss reasons why the limiting asymmetric equilibrium may be more appealing as a prediction of behavior.

2.5.1 Symmetric Equilibrium

Suppose that each of the n firms has potential surplus s_1. In a symmetric equilibrium each firm must earn zero profit. (Strictly positive profits would imply that all firms always advertise, but this is impossible because under the Bertrand competition that would ensue none of the firms would even cover its ad cost.) The best surplus offer made must be $\bar{\sigma} = s_1 - A$ which wins with probability one and earns zero profit. (If $\bar{\sigma}$ were lower, then any firm could earn strictly positive profits by overcutting it.) The worst offer (which wins only if no other firm advertises) must still be the monopoly offer $\underline{\sigma} = 0$; standard arguments rule out atoms or gaps over the support $(0, s_1 - A]$. If $G_i(\sigma)$ is the common mixed strategy distribution over surplus offers, then Firm 1's probability of winning the sale with offer σ is $G_{-1}(\sigma) = G_i(\sigma)^{n-1}$

[10]Partially symmetric cases (where only some firms are identical) are treated in the appendix.

[11]More properly, there is a collection of such equilibria, differing only in the identities of the firms playing the roles of Firm 1 and Firm 2. There are also additional equilibria in which an arbitrary subset of $\tilde{n} < n$ of the firms play a version of the symmetric equilibrium (with $\tilde{n} \geq 2$ replacing n), while the $n - \tilde{n}$ others sit out (never advertise). For details, see the appendix.

(where $G_{-1}(\sigma)$ is the distribution of the best rival offer faced by Firm 1). Since the gross profit on an offer σ is $s_1 - \sigma$, Firm 1's zero-profit indifference condition then becomes $(s_1 - \sigma) G_{-1}(\sigma) - A = 0$ (and similarly for each other firm). This pins down a firm's symmetric equilibrium surplus offer distribution as $G_i(\sigma) = \left(\frac{A}{s_1-\sigma}\right)^{\frac{1}{n-1}}$, for $i \in \{1, ..., n\}$. Under these symmetric strategies, the probability mass $G_i(0)$ must reflect a failure to advertise, not an atom on zero-surplus offers (as such an atom would be profitably undercut). Hence we have the following:

Proposition 5 *In the symmetric equilibrium with n firms each delivering potential surplus $s_1 > A$, expected profit for each firm is zero, and the equilibrium offer distribution is $G_i(\sigma) = \left(\frac{A}{s_1-\sigma}\right)^{\frac{1}{n-1}}$ with support on $[0, s_1 - A]$, $i \in \{1, ..., n\}$. Each firm refrains from advertising with probability $G_i(0) = \left(\frac{A}{s_1}\right)^{\frac{1}{n-1}}$ and the consumer's best offer has distribution $G_{\max}(\sigma) = \left(\frac{A}{s_1-\sigma}\right)^{\frac{n}{n-1}}$.*

2.5.2 Limiting Asymmetric Equilibrium

The limiting asymmetric equilibrium of the symmetric model follows trivially from Proposition 1 by taking limits as $s_i \to s_1$ for all i. We obtain that Firm 1 always advertises, Firm 2 advertises with probability $a_2 = \frac{s_1 - A}{s_1}$, both earn zero profits, and all other firms sit out. Firm 1's surplus offers follow $G_1(\sigma) = \frac{A}{s_1-\sigma}$, with $G_1(0) = \frac{A}{s_1}$ representing an atom at the monopoly offer $\sigma = 0$. Firm 2's offers follow $G_2(\sigma) = \frac{A}{s_1-\sigma}$ – identical to Firm 1, except that $G_2(0) = \frac{A}{s_1}$ represents not advertising. It is straightforward to confirm that these strategies do constitute an equilibrium of the symmetric game.

2.5.3 Arguments against the Symmetric Equilibrium

While it is commonplace to focus on symmetric equilibria when a game is symmetric, there is not always a compelling rationale for doing so. In our case, we will argue that the symmetric equilibrium is less attractive than the limiting asymmetric one because it is unstable and generates perverse comparative statics. Other papers have noted similarly counterintuitive comparative statistics in symmetric equilibria, but we also stress the underlying instability of the symmetric equilibria. The same features hold in related games such as Varian's Model of Sales.

(In)stability of the Symmetric Equilibrium

Symmetry can be a useful simplifying assumption, but in reality we would generally expect that firms differ at least a little bit in their costs, qualities, or both. Suppose that we start from a symmetric situation and then perturb Firm i's surplus to $s_i = s + \zeta_i$, where ζ_i is a publicly observed idiosyncratic shock. Then generically, the surpluses will all be distinct, and so the unique equilibrium is the one characterized by Proposition 1. Thus, for a small perturbation, outcomes in the perturbed model will not be close to the symmetric equilibrium above, which has no counterpart when surpluses are unequal, but will be close to outcomes in the limiting asymmetric equilibrium.[12] In this sense, the symmetric equilibrium is unstable.[13]

Unappealing Welfare Implications of the Symmetric Equilibrium

Because firms earn zero profits in the symmetric equilibrium, expected social surplus and consumer surplus are equal. The following unintuitive and implausible property about the symmetric equilibrium follows from the fact that the consumer's best offer distribution $G_{\max}(\sigma)$ is increasing in n:

[12] To be more precise, there are $n(n-1)$ limiting asymmetric equilibria, depending on which firms take the roles of Firm 1 and 2, and the equilibrium of the perturbed game must be close to one of these. Furthermore, equilibrium outcomes, such as profits, total advertising, and the distribution of surplus offers, will be close to the values they take in all of the limiting asymmetric equilibria.

[13] This argument does depends on the assumption that at least some of the differences among firms are public rather than private; we would argue that this is reasonable. Correlated deviations from symmetry do not upend this argument unless they are such that the top two (or more) firms remain identical; it is hard to see why this should be the case.

Proposition 6 *In the symmetric equilibrium, expected social surplus and consumer surplus are decreasing in the number of firms n.*

Thus, more competition among firms reduces welfare and makes consumers worse off. This surprising property is driven by the indifference conditions underlying the equilibrium. As n rises, the distribution of the best offer from any collection of $n-1$ of the firms must remain constant (so as to keep the remaining firm indifferent to competing). This is only possible if each individual firm competes *less* vigorously as n rises. But then a consumer's overall best offer, the max of the best offer from the first $n-1$ firms and the final firm's offer, must grow statistically worse with n.[14]

In contrast, in the limiting asymmetric equilibrium social surplus and consumer surplus are unaffected by additional firms, since competition by the two active firms is already sufficiently fierce to foreclose the market to everyone else. The result that additional competition *does* improve consumer surplus – but that all of the gains are made in the shift from monopoly to duopoly – is admittedly extreme, but it is typical of Bertrand price-setting games. Just as in other Bertrand-like games, we would expect this conclusion to soften into a more gradual consumer surplus improvement with n under softer price competition (such as discussed in Section 2.6.3, for example).[15]

Thus we argue that the symmetric equilibrium is a less appealing prediction of behavior than the limiting asymmetric one on the grounds of both stability and

[14]This logic has some precedent in mixed strategy equilibria. The result is reminiscent of the Palfrey and Rosenthal (1984) binary public good game whereby acting provides a public value v to all players at cost c to the players that choose to act. Sharkey and Sibley (1993) (for the symmetric case here discussed) already noted the anti-competitive effect of entry on the equilibrium price distribution per firm, and Stahl (1994) shows that social surplus can decrease. Indeed, Stahl (1994) analyzes an advertising cost function that encompasses both the Butters (1977) case and ours (as a limit case) and finds that with a relatively flat marginal cost of advertising, seller entry can decrease social surplus.

[15]In the appendix, we compare welfare in the symmetric and limiting asymmetric equilibria more closely, with a focus on decomposing sources of inefficiency.

unintuitive comparative statistics. The two are likely linked. We recall Samuelson (1941): "[T]he problem of stability of equilibrium is intimately tied up with the problem of deriving fruitful theorems in comparative statics."

2.6 Extensions

2.6.1 Downward-sloping demand

While we have assumed the firms face a consumer with unit demand, our key results continue to apply if her demand is downward-sloping. In particular, the top firm always advertises and earns a positive profit, the second-best firm advertises with positive probability and earns zero profit, and all other firms sit out. Below we sketch these results and point out some new wrinkles that emerge.

For simplicity, we assume the consumer has the same demand $D(p)$ for any firm's product, while the firms have heterogeneous, constant marginal costs $c_1 < c_2 < ... < c_n$.[16,17] The demand function $D(p)$ is assumed to be twice continuously differentiable and (-1)-concave.[18] Let p_i^m be Firm i's monopoly price against this demand curve. To rule out uninteresting cases, we assume that $(p_1^m - c_2) D(p_1^m) > A$ which ensures that at Firm 1's monopoly price at least the top two firms could cover their ad costs with a sale. Let $\tilde{\pi}_i(p) = (p - c_i) D(p)$ be the "pseudo-profit" function representing the gross profit Firm i could earn if it faced the consumer as a monopolist. The concavity assumption ensures $\tilde{\pi}_i(p)$ is strictly quasi-concave with $p_i^m \geq p_1^m$, so that Firm i's pseudo-profit is unambiguously increasing in price over the interval $p \in (c_i, p_1^m)$.

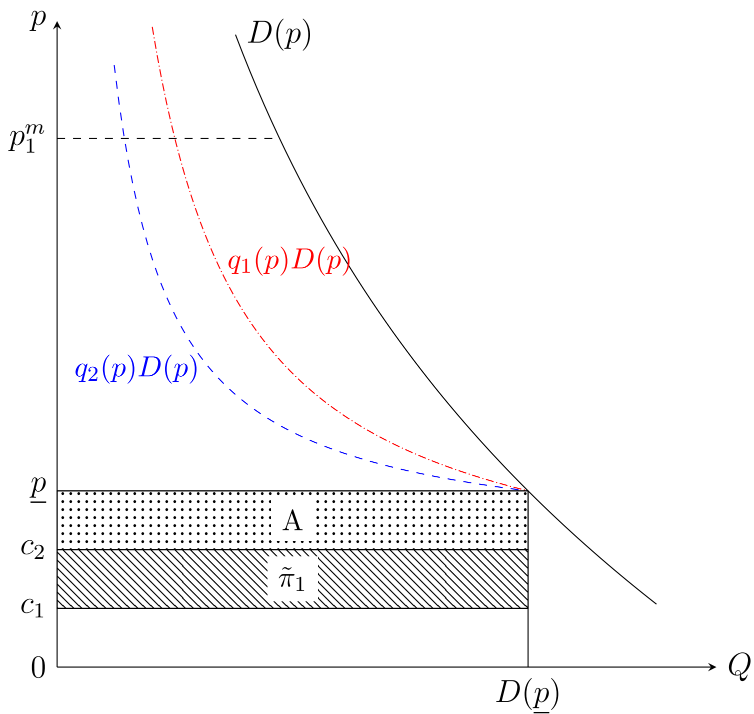
---

[16] The same analysis would apply, with some relabeling, if the consumer's demand for different products differed only by a firm-specific quality shift term.

[17] Of course, it would be interesting to study more general asymmetries in the demand faced by firms. However, when those asymmetries cannot be summarized by a one-dimensional parameter (implying that firm competitiveness cannot be unambiguously ranked), the analysis becomes quite complex. See the next section, with heterogeneous ad costs, for a sense of the issues that arise.

[18] This ensures that $D(p)$ is "more concave" than a rectangular hyperbola, and that marginal revenue slopes down. See Caplin and Nalebuff (1992) for more details on ρ-concavity.

Figure 2.3: Downward-sloping Demand



Determination of equilibrium sales probabilities $q_i(p)$ for downward-sloping demand.

Absent quality differences across firms, the consumer will simply choose her lowest price offer, so we present the analysis in terms of price offers rather than surplus offers. The main modification to our earlier set-up is that Firm i's expected profit when advertising p_i is now

$$\pi_i(p_i) = \tilde{\pi}_i(p_i) \Pr\big(i \text{ sells} \mid p_i\big) - A = (p_i - c_i) D(p_i) \Pr\big(i \text{ sells} \mid p_i\big) - A\ ;$$

where the only difference is the inclusion of the term $D(p_i)$. Lemma 1 goes through essentially unchanged; namely, if Firm i can break even at its lowest price offer, then a more efficient firm $j < i$ can survive at a slightly lower price. Lemmas 2 through 6 continue to hold as well, as does the argument that Firms 1 and 2 must mix over common support $[\underline{p}, \bar{p})$ (plus an atom at $\bar{p}$ for Firm 1), establishing the claims we made above.

As usual, the lower bound of the support is pinned down by the requirement that Firm 2 wins for sure and earns zero profit when offering $\underline{p}$; thus we have $\tilde{\pi}_2\left(\underline{p}\right) = A$ (which uniquely ties down $\underline{p}$). Since $\underline{p}$ is also in Firm 1's support, this establishes its equilibrium profit as $\pi_1^* = \tilde{\pi}_1\left(\underline{p}\right) - A = (c_1 - c_2)\, D\left(\underline{p}\right)$ (which reduces to our Lemma 7 result for unit demand). However this is smaller than the rents of $(c_1 - c_2)\, D\,(c_2)$ that Firm 1 would receive in the standard Bertrand model without advertising.[19] Furthermore, the upper bound $\bar{p}$ must be Firm 1's monopoly price p_1^m, since this offer wins for Firm 1 only when Firm 2 does not advertise; in contrast with the unit-demand case, this least competitive offer leaves the consumer a strictly positive surplus.

As with our earlier analysis, the firms' profits, the bounds on prices, and the two indifference conditions completely pin down the equilibrium mixed strategies and Firm 2's chance of advertising. These conditions may be succinctly illustrated on a diagram of each firm's residual demand $D_i^r\,(p) = q_i\,(p)\, D\,(p)$, where $q_i\,(p)$ is its equilibrium probability of a sale when offering p. As shown in Figure 2.3, each firm enjoys the full demand $D\left(\underline{p}\right)$ when offering $\underline{p}$, with gross profit of $\tilde{\pi}_2\left(\underline{p}\right) = A$ for Firm 2, and $\tilde{\pi}_1\left(\underline{p}\right) = A + \pi_1^*$ for Firm 1, as depicted. At higher prices, competition from Firm 2 must pull back Firm 1's residual demand just enough so that its gross profits lie on the iso-profit hyperbola $(p - c_1)\, D_1^r\,(p) = A + \pi_1^*$. Similarly, Firm 2's residual demand is determined by the iso-profit hyperbola $(p - c_2)\, D_2^r\,(p) = A$.[20] Then because Firm 2's chance of winning depends on Firm 1's price distribution $F_1\,(p)$ through the

[19] Due to the need to cover A, Firm 2 cannot compete down to marginal cost (as would be socially efficient); consequently quantity consumed is inefficiently low. Firm 1's per-unit profit margin is fixed at $c_1 - c_2$.

[20] As Figure 2.3 shows, $q_1\,(p) = D_1^r\,(p)\,/D\,(p) > q_2\,(p) = D_2^r\,(p)\,/D\,(p)$, so that Firm 1 has a greater chance of winning at any price than Firm 2. This must be the case in equilibrium because $q_1\,(p)$ is a convex combination of $q_2\,(p)$ and $D\left(\underline{p}\right)/D\,(p) > 1$. The equilibrium D^r's can be derived graphically using an analogous device to that used in the Appendix, where we find the equilibrium for the symmetric case.

relation $q_2(p) = 1 - F_1(p)$, $F_1(p)$ can be inferred from the ratio of Firm 2's residual demand to total demand. Likewise, Firm 1's chance of winning depends on Firm 2's price distribution, conditional on advertising, according to $q_1(p) = (1 - a_2) + a_2\left(1 - F_2(p)\right)$; thus the ratio $D_1^r(p)/D(p)$ pins down Firm 2's equilibrium strategy.[21]

All of this analysis presumes that each firm must offer a simple linear price. However, since our focus is on personalized offers made to a consumer based on detailed information about his tastes, it is natural to consider the possibility that firms can craft nonlinear price offers. If, for example, firms can offer two-part pricing, with a fixed fee and a per-unit price, it is not hard to see that the analysis is even simpler than with linear prices – in fact, it collapses back to the single-unit analysis. To see why, note that for standard reasons, each firm will always wish to set its per-unit price efficiently, at marginal cost, and take any profits through the fixed fee. But then we may simply apply our standard model, substituting in for the surplus s_i the total consumer surplus generated by purchasing the socially optimal quantity from Firm i, and letting the fixed fee play the role of p_i.

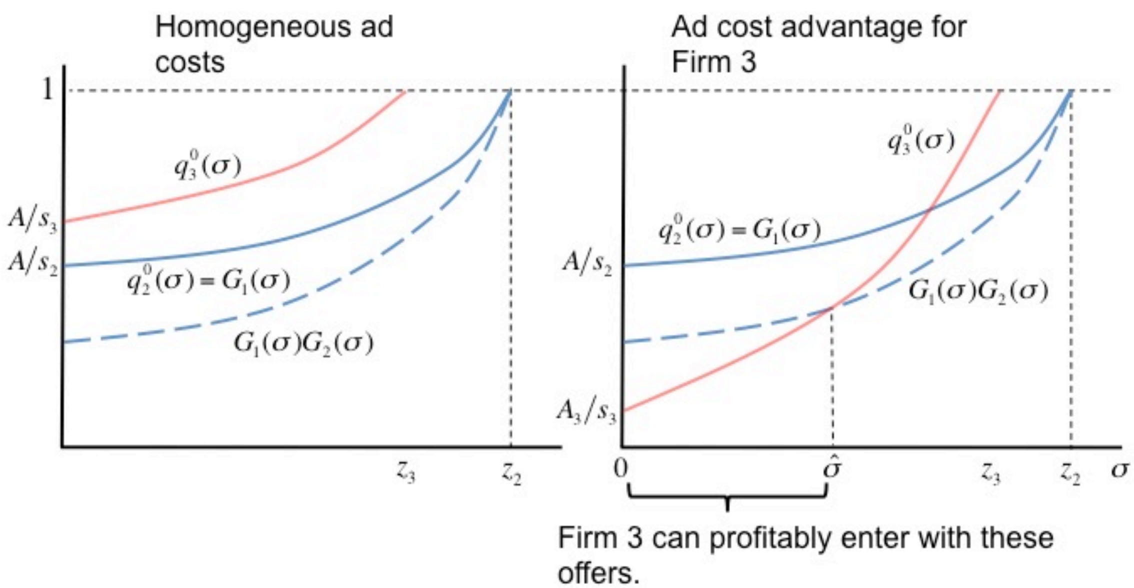
2.6.2 Competition when Firms have Different Costs of Targeted Advertising

Thus far we have assumed that all firms face the same cost A to reach the consumer with a targeted ad. However, one can imagine circumstances where this ad cost varies across firms, perhaps because they deliver offers through different channels

[21] In particular, for Firm 1 we have $F_1(p) = 1 - A/\left((p - c_2)D(p)\right) = 1 - \tilde{\pi}_2\left(\underline{p}\right)/\tilde{\pi}_2(p)$ for $p \in [\underline{p}, p_1^m)$, with a mass point of size $A/\tilde{\pi}_2(p_1^m) = \tilde{\pi}_2\left(\underline{p}\right)/\tilde{\pi}_2(p_1^m)$ at its monopoly price p_1^m. Next, using $q_1(p) = (A + \pi_1^*)/\left((p - c_1)D(p)\right) = \tilde{\pi}_1\left(\underline{p}\right)/\tilde{\pi}_1(p)$, we pin down a_2 (from Firm 1's chance of winning at its monopoly price) as $a_2 = 1 - q_1(p_1^m) = 1 - \tilde{\pi}_1\left(\underline{p}\right)/\tilde{\pi}_1(p_1^m)$ and therefore Firm 2's conditional-on-advertising price distribution as $F_2(p) = \left(1 - \tilde{\pi}_1\left(\underline{p}\right)/\tilde{\pi}_1(p)\right)/\left(1 - \tilde{\pi}_1\left(\underline{p}\right)/\tilde{\pi}_1(p_1^m)\right)$ for $p \in [\underline{p}, p_1^m)$. All of these expressions reduce to our earlier results for the case of unit demand.

(such as email versus postal mail) or purchase data from different data brokers.[22] This subsection discusses how competition in our model looks like under asymmetric ad costs. While the main results do not change too much, it does become possible for more than two firms to advertise in equilibrium.

Figure 2.4: Probability of Third Firm Entry



In a candidate equilibrium where only Firms 1 and 2 are active, $G_{\max}(\sigma) = G_1(\sigma)\, G_2(\sigma)$ represents Firm 3's chance of winning the sale if it enters with an offer of σ. Firm 3 needs to win with probability $q_3^0(\sigma)$ or greater to break even. If ad costs are homogeneous, as on the left, this is impossible, as the curves will be ranked as shown. But if Firm 3 has a sufficiently large ad cost advantage $A_3 < A_1 = A_2 = A$ (intercept $q_3^0(0) = \frac{A_3}{s_3} < G_1(0)\, G_2(0) = \frac{A}{s_2}\frac{A+s_1-s_2}{s_1}$ as shown on the right), then it can profitably enter with any surplus offer in the range $[0, \hat{\sigma}]$, disrupting the Firm 1/Firm 2 candidate equilibrium.

Suppose Firm i has targeted ad cost $A_i > 0$, and let $z_i = s_i - A_i$ denote Firm i's surplus from a sale net of ad costs. These net surpluses will play an important role when ad costs are heterogeneous; without loss of generality, relabel the firms so that Firms 1 and 2 have the largest and second-largest *net* surpluses: $z_1 > z_2 > z_3 > ... > z_n$. We claim that under this reordering, the following features of our earlier analysis

[22]We thank a referee for raising this point.

are preserved: 1) Firm 1 always advertises and Firm 2 does with positive probability less than one; 2) Firm 1 earns a positive profit equal to its *net* surplus advantage over Firm 2, $\pi_1 = z_1 - z_2$, while all other firms earn zero profit; and 3) the support of surpluses offered to consumers is $[0, z_2]$. The logic mirrors our earlier analysis.[23,24]

The non-identical ad cost case begins to look different when Firm 3 (or any other firm) can profitably enter the market above (i.e., advertise with positive probability). To explore this question, consider a firm's "break-even win probability" $q_i^0(\sigma)$: the chance of winning the sale with surplus offer σ at which Firm i earns zero expected profit. This break-even probability is defined implicitly by $q_i^0(\sigma)(s_i - \sigma) - A_i = 0$. Given our results above, competition from Firm 1 implies that Firm 2's win probabilities in a candidate equilibrium will lie on its break-even line: $G_1(\sigma) = q_2^0(\sigma) = \frac{A_2}{s_2 - \sigma}$. If ad costs are identical, then Firm 3's break-even line lies strictly above Firm 2's as in the left panel of Figure 2.4 – at any surplus offer it would need a strictly better chance of winning than $q_2^0(\sigma)$ to be profitable. But if Firm 3 were to enter with an offer of σ, it would actually win with probability $G_1(\sigma) G_2(\sigma)$ (the dashed line in the figure). This is strictly worse than $q_2^0(\sigma) = G_1(\sigma)$ (since Firm 3 would compete against not just Firm 1 but also Firm 2) and so Firm 3 would

[23] Firm 1 can earn at least $z_1 - z_2$ by overcutting the best surplus offer among its rivals and winning for sure. But it cannot earn strictly more, or else Firm 2 could use the same trick to earn a positive profit, and positive profits for any firm other than Firm 1 are impossible for the reasons laid out in Lemma 2. Thus Firm 1 earns $\pi_1 = z_1 - z_2$, always advertises, and its best surplus offer (which wins with probability one) must be $\bar{\sigma} = s_1 - A_1 - \pi_1 = z_2$. Firm 1 would not make such a generous offer unless it faced competition all the way up to $\bar{\sigma} = z_2$; only Firm 2 can offer that much surplus without losing money, so Firm 2 must advertise with positive probability. For the same reasons as earlier, the one firm that does not always face competition (Firm 1) must have an atom at the worst surplus offer $\underline{\sigma}$ consumers receive (so that other advertising firms – who do always face competition – win often enough with their worst offers to cover their ad costs). Since this offer wins only if there are no competing ads, it will be set at the monopoly level, $\underline{\sigma} = 0$.

[24] If only Firms 1 and 2 are active in equilibrium (as will be the case if no lower-ranked firm has a sufficiently large ad cost advantage), it is straightforward to confirm that a consumer's best-surplus-offer distribution generalizes to $G_{\max}(\sigma) = G_1(\sigma) G_2(\sigma) = \frac{A_2}{s_2 - \sigma} \frac{A_2 + s_1 - s_2}{s_1 - \sigma}$. Thus the consumer surplus conclusions of Proposition 4 generalize with one nuance: while consumers benefit from a reduction in A_2, they are unaffected (at the margin) by changes in A_1.

lose money. However, if Firm 3 can make up for a smaller surplus by advertising sufficiently more cheaply – that is, if $s_2 > s_3$ and $z_2 > z_3$ but $A_3 < A_2$ – then its break-even line $q_3(\sigma) = \frac{A_3}{s_3 - \sigma}$ may cross that of Firm 2 from below (as in the right panel of Figure 2.4). In this case, Firm 3 cannot compete with Firm 2 on generous surplus offers that are likely wins, but it has a comparative advantage in making miserly offers that earn large profit margins but rarely win – it does not need to win as often as Firm 2 would to break even. Firm 3's ad cost advantage must be great enough to make such offers profitably despite the fact that it can win only with probability $G_1(\sigma) G_2(\sigma)$, due to the combined competition of Firms 1 and 2. In this case, as illustrated in the right panel of Figure 2.4, Firm 3 can enter the market (i.e. advertise) over a range of high-price, low-surplus offers. In the appendix, we provide an example of an equilibrium like this: Firm 3 mixes over low-surplus offers (displacing Firm 2), Firm 2 mixes over high-surplus offers, and collectively they provide competitive discipline for Firm 1, which mixes over the entire range. Under the right conditions many additional firms could enter the market in this way, each carving out its own niche, although they must all earn zero profits.

2.6.3 Competition when Information about the Consumer's Tastes is Noisy

We have assumed throughout that firms know the consumer's tastes perfectly, insofar as her reservation prices at each firm are common knowledge in the price competition game – this simplifying assumption has been very helpful in obtaining the clean, sharp results presented thus far. As data mining improves, we may indeed be approaching such a brave new world in which every consumer's heart is laid bare to the market, but we are probably not there quite yet. A more realistic assumption might be that the consumer's reservation prices have an observable component, which firms can infer from data mining, and an idiosyncratic component that the consumer knows but the firms do not.[25] We will argue informally that our model provides a reasonable approximation of this more complex setting when the firms' information is good but not perfect.

Suppose the consumer's reservation value for Firm i's product is actually $\tilde{r}_i = r_i + \mu\varepsilon_i$, where r_i is common knowledge, $\mu \geq 0$ is a scale parameter, and ε_i is a taste shock observed only by the consumer. In the spirit of discrete choice modeling, suppose these taste shocks are drawn i.i.d. from a distribution known to the firms. If there were no targeted advertising costs to consider, we would then have a straightforward discrete choice model of price competition as studied by Anderson, de Palma, and Thisse (1992). A general intuition is that the "noise" in the consumer's preferences tends to soften price competition, permitting firms with lower observable quality to nevertheless carve out some market share and profits. In effect, ad costs would append a simultaneous entry decision to the price competition game: Anderson and de Palma (2001) study a two-stage game with entry decisions preceding pricing. In broad terms, for μ large enough we would expect to see similar results to other entry games: with soft price competition, a larger number of actively advertising firms can be supported in equilibrium as ad costs go down, and there may be multiple equilibria in which different combinations of firms enter.

When the noise in consumer tastes is small, an argument from upper-hemicontinuity of the equilibrium correspondence suggests outcomes will approximate those in our standard model.[26] As $\mu \to 0$ any sequence of equilibria of the μ-game must converge to some equilibrium of the $\mu = 0$ game, but the latter is our standard model which has a unique equilibrium.[27]

[25] We thank a referee for suggesting this line of thought.

[26] In their classic paper on equilibrium in games with discontinuous payoffs, Dasgupta and Maskin (1986, p38) make essentially the same point, arguing that equilibria of a limit game can be constructed as a limit of equilibria of a perturbed game in which expected payoffs have been made continuous by adding exogenous uncertainty.

[27] To flesh out the intuition a bit, for μ small enough, price competition is sharp enough that

2.7 Conclusion

Our results indicate that much of our intuition about standard asymmetric Bertrand competition (e.g. Lederer and Hurter, 1986) still applies when firms must pay to advertise: only the most and second-most competitive firms are relevant, and only the former earns a profit (equal to its competitive advantage over the latter). However, some of the underlying details are quite different: pricing is in mixed strategies, and equilibrium is inefficient due to both wasteful advertising and sales to the second-best firm. Our equilibrium gracefully limits to the conventional Bertrand outcome as ad costs vanish. When firms are homogeneous and ad costs are positive, our equilibrium selection offers an (arguably) more appealing alternative to the counter-intuitive comparative statics of the usual symmetric equilibrium.

In assuming that firms must pay to advertise and in the result that equilibrium prices follow Pareto distributions, our framework resembles Butters' (1977) celebrated model.[28] However, where Butters assumes that ads are matched to individual consumers randomly, we assume a firm knows exactly whom it is targeting with an offer. A secondary distinction is that Butters focuses on symmetric firms, while we stress the importance of asymmetries.

In closing, we note a few directions for future research. In some settings, it may be reasonable to think that consumers are already aware of firms' list prices, but firms can pay to send them targeted ads with personalized discounts. In this case, firms must anticipate the outcome of this targeting game (including which consumers

---

there is only room for one firm to make a positive profit. In particular, a version of Lemma 2 still applies: if two firms were profitable, both would always advertise. But each firm's residual demand is sufficiently elastic that, at their least competitive price offers, it is impossible for each of them to win often enough to cover the ad cost and yet neither has an incentive to undercut the other. Thus we have Firm 1 always advertising and earning a profit close to $s_1 - s_2$, and Firm 2 advertising with positive probability and earning zero profit. Any strictly lower-surplus firm would make strictly less than Firm 2's gross profit were it to advertise, and so could not cover its ad cost.

[28]Similar assumptions and results appear in models based on Varian's (1980) Model of Sales, such as Baye and Morgan (2001) and in various search models in the vein of Stahl (1989).

will be fought over twice) when setting their initial list prices. In a companion paper (2014), we study this setting, incorporating the model in this paper as a second stage.

Our assumptions make sense when firms have access – at a price – to the same extensive data about individual consumer tastes. However, some of the most interesting applications of targeting arise when one firm has better information about a consumer than its rivals do. For example a grocery store may be able to use loyalty card data (that its rivals do not see) to link a consumer to his record of past purchases, giving the store an advantage in crafting personalized offers for him. In our model, consumer taste is summarized by willingness to pay for a single product, and so targeting reduces to a personalized price. In reality, consumer tastes are more complex than this, and effective targeting might involve understanding which product to offer, when to offer a discount (based on forecasting when the consumer will need to restock), whether to bundle products together, and so forth. The supply side of targeted ad provision (which we have treated as a reduced form cost) deserves additional attention. Web search, social media, and advertising platforms are all important spaces where a consumer's tastes are partially observed (*via* cookies on web-sites or by linking social network data to consumption choices, for example) and matched to firms wishing to reach him (often through real-time auctions for advertising placement). While the literature on matching firms to consumers is growing rapidly (see e.g. Athey and Ellison, 2011), much of it suppresses firms' competition in product prices in order to focus on how they compete in markets for an advertising platform's targeting service. Because the targeted ad prices generated by such markets can vary across firms (as in the case of position auctions), our preliminary skirmish with product-price competition in this case (Section 2.6.2) deserves a deeper look.[29] Of course, these suggestions are just the tip of the iceberg – data-rich process-

[29]In addition to the differences in ad prices that firms may face, they can sometimes opt into different *types* of ad pricing, such as cost-per-click, cost-per-impression, or cost-per-action.

ing is quickly changing the landscape of advertising and pricing, and there is much room for future work.

Bibliography

[1] Anderson, Simon, Alicia Baik, and Nathan Larson (2015). "How Targeted Advertising Affects Price Competition, Profits, and Consumers."

[2] Anderson, Simon and André de Palma. 1988. "Spatial Price Discrimination with Heterogeneous Products," *Review of Economic Studies*, 55, 573-592.

[3] Anderson, Simon and André de Palma. 2001. "Product diversity in asymmetric oligopoly: Is the quality of consumer goods too low?," *The Journal of Industrial Economics*, 49, 113-135.

[4] Anderson, Simon, de Palma, André, and Jacques-François Thisse. 1992. *Discrete Choice Theory of Product Differentiation*, MIT Press.

[5] Athey, Susan and Glenn Ellison. 2011. "Position Auctions with Consumer Search," *Quarterly Journal of Economics*, 126, 1213-1270.

[6] Baye, Michael R., Dan Kovenock, and Casper G de Vries. 1992. "It Takes Two to Tango: Equilibria in a Model of Sales," *Games and Economic Behavior*, 4, 493-510.

[7] Baye, Michael R., Dan Kovenock, and Casper G de Vries. 1996. "The All-Pay Auction with Complete Information," *Economic Theory*, 8, 291-305.

[8] Baye, Michael R. and John Morgan. 2001. "Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets," *American Economic Review*, 91, 454-474.

[9] Bester, Helmut and Emmanuel Petrakis. 1995. "Price Competition and Advertising in Oligopoly," *European Economic Review*, 39, 1075-1088.

[10] Bester, Helmut and Emmanuel Petrakis. 1996. "Coupons and Oligopolistic Price Discrimination," *International Journal of Industrial Organization*, 14, 227-242.

[11] Butters, Gerard R. 1977. "Equilibrium Distributions of Sales and Advertising Prices," *Review of Economic Studies*, 44, 465-491.

[12] Dasgupta, Partha and Eric Maskin. 1986. "The Existence of Equilibrium in Discontinuous Economic Games, II: Applications," *Review of Economic Studies*, 53, 27-41.

[13] Grossman, Gene and Carl Shapiro. 1984. "Informative Advertising with Differentiated Products," *Review of Economic Studies*,51, 63-81.

[14] Hillman, Arye L. and John G. Riley. 1989. "Politically Contestable Rents and Transfers," *Economics and Politics*, 1, 17-39.

[15] Hotelling, Harold. 1929. "Stability in Competition," *Economic Journal*, 39, 41-57.

[16] Janssen, Maarten C.W., José Luis Moraga-Gonzalez and Matthijs R. Wildenbeest. 2005. "Truly Costly Sequential Search and Oligopolistic Pricing," *International Journal of Industrial Organization*, 23, 451-466.

[17] Kocas, Cenk and Tunga Kiyak. 2006. "Theory and Evidence on Pricing by Asymmetric Oligopolies," *International Journal of Industrial Organization*, 24, 83-105.

[18] Lederer, Phillip J. and Arthur P. Hurter, Jr. 1986. "Competition of Firms: Discriminatory Pricing and Location," *Econometrica*, 54, 623-640.

[19] Mas-Colell, Andreu, Michael D. Whinston and Jerry R. Green. 1995. *Microeconomic Theory*, Oxford: Oxford University Press.

[20] Narasimhan, C. 1988. "Competitive Promotional Strategies," *Journal of Business*, 61, 427-449.

[21] Palfrey, Thomas and Howard Rosenthal. 1984. "Participation and the Provision of Discrete Public Goods: a Strategic Analysis," *Journal of Public Economics*, 24, 171-193.

[22] Robert, Jacques and Dale Stahl. 1993. "Informative Price Advertising in a Sequential Search Model," *Econometrica*, 61, 657-686.

[23] Samuelson, Paul. 1941. "The Stability of Equilibrium: Comparative Statics and Dynamics," *Econometrica*, 9, 97-120.

[24] Shaffer, Greg and Z. John Zhang. 1995. "Competitive Coupon Targeting," *Marketing Science*, 14, 395-416.

[25] Shaffer, Greg and Z. John Zhang. 2002. "Competitive One-to-One Promotions," *Management Science*, 48, 1143-1160.

[26] Sharkey, William W. and David Sibley. 1992. "A Bertrand Model of Pricing and Entry," *Economics Letters*, 41, 199-206.

[27] Stahl, Dale. 1989. "Oligopolistic Pricing with Sequential Consumer Search," *American Economic Review*, 79, 700-712.

[28] Stahl, Dale. 1994. "Oligopolistic Pricing and Advertising," *Journal of Economic Theory*, 64, 162-177.

[29] Tirole, Jean. 1988. *The Theory of Industrial Organization*, The MIT Press.

[30] Varian, Hal R. 1980. "A Model of Sales," *American Economic Review*, 70, 651-659.

Appendices

.1 Appendix

.1.1 Supporting Analysis for the Symmetric Model

Welfare in the Symmetric and Limiting Asymmetric Equilibria

We start by computing expected social surplus for the symmetric equilibrium when all firms have surplus s_1, using the equilibrium strategies in Proposition 5. Because it does not matter which product the consumer buys, expected social surplus can be decomposed into the expected social value of receiving at least one offer minus total expected advertisement costs:

$$\begin{aligned} SS &= s_1 \Pr(\text{consumer gets an offer}) - \text{Ad costs} \\ &= s_1 \left(1 - G_i(0)^n\right) - An\left(1 - G_i(0)\right) \\ &= s_1 \left(1 - \left(\frac{A}{s_1}\right)^{\frac{n}{n-1}}\right) - An\left(1 - \left(\frac{A}{s_1}\right)^{\frac{1}{n-1}}\right) \end{aligned}$$

Next, we claim the following.

Proposition 7 *With $n = 2$ firms, the symmetric and limiting asymmetric equilibria have the same expected social surplus and the same expected consumer surplus. With more than two firms, expected social and consumer surplus are both strictly higher in the asymmetric equilibrium.*

Recall that in both cases, social and consumer surplus are equal because firms earn zero profits. The first part follows from the observation that when $n = 2$, the distribution of a consumer's best offer is the same under either equilibrium: $G_{\max}(\sigma) = \left(\frac{A}{s_1 - \sigma}\right)^2$. The rest follows from Proposition 6 (and the fact that the asymmetric equilibrium does not change with n). Regardless of n, second-best social surplus would be $s_1 - A$: the consumer must get some product (it doesn't matter which) and this requires sending at least one ad. For both equilibria (symmetric and asymmetric), inefficiency relative to this second-best benchmark may be decomposed into two components, excessive advertising costs and misallocation. For the symmetric equilibrium these costs are

$$\begin{aligned}\text{Social cost of wasteful advertising (Symm.)} &= An\left(1-G(0)\right)-A \\ \text{Social cost of misallocation (Symm.)} &= s_1G(0)^n.\end{aligned}$$

The former is just total expected advertising minus the necessary amount A; the latter reflects failures to make a sale (the only type of misallocation, since products are identical).[30] Under the asymmetric equilibrium, there is no misallocation since the consumer always receives at least one ad and makes a purchase. Thus all inefficiency is due to (socially unnecessary) advertising by Firm 2; this has cost a_2A; hence

$$\text{Social cost of wasteful advertising (Asymm.)} = A(1-\frac{A}{s_1})$$

With two firms, the symmetric and asymmetric equilibria both have total avoidable inefficiency equal to $A(1-\frac{A}{s_1})$, but for slightly different reasons: the latter has more wasteful advertising, but under the former, a sale may be lost due to miscoordinated advertising. It is not a coincidence that these two effects happen to balance out. At a technical level, competitive forces ensure that the consumer's best offer distribution must be $G_1(\sigma)\,G_2(\sigma) = G(\sigma)^2$ under either equilibrium. The difference between equilibria amounts to an interpretation of $G_1(0)$: if this is the probability of an advertised zero-surplus offer, we have the asymmetric equilibrium; if it reflects a failure to advertise at all, we have the symmetric one. In this situation, Firm 1 must be indifferent between advertising $\sigma = 0$ and not advertising (since it must earn zero profit either way). Furthermore, its private incentives on a $\sigma = 0$ offer are aligned

[30]Thus using $G(0) = A/s_1$, total avoidable inefficiency may be written $\chi(A) = s_1\left(A/s_1\right)^{\frac{n}{n-1}} + An\left(1-\left(A/s_1\right)^{\frac{1}{n-1}}\right) - A$.

with social welfare (since it would captures the full surplus from a sale), so, its private indifference implies that the shift would be welfare-neutral.

Asymmetric Equilibria under Symmetry

There are additional equilibria in which an arbitrary subset of $\tilde{n} < n$ of the firms play a version of this equilibrium (with $\tilde{n} \geq 2$ replacing n), while the $n - \tilde{n}$ others sit out (never advertise). As the argument that leads up to (5) makes clear, the equilibrium offer distribution for the remaining (potentially active) firms is symmetric. There remains the possibility that at most one of them advertises a zero-surplus offer with positive probability. Indeed, if two or more firms were to advertise a zero-surplus offer with positive probability then one could profitably undercut and gain a positive sales increase probability from an infinitesimal price cut. To see that one firm could use a zero-surplus advertisement, recall that the probability mass $G_i(0) = \left(\frac{A}{s_1}\right)^{\frac{1}{\tilde{n}-1}}$ in (5) may include a zero-surplus advertisement for some i. This leaves an indeterminacy. For arbitrary $a_i \in \left[1 - \left(\frac{A}{s_1}\right)^{\frac{1}{\tilde{n}-1}}, 1\right]$, any strategy profile in which Firm i refrains from advertising with probability $1 - a_i$, advertises a zero surplus offer with probability $a_i - \left(1 - \left(\frac{A}{s_1}\right)^{\frac{1}{\tilde{n}-1}}\right)$, and the remaining firms refrain from advertising with probability $\left(\frac{A}{s_1}\right)^{\frac{1}{\tilde{n}-1}}$, is an equilibrium. Thus it remains true and consistent with our earlier analysis that at most one firm can have an atom of ads at $\sigma = 0$, but it is no longer necessary that any firm does so, since they all earn zero profit and so are indifferent between advertising and not. Notice though that this indeterminacy has no bearing on equilibrium payoffs.

Pulling this together, there is thus an equilibrium under symmetry at which only two firms are active: one advertises with probability 1, the other with probability $1 - \left(\frac{A}{s_1}\right)$, and for both the offer distribution is $G(\sigma) = \frac{A}{s_1 - \sigma}$. But this is identical to the limiting equilibrium, under asymmetric costs and valuations, as those asymmetries vanish. That is to say, a perturbation approach of beginning with strictly differentiated firms and taking limits as the gaps among the top n firms vanish will select this asymmetric two-firm equilibrium in the symmetric limit, not the symmetric n-firm equilibrium.

.1.2 Other Symmetric Cases

We have previously considered the case in which there are no ties in the surpluses s_i that firms offer and the symmetric case in which all surpluses are equal. This section evaluates the remaining cases in which some subsets of the firms are identical; thus we set $s_1 \geq s_2 \geq ... \geq s_n$. There are three main cases to consider, depending on the highest rank at which any firms tie.

Low Ties

The easiest to dispense with is the case in which any ties are among firms at the level of Firm 3 or worse; that is, $s_1 > s_2 > s_3 \geq ... \geq s_n$. It should be clear that this will not affect the equilibrium outcome – a few of the supporting lemmas must be amended slightly, but Proposition 1 still applies.

Dominant Firm and Fringe Firms

Next suppose that m firms tie for the second-ranked spot (whether or not there are ties below the second-ranked position will be irrelevant) : $s_1 > s_2 = s_3 = ... = s_{m+1} > s_{m+2} \geq ...$. It is straightforward to prove that *any equilibrium must have strictly positive profits for Firm 1, zero profits for the other firms, including the m runners-up, and only Firm 1 and some subset of the runners-up advertising with positive probability.* As earlier, let $\Delta_{12} = s_1 - s_2$ be the advantage of Firm 1 over the runners-up, and let $G_i(\sigma)$ be the distribution of the surplus offered by Firm i, with a failure to advertise included as an offer of $\sigma_i = 0$. Likewise, define $G_{-i}(\sigma) =_{\{j \leq m+1 \,:\, j \neq i\}} G_j(\sigma)$, the distribution of the best opponent surplus offer faced by Firm i. Arguments similar to those earlier can be used to establish that each of these "best opponent" distributions has support on $[0, s_2 - A]$. Similar arguments establish that Firm 1's equilibrium profit is $\pi_1 = \Delta_{12}$: any firm can win with probability one by advertising the upper bound surplus and when Firm 1 does so it charges a price that is Δ_{12} higher than the other firms, thereby earning $\pi_2 + \Delta_{12} = \Delta_{12}$. Before examining other possibilities, first consider the candidate equilibrium in which the m runners-up behave symmetrically. Firm 1's indifference over its mixed strategy support implies that its probability of winning with an offer of σ_1 is no different now that it has m rivals than it was when it faced one (under the assumptions of Proposition 1); that is,

$$G_{-1}(\sigma) = \left(G_2(\sigma)\right)^m = \frac{A + \Delta_{12}}{s_1 - \sigma}$$

and so $G_2(\sigma) = \left(\frac{A+\Delta_{12}}{s_1-\sigma}\right)^{1/m}$. Similarly, indifference for each runner-up implies that it must face the same best-opponent distribution that Firm 2 did in Proposition 2; this implies $G_{-i}(\sigma) = G_1(\sigma)\left(G_2(\sigma)\right)^{m-1} = \frac{A}{s_2-\sigma}$ for each $i \in \{2, ..., m+1\}$, and so

$$G_1(\sigma) = \left(\frac{A}{s_2 - \sigma}\right)\left(\frac{A + \Delta_{12}}{s_1 - \sigma}\right)^{-\frac{m-1}{m}}$$

The consumer's best offer is then distributed according to $G_{\max}(\sigma) = \left(\frac{A}{s_2-\sigma}\right)\left(\frac{A+\Delta_{12}}{s_1-\sigma}\right)^{1/m}$. Notice that at both the top and second-ranked firms the consumer has a positive chance of not being offered a strictly positive offer. To complete the description of equilibrium, we must establish whether the probability $G_1(0) > 0$ reflects Firm 1 advertising a zero surplus offer or not advertising, and similarly for $G_2(0)$. Because Firm 1 earns positive profits, it must advertise with probability one, and so the probability mass $G_1(0)$ must represent an atom of advertised zero surplus offers. There cannot be more than one firm advertising an atom of zero surplus offers, as each would have a strict incentive to undercut, and so we must have $G_2(0) = 1 - a_2$ for each of the second-ranked firms.

Having established this template for a symmetric equilibrium, and noting that each of the tied firms is indifferent to not advertising, it is straightforward (cf. Section .1.1) to show that there is a family of additional equilibria in which a subset $\tilde{m} < m$ of the tied firms advertise using the strategies above (with $\tilde{m} \geq 2$ substituted for m), and the remainder "sit out." *A priori*, it is not clear which of these equilibria should be preferred over the others; absent a reason to distinguish between the tied firms, one might argue for the "equal treatment" – and hence symmetric – equilibrium in which they all advertise. However, once again such an equilibrium is unstable. Our preferred approach is to begin with the generic case of unequal $\{s_2, ..., s_{m+1}\}$ and select the limiting equilibrium as differences between these firms vanish. As per our earlier analysis, this approach selects a limit equilibrium in which one firm (Firm 2) advertises and the other $m-1$ runners-up sit out. These two alternative equilibrium selections agree on firm profits, but disagree on price distributions, probabilities of advertising for the runners-up, and consumer surplus. In particular, the consumer is better off in the equilibrium where only Firms 1 and 2 are active.

Top Tie

Finally, suppose that m firms tie for the top spot: $s_1 = s_2 = ... = s_m > s_{m+1} \geq ...$. In this case, only firms at the top will ever advertise, and they all must earn profit zero. Indeed, if one of the lower-ranked firms j were to advertise in equilibrium, then it would have to be the case that all m top firms earn strictly positive profits. (If not, a top firm earning zero could profitably deviate to undercutting j's best offer.) But then all m top firms would have to be advertising with probability one, and this is impossible for the reasons laid out in Lemma 2.[31] Consequently, ties below the top

[31]The arguments in Lemma 2 rule out strictly positive profits for more than one of the top firms. Furthermore, if any single firm, say Firm 1, were to earn strictly positive profits in equilibrium, then any of the other top firms could undercut Firm 1's lowest advertised price and earn strictly positive

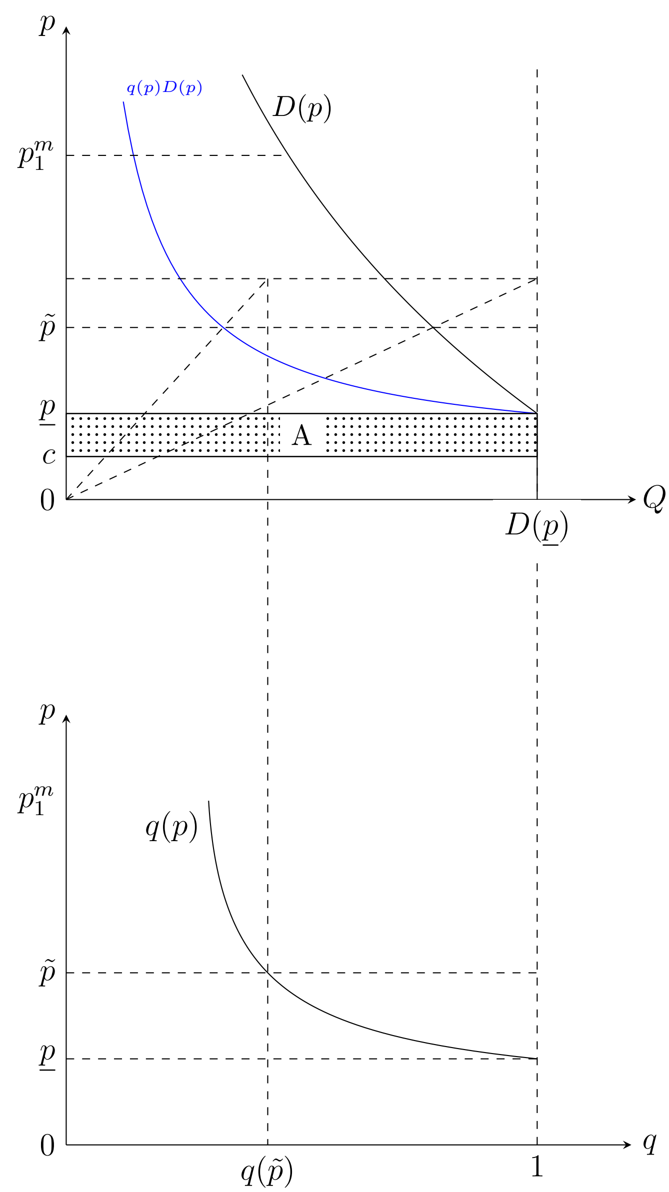
level will be irrelevant. Thus the analysis of Section 2.5 covers this case.

.1.3 Downward-Sloping Demand: Additional Analysis

---

profits as well.

Figure 5: Symmetric Downward-sloping Demand



We next derive the symmetric equilibrium under symmetric costs, c, and we show graphically how to tie down the equilibrium sales probability that underpins the equilibrium price distribution. Because all firms are indifferent between advertising and not, all earn zero profit. This ties down the lowest price in the support as the (unique) solution to $\left(\underline{p}-c\right) D\left(\underline{p}\right)=A$. Let $q\left(p\right)$ be the common equilibrium probability of a sale, which is determined by

$$(p-c)\, D\,(p)\, q\,(p)=A \quad \text{for all } p\in[\underline{p},p^{m}),$$

where p^{m} is the (common) monopoly price. The solution is illustrated in Figure 5. The top panel shows the price support as determined from the demand curve, along with the iso-profit line passing through $\left(\underline{p}, D\left(\underline{p}\right)\right)$ that determines the equilibrium $q\left(p\right)$. The value of q as a function of p is shown in the lower panel by the device of associating $D\left(\underline{p}\right)$ to the maximum possible value $q=1$. Then the other q values can be found by drawing a ray from the origin through the demand $D\left(p\right)$ associated to any specific price in the support, finding where the ray reaches the value of $D\left(\underline{p}\right)$ and then finding q from where the ray through the rectangular hyperbola reaches the same height. That is, we simply use similar triangles to find the ratio q which is the ratio of the horizontal distance to the rectangular hyperbola over the distance to the demand curve. (Notice that a similar device can be used in the asymmetric case of the Figure to tie down the equilibrium q's there.)

Hence, we tie down the equilibrium conditional price distribution by $q\left(p\right)=\left(1-aF\left(p\right)\right)^{n-1}$. Note that the common equilibrium advertising probability, a, is tied down from the condition that the monopoly price returns zero profit, i.e.,

$$(1-a)^{n-1}\,\pi^{m}=A$$

or $a=1-\left(\frac{A}{\pi^{m}}\right)^{\frac{1}{n-1}}$. Then the probability that there is no ad at all (and so the market is not served) is

$$(1-a)^{n}=\left(\frac{A}{\pi^{m}}\right)^{\frac{n}{n-1}}$$

which increases in n, reflecting the earlier result (see Proposition 6) that social surplus worsens with more competition. As we argued for the rectangular demand case, the symmetric equilibrium is unstable with respect to cost heterogeneity.

.1.4 Advertising by Three Firms when Ad Costs Differ

While it is beyond our scope to give a general analysis of strategies when more than two firms are active, a worked example may help to suggest what the general case looks like. There are three firms with surpluses $s_1 > s_2 > s_3$. Firms 1 and 2 have common ad cost $A_1 = A_2 = A$, while the low surplus Firm 3 has a cost advantage in advertising: $A_3 < A$. We assume this cost advantage is not too great, so that the net surplus ranking $z_1 > z_2 > z_3$ mirrors the gross surpluses. Thus Firms 1 and 2 will be active in equilibrium, and it remains to be seen whether Firm 3 will be.

As a stepping stone toward finding the equilibrium, it is helpful to set up the equilibrium that would prevail if only Firms 1 and 2 were present and then ask whether Firm 3 can disrupt it. Consulting earlier results, the distributions of surplus offers by Firms 1 and 2 respectively would then be $G_1(\sigma) = \frac{A}{s_2-\sigma}$ and $G_2(\sigma) = \frac{A+\Delta s}{s_1-\sigma}$, where $\Delta s = s_1 - s_2$, with the distribution of the best surplus offer between them given by $G_{\max}(\sigma) = G_1(\sigma)\, G_2(\sigma) = \frac{A}{s_2-\sigma}\frac{A+\Delta s}{s_1-\sigma}$. This tells us how often Firm 3 would win if it were to enter with a surplus offer σ; if $G_{\max}(\sigma)$ ever lies above Firm 3's zero-profit line $q_3(\sigma) = \frac{A_3}{s_3-\sigma}$, then Firm 3 can enter profitably and disrupt the two-firm equilibrium. A sufficient condition for this is $q_3(0) < G_{\max}(0)$, or $\frac{A_3}{s_3} < \frac{A}{s_2}\frac{A+\Delta s}{s_1}$, as illustrated in Figure 2.4. For the purpose of the example, suppose this condition does hold. As indicated in the figure, let $\hat{\sigma}$ be the largest surplus offer Firm 3 could make without losing money, given the strategies Firms 1 and 2 would use if Firm 3 were not present.

In the full game with three firms, write $H_i(\sigma)$ for the distribution of surpluses that Firm i offers, with the understanding as before that $H_i(0) > 0$ represents an atom at $\sigma = 0$ for Firm 1 and declining to advertise for Firms 2 and 3. We claim there is an equilibrium with the following features:

1. Firm 2 specializes in low-price, high-surplus offers with support $\sigma \in [\hat{\sigma}, z_2]$, while Firm 3 specializes in high-price, low-surplus offers with support $\sigma \in [0, \hat{\sigma}]$. Firm 1 mixes over the entire range $\sigma \in [0, z_2]$.

2. At surplus offers above $\hat{\sigma}$, where Firm 3 does not compete, Firms 1 and 2 behave just as they would have done if Firm 3 were absent. That is, $H_1(\sigma) = G_1(\sigma)$ and $H_2(\sigma) = G_2(\sigma)$ for $\sigma \in [\hat{\sigma}, z_2]$.

3. Any probability that Firm 2 would have assigned to low-surplus offers $\sigma \in [0, \hat{\sigma}]$ is simply reassigned to not advertising. Consequently, Firm 2 advertises less often than it would have done if Firm 3 were absent.

4. Firm 1 is forced to compete more aggressively over the low-surplus offers than it would have done otherwise; that is, $H_1(\sigma) < G_1(\sigma)$ for $\sigma \in [0, \hat{\sigma})$. In particular, it competes just hard enough to drive Firm 3's profit on such offers down to zero. This requires it to more than compensate for the loss of competition from Firm 2 over this range (since Firm 3 could have made positive profits against the competition of both Firms 1 and 2, under their original strategies).

5. Collectively, Firms 2 and 3 provide Firm 1 exactly as much competition at every surplus level as Firm 1 would have faced in the absence of Firm 3. That is, $H_2(\sigma) H_3(\sigma) = G_2(\sigma)$ for $\sigma \in [0, z_2]$. In particular, the probability that no competitor to Firm 1 advertises (given by $H_2(0) H_3(0)$) does not change when Firm 3 is present.

6. While the presence of Firm 3 does not affect any firm's profit, it does make consumers better-off. This is driven entirely by the first-order stochastic improvement in Firm 1's least competitive ($\sigma \in [0, \hat{\sigma})$) offers. Consequently, the presence of Firm 3 improves welfare as well.

In order to verify these claims, first recall that in the candidate equilibrium with only Firms 1 and 2, the equilibrium strategies can be summarized by the surplus offer distributions $G_1(\sigma) = \frac{A}{s_2-\sigma}$ and $G_2(\sigma) = \frac{A+\Delta s}{s_1-\sigma}$, with $\Delta s = s_1 - s_2$ (see Proposition 1). Defining $G_{\max}(\sigma) = G_1(\sigma) G_2(\sigma)$ and $q_3^0(\sigma) = \frac{A_3}{s_3-\sigma}$ as in the text, we have $q_3^0(0) < G_{\max}(0)$ and $q_3^0(z_3) = 1 > G_{\max}(z_3)$. Then let $\hat{\sigma}$, defined by $q_3^0(\hat{\sigma}) = G_{\max}(\hat{\sigma})$ be the point on $[0, z_2]$ at which the two functions cross; one can confirm that this crossing is unique. We claim the following surplus offer distributions represent equilibrium strategies for Firms 1, 2, and 3.

$$H_1(\sigma) = \begin{cases} \frac{q_3^0(\sigma)}{G_2(\hat{\sigma})} = \frac{A}{s_2-\hat{\sigma}} \frac{s_3-\hat{\sigma}}{s_3-\sigma} & \text{if } \sigma \in [0, \hat{\sigma}) \\ G_1(\sigma) & \text{if } \sigma \in [\hat{\sigma}, z_2] \end{cases}$$

$$H_2(\sigma) = \begin{cases} G_2(\hat{\sigma}) & \text{if } \sigma \in [0, \hat{\sigma}) \\ G_2(\sigma) & \text{if } \sigma \in [\hat{\sigma}, z_2] \end{cases}$$

$$H_3(\sigma) = \begin{cases} \frac{G_2(\sigma)}{G_2(\hat{\sigma})} = \frac{s_1-\hat{\sigma}}{s_1-\sigma} & \text{if } \sigma \in [0, \hat{\sigma}) \\ 1 & \text{if } \sigma \in [\hat{\sigma}, z_2] \end{cases}$$

To confirm this, write $H_{-1}(\sigma) = H_2(\sigma) H_3(\sigma)$ for the distribution of the most generous rival surplus offer faced by Firm 1, and similarly for $H_{-2}(\sigma)$ and $H_{-3}(\sigma)$. By construction, $H_{-1}(\sigma) = G_2(\sigma)$ for $\sigma \in [0, z_2]$. But we know from the Firm 1/Firm 2 equilibrium that this makes Firm 1 indifferent over all surplus offers in $[0, z_2]$ (and thus willing to mix over this range). Firm 3 faces best opponent offer distribution

$$H_{-3}(\sigma) = \begin{cases} q_3^0(\sigma) & \text{if } \sigma \in [0, \hat{\sigma}) \\ G_1(\sigma) G_2(\sigma) & \text{if } \sigma \in [\hat{\sigma}, z_2] \end{cases}$$

But as we argued earlier, any offer σ by Firm 3 that wins with probability $q_3^0(\sigma)$ earns it zero profit, while for $\sigma > \hat{\sigma}$, an offer that wins with probability $G_1(\sigma) G_2(\sigma)$ loses money. Thus mixing over $[0, \hat{\sigma})$ is a best reply for Firm 3. Finally, Firm 2 faces best opponent offer distribution

$$H_{-2}(\sigma) = \begin{cases} \frac{q_3^0(\sigma)}{G_2(\hat{\sigma})} \frac{G_2(\sigma)}{G_2(\hat{\sigma})} & \text{if } \sigma \in [0, \hat{\sigma}) \\ G_1(\sigma) & \text{if } \sigma \in [\hat{\sigma}, z_2] \end{cases}$$

From the Firm 1/Firm 2 equilibrium, we know that any offer $\sigma \geq \hat{\sigma}$ will earn Firm 2 zero profit. Furthermore, the threshold $\hat{\sigma}$ is defined by the fact that $q_3^0(\sigma) < G_1(\sigma) G_2(\sigma)$ for $\sigma < \hat{\sigma}$. Thus for $\sigma < \hat{\sigma}$,

$$H_{-2}(\sigma) = \frac{q_3^0(\sigma)}{G_2(\hat{\sigma})} \frac{G_2(\sigma)}{G_2(\hat{\sigma})} < G_1(\sigma) \left(\frac{G_2(\sigma)}{G_2(\hat{\sigma})} \right)^2 < G_1(\sigma)$$

Because Firm 2 needs to win with probability $G_1(\sigma)$ to break even, any offer $\sigma < \hat{\sigma}$ will lose money. So mixing over $[\hat{\sigma}, z_2]$ is a best reply for Firm 2. Thus the stipulated strategies are a Nash equilibrium, as claimed.

The probabilities that Firm 2 and 3 advertise can be recovered as $a_2 = 1 - H_2(0) = 1 - G_2(\hat{\sigma})$ and $a_3 = 1 - H_3(0) = \frac{\hat{\sigma}}{s_1}$ respectively. This represents a cutback for Firm 2 since it would have advertised with probability $1 - G_2(0)$ if Firm 3 were absent. However, the probability that neither competitor to Firm 1 advertises, $(1 - a_2)(1 - a_3) = G_2(\hat{\sigma}) q_3^0(0) = \frac{A + \Delta s}{s_1} = G_2(0)$, is exactly what it would have been if Firm 3 were absent. (Of course this should not be too surprising since the aggregate competition for Firm 1 is pinned down by its indifference condition, regardless of how many rivals it has.) For Firm 1, because $q_3^0(\sigma) < G_1(\sigma) G_2(\sigma)$ for $\sigma < \hat{\sigma}$, we have

$$H_1(\sigma) = \frac{q_3^0(\sigma)}{G_2(\hat{\sigma})} < G_1(\sigma) \frac{G_2(\sigma)}{G_2(\hat{\sigma})} < G_1(\sigma) \quad \text{for } \sigma < \hat{\sigma}$$

So over the range of high-price, low-surplus offers $\sigma \in [0, \hat{\sigma})$, the presence of Firm 3 induces Firm 1 to shift weight toward more competitive offers. In particular, the probability that Firm 1 advertises its monopoly price declines from $G_1(0)$ to $H_1(0)$.

Because $H_2(\sigma) H_3(\sigma) = G_2(\sigma)$, the best offer from Firms 2 and 3 is statistically equivalent to the best offer from Firm 2 in a game where Firm 3 is absent. Consumer surplus can be derived from the distribution of the consumer's best offer, $H_{\max}(\sigma) = H_1(\sigma) H_2(\sigma) H_3(\sigma) = H_1(\sigma) G_2(\sigma)$. But then since $H_1(\sigma) \leq G_1(\sigma)$ (strictly on $[0, \hat{\sigma})$), $H_{\max}(\sigma) \leq G_1(\sigma) G_2(\sigma) = G_{\max}(\sigma)$. Thus consumer surplus improves, and this improvement can be attributed to more competitive offers by Firm 1 in the range $\sigma \in [0, \hat{\sigma})$ where it must compete with Firm 3. Since total firm profits do not change, welfare must rise as well. This improvement involves several effects. Allocative efficiency tends to rise, as Firm 1 competes harder and wins more often, but there is a countervailing effect because some of Firm 2's wins shift to the lower-surplus Firm 3. It is not too hard to show that total advertising is greater when Firm 3 is present. (Simple algebra establishes that if $(1 - a_2)(1 - a_3) = 1 - a_2^{old}$, then $a_2 + a_3 > a_2^{old}$.) However, advertising shifts from the higher cost Firm 2 to the lower cost Firm 3.

The equilibrium has a certain intuitive appeal: in effect the runner-up firms keep the top firm honest and discipline its profits by providing a sort of competitive upper envelope, each one turning up the heat on Firm 1 along the range where it has a comparative advantage. Furthermore, the implication that adding additional firms to the mix can only have a neutral to positive effect on welfare seems likely to be general: we know that total firm profits cannot change, nor can the most competitive non-Firm 1 offers (by Firm 1's indifference condition), but the additional competition may induce better offers out of Firm 1, and hence greater consumer surplus. This would not be obvious *a priori*, since additional active firms also could be associated with misallocation of the sale to a lower-surplus firm and socially excessive advertising.

Chapter 3

Personalized Advertising: A Theoretical and Empirical Analysis of the Customer Purchase Path

3.1 Introduction

Technological advancements are revolutionizing the degree to which retailers are able to personalize advertisements relevant to consumers' revealed preferences. Increased capacity to track customer online, purchase, and geo-location behavior as well as the decreased costs of storing and analyzing information have greatly increased retailers' capacity to track and identify customers' behavior and send messages relevant to them. Companies like Epsilon and Axciom have thousands of attributes attached to nearly all households with demographic data and psychological traits gathered from deep purchase history databases (The Aisles Have Eyes, 2017).

Consumers are also more likely to respond positively to a personalized advertisement relative to a generic advertisement. Overloaded with information, customers have little capacity and desire to internalize messages which are not relevant to their specific tastes. Additionally, with an abundance of options easily available online and in the store, customers are more amenable to personalized recommendations. As a customer analytics partner wrote, personalization "lies at the center of omnichannel marketing strategies [for retailers]" and that they will work to "curate the right product information at the right time in the shopping process." (The Aisles Have Eyes, 2017).

The retailing environment is additionally providing the catalyst for increasing use of personalized advertisements. Retailing is in the midst of a significant transformation akin to the evolution from competitive merchant markets in the 19th century to mass production and retail markets in the early 20th century. Just as small merchants were replaced by more operationally efficient retailers with greater product variety, the same brick and mortar retailers are threatened to be replaced by more nimble and responsive online retailers like Amazon. Within this highly competitive context, brick and mortar retailers are actively seeking ways to differentiate themselves along customers' purchase path.

When analyzing targeted advertisements, theoretical and empirical models have focused on the consumer's direct response to the targeted product, either through the purchase of the product or the click through rate of the online advertisement. While this metric is important for understanding the ROI of an advertisement, it misses the impact of the advertisement on the consumer's purchase path at the store.

In this paper, I develop a theoretical model of personalized advertisements and their impact on customers' purchase paths. I am the first to formally model the effect of advertisements on expanding customers' consideration set within a bounded rationality context. Utilizing Eliaz and Spiegler (2011) as the foundation for the model, customers decide whether to consider advertisements prior to evaluating the products in the advertisements relative to their status quo product. I extend the model to account for customers' status quo purchase paths in the grocery context. I also allow personalized advertisements to be more likely to be considered by customers. The model predicts that personalized advertisements that promote products outside of customers' status quo shopping paths will increase consumers' consideration sets the most and have the greatest impact on consumers' sales. Personalized advertisements along customers' purchase path will have little effect on expanding the consideration set or on sales unless they shift customers' store ordering. Finally, generic advertisements will have little impact on consideration sets or sales because they are less likely to be considered.

With a unique dataset from a single unidentified national grocery retailer, I am able to empirically evaluate the impact of personalized and generic advertisements on the overall purchase path of customers. The two targeting tools used by the retailer are reward and promotional coupon campaigns. Rewards offer discounts for a broad set of frequently purchased products. Since the products are frequently purchased, they have a higher redemption rates. Promotion campaigns recommend and offer discounts for a specific set of products relevant to a targeted customer group. The same discounts are given to a group of customers selected for a campaign.

The data contains the complete purchase history of $2,500$ households in 582 store locations over a two year period. A subset of households receives one or both of each type of targeted coupon campaign. After 33 weeks of the sample period, the retailer began sending reward and promotion coupon campaigns to a subset of its customers. The first weeks of the sample serve as a control period during which no household receives any targeted campaigns. While both campaign types are primarily sent to households exhibiting loyalty through frequent visits or high weekly sales, the retailer intentionally uses test and control groups within loyalty segments to assess the effectiveness of the campaigns.

As predicted in the theoretical framework, I find that promotional campaigns increase sales more on average than reward campaigns both when coupons are redeemed and when they are not redeemed. Additionally, I find that promotional campaigns increase sales in the departments of the promoted products, confirming the theoretical prediction that customers increase their consideration set when they leave their status quo purchase path to evaluate a new product. Reward campaigns have a higher redemption rate, but when customers do not redeem products, customers on average spend less at the store. Additionally, redemptions in the department of the promoted products increase sales by less than promoted products. Generic mailers and displays have little impact on store sales. I also identify the impact of the marketing tools on store visit and find that promotion campaigns are the most effective, followed by reward campaigns and then mailers. I control for endogeneity of the targeted advertisements and selection of store visits with control functions within a sample selection model.

3.2 Literature Review

3.2.1 The Evolution of the Economics of Advertising

Advertising reflects the character of the nexus between firms and consumers. As such, as the nature of the market interaction has evolved since the 19th century to now, the role of advertising has changed. Within the context of competitive merchant markets of the 19th century, economists were naturally skeptical about the role of advertising. Bagwell (2007) outlines an overall disregard for advertising by early economists in his comprehensive review, *The Economic Analysis of Advertising*. Bagwell attributes the dearth of attention to a general focus on perfect competition for firms and conventional assumptions of perfect information with regard to prices and quantities for customers.

However, at the turn of the 20th century, with the development of mass production and retail markets, brand advertising played a more important role in bridging the gap between the consumer and the more productive firm. Bagwell explains that the economies of scale now achievable through rapid innovations were only realized through matched demand. This laid the groundwork for mass marketing activities.

Much of what has defined our understanding of advertising stems from foundational work from this time. Bagwell (2007) outlines the three main views of advertising as persuasive, informative, and complementary. Popular in the early 20th century, the persuasive view of advertising suggests that advertisements alter customer preferences and create brand loyalty. The implication is that advertising can have anti-competitive effects by adjusting customers' view of reality. With the development of the economics of information in the 1960s, informative advertising became the more prominent theoretical approach to advertising (see Renault's chapter in the Handbook of Media Economics, 2015). The view that advertising is informative relies on the assumption that consumers have imperfect information about their options. This market inefficiency is resolved by firms sending more information about the price and location of their products. Finally, the complementary view of advertising argues that advertising enhances the value of the advertised product, directly increasing the well-being of the consumer. I will expand on the literature of persuasive and informative advertising below.

Reflective of the prominence of big business and mass marketing, economic modeling of advertising from this period highlights the one-way communication from firms to consumers. Because firms monopolized information about their products, consumers relied on the firm to communicate the price, location, and characteristics of the products. This information asymmetry implicitly characterizes consumers' dependence on firms for advertisements to better inform their decision making.

Assumptions of fully rational customers may have also been reflective of the period. Relative to now, consumers faced a limited set of options and accessed most of their information through mass media such as the radio, TV, newspapers, and billboards. Even though customers have imperfect information which may requires them to engage in costly search, rational customers are assumed to be able to process information received from advertisements, internalize the value of the advertised product relative to alternative options and optimally choose the product that best suits their well-defined preferences. Preferences are assumed to be complete and transitive, enabling the employment of the utility framework.

As access to information has exploded with the increasing connectedness through the internet and smartphones, market structure is changing again. A 2015 Pew Research report finds that one-fifth of Americans report going online 'almost constantly' and 73 percent of Americans go online every day. Internet connectedness is even more pronounced with young adults (18-35), of which 36 percent report going online almost constantly and 50 percent report going online multiple times a day. In this context, consumers have no shortage of information. Rather, as Anderson and de Palma (2012) explain, they lack the attention needed to process the information. In this context, firms now have to compete for the attention of customers.

Additionally, firms no longer monopolize information about their products and customers no longer readily accept advertising messages. Customers not only can easily search for information about products online and can evaluate through other customer reviews. As a result of increased access to information, customers no longer trust advertisements. In fact, according to an AdAge article, fewer than 25 percent of consumers trust advertisements in print and even fewer in digital forms.

However, even as the complexities of choices have increased, many models have maintained an assumption of consumer rationality. As Spiegler (2011) notes, while the "rational choice paradigm allows preferences to be defined over very general domains," many I.O. applications narrowly define the consequences of rationality, "fully specified by the amount of money the consumer pays and the quantity or quality of the product he consumes." Additionally, even when imperfectly informed, rational agents are assumed to have perfect ability to accurately calculate Bayesian inferences according to correct knowledge of the market. While the motivation of these characteristics may have appropriately described the context in which they were developed, the new realities of limitless product choices and unbounded information to process challenge full rationality assumptions undergirding many advertising models in this new context.

Theories of bounded rationality in consumer choice are emerging as a way to address the shifting realities facing consumers as they interact with firms. In his survey, *Bounded Rationality and Industrial Organization*, Spiegler (2011) identifies three motivations for market models with bounded rational consumers. First, it accounts for the observations of changing realities that have traditionally motivated economic thought. Second, it addresses a growing sentiment that certain economic phenomena including advertising are not adequately captured with standard rational-choice models. Third, experimental psychologists have substantially shown that decision makers deviate from the model of rational choice practiced by most economists.

It is within the context of over-abundance of information that I address personalized marketing tools. Using the framework of Eliaz and Spiegler (2011), I account for limited consideration sets biasing the status quo. Marketing tools are used to expand consumers' consideration of new products. However, this consideration is bounded by the consumers' consideration function. As far as I have been able to see, I am the first to directly apply bounded rationality to evaluation of personalized advertising. This represents a deviation from much of the targeted advertising literature which has derived foundational assumptions from standard informative advertising framework in which consumers are fully rational. I will outline this literature below.

In the rest of this section, I will survey seminal works in the persuasive and informative advertising literature. Next, I will outline important papers accounting for competition for attention in the information age. I will then discuss work on targeted advertising. Next, I will summarize the bounded rationality literature as it relates to advertising. Finally, I will discuss marketing literature related to targeted advertising and consideration sets.

3.2.2 Persuasive Advertising

Persuasive advertising models argue that advertising had an anti-competitive effect by increasing brand loyalty and manipulating consumer preferences. The papers described here are also highlighted in the more comprehensive survey of persuasive advertising literature in Bagwell (2007). After Chamberlin (1933) first identified a persuasive (as well as informative) role in advertising by noticing that advertisements that altered tastes shifted the demand curve out and added some inelasticity, his contemporary Braithwaite (1928) expounded more deeply into how advertising had such anti-competitive effects. Observing the advertising activity of manufacturers trying to emerge out of competition by differentiating themselves from one another, Braithwaite saw advertising as a means of convincing customers of more quality differentiation than actually existed. She explained that advertising was a "selling cost" which persuaded products had more value than they actually possessed. Citing campaigns such as the American Face Brick Campaign meant to combat "propaganda advocating the use of lumber throughout the building of a house" and the Greeting Card Association's campaign to extend sales beyond the holiday season, Braithwaite argues that manufacturers shifts out demand, but at a cost to the consumer which can only be offset by economies of scale dropping prices (1928, p. 22). Advertising also deters entry by giving incumbent firms reputations against which entrants have a hard time competing.

Kaldor (1950) expands on this persuasive advertising view by distinguishing between the direct and indirect effects of advertising. The direct effect of advertisements are to transmit information including the price and quantities to the customers. However, while it does convey information, he notes that the information is given by a biased party and therefore is meant to be persuasive rather than purely informative. This also undermines the efficiency of the transmission of information, he argues.

The indirect effect of advertisement is greater concentration of markets caused by economies of scale of advertising which benefit larger firms and impedes smaller firms (1950, pg. 17).

In the context of growing market concentration and increasingly sophisticated advertising tools meant to build brand loyalty, economists studying advertising in the early 20th century identified a persuasive role in advertising. They argued that advertising created little value except in potentially expanding economies of scale for the production of manufacturers trying to gain market share. However, this gain was limited compared to the lost consumer surplus from expanding the demand beyond customers' unadulterated preferences.

While the context is different, building loyalty remains a goal amongst manufacturers and retailers today. In fact, the cited goal of the reward campaigns is to maintain and build retailer loyalty. However, major differences exist between the examples given in the cited literature and the context of my analysis. First, the advertisement mechanism of the coupon naturally benefits the customer more than a jingle on the radio or a featured advertisement in a newspaper. While both intend to increase demand, the reward coupon does so by decreasing the price for the customer while the latter does so by increasing the perceived value of the products. Second, the reward coupon campaigns are more customer-focused by signaling to the customer that the retailer has identified products most important to her while the persuasive advertisements listed above try to convince customers to value the products being advertised. This signal may in fact cause the customer to be more loyal, but the loyalty is derived from highlighting what is valuable to the customer which may be more informative than persuasive.

3.2.3 Informative Advertising

As Bagwell (2007) notes, the view that advertising is pro-competitive gained a lot of prominence in the 1960's under the leadership of the Chicago School of economics. Instead of seeing advertising as anti-competitive, economists who advocated this view argued that advertising benefited customers and enabled them to make better choices. The significant increase in available choices relative to the early 20th century caused economists to see consumers' imperfect information as a significant hindrance to efficient market outcomes.

Renault's chapter in the *Handbook of Media Economics* (2015) stresses that costly consumer search primarily motivated this shift in analysis. Diamond's paradox in 1971 highlights how market power can develop when it is hard for consumers to learn about the options available to them. The surprising result of consumers paying monopoly prices because of costly search is remedied through informative advertising, according to Renault. By providing information about the product, the price, and location, customers are more able to identify the options available to them. Stigler (1961) formally outlines how informative advertising enables consumers to make better decisions. Stigler argues "price dispersion is a manifestation–and, indeed, it is the measure–of ignorance in the market" caused by costly search (p. 214). Informative advertisements, then, reduce consumers' search costs by providing valuable information about the existence, location, and price of the products directly to customers.

As Renault (2015) argues, the link between costly search and advertisements is central to the view of informative advertising. I agree with Renault that costly search is certainly still a key impediment to competitive markets. However, I will argue that the source of the cost is different from the sources which motivated Stigler. Whereas the source of the cost was time, effort, and lack of access to information to examine options (for example, needing to go through a listing in the Yellow Pages) in the

Stigler model, the cost today is much more internal to the customer. As Van Zandt (2004) and Anderson and de Palma (2009) argue, information congestion is the major source of cost for consumer search in the Information Age. With limited attention, customers are unable to process each of the advertising messages they receive. As a result, they internalize only a fraction of the messages sent. I argue that this information congestion is a cause for limited consideration sets in my model.

Butters (1977) builds on the foundational concepts of costly search and informative advertising, similarly finding that increased costs of search and costs of advertisements increase the price dispersion in the market. He is the first to derive equilibrium outcomes in this framework. In his model, informative advertisements provide direct information about the product and its price for homogenous goods to customers. He identifies diminishing returns to advertisements which imply an increasing marginal cost of reaching the next customer even though the marginal cost of more advertisements is constant (see Stahl, 1994). This observation is important in the context of technologies which enable targeted advertisements because the targeting ability enables firms to decrease costs associated with wasted advertisements.

Economics of Targeted Advertising

At the heart of targeted advertisements is the efforts by firms to better match their advertisements to the preferences of buyers. Some of the earliest papers to address the importance of this match are Nelson (1970, 1974), which expands the perspective of how advertisements can be informative by addressing the difference between search and experience goods. He argues that the quality of search goods can be ascertained prior to purchase (after costly search) whereas the quality of experience goods can only be evaluated after consuming the good. As a result, advertising can provide helpful information about experience goods to consumers before they purchase the products. The most relevant informative effect that Nelson (1974) highlights is the *match-products-to-buyers effect*. The *match-products-to-buyers effect* notes that individuals have different utilities for different products. As Nelson (1974) points out, "an esoteric, high-price soup gets advertised in the *New Yorker*, while Campbell's soup displays its wares in *Good Housekeeping*." The advertisement, then, sends a signal through the advertisement about what type of product might best match the customers' preferences.

Grossman and Shapiro (1984) build on this *match-products-to-buyers effect* by including markets of horizontal product differentiation in their equilibrium model of advertising. Abstracting away from search costs, Grossman and Shapiro (1984) assume that customers only learn about products through advertisements (assuming search costs are too high compared to consumer surplus from the products). The number of advertisements is chosen by the firm, but the delivery is randomly allocated to customers who are able to process the advertisements without cost. In Grossman and Shapiro (1984), heterogeneous consumers have Lancaster (1975) preferences uniformly distributed on a unit circle and seek to purchase products which are closest to their most preferred brand. Advertising improves the potential matching of consumers and products, but firms cannot target consumers. More advertisements benefit customers by giving them more options from which to choose the lowest price. While firms are not able to endogenously target the advertisements to a particular segment of the market, they identify the fact that heterogeneous customers will naturally gravitate to the product that best matches their preferences.

The ability to target advertisements offers firms two key benefits. First, firms are able to reduce advertisement costs by concentrating advertising to subsets of the population that are most likely to positively respond thereby reducing the overall cost of advertising. Second, firms are able to price discriminate to customers and charge higher prices to those who have a higher willingness to pay for their product. Bester and Petrakis (1996) are the first to account for this second benefit in their coupon informative advertising model. Customers are divided between two regions and know all the regular prices of goods in both regions. There are two levels of prices in each region, either the regular price known to all or the coupon discounted price known to those reached by the far firm with probability $\phi \in [0, 1]$. Firms seek to attract customers from the far region through coupon price discounts to induce them to incur the transport cost needed to purchase from them, but they will charge regular prices to customers in their local region. Their model is a simpler version of targeted advertising with price discrimination. Price targeting gives the local customers higher prices and the reach customers lower prices, enabling the firm to extract more surplus from loyal customers. While I also evaluate coupons in this paper, I find that the store sends discounts to its "local" customers through reward and promotion coupons. I find that one reason for this is the spillover effects of the promotion coupons and the increase likelihood of these customers coming to the store in the weeks of the campaigns. Esteban et al. (2001) is the first to enable firms to endogenously choose the level of advertisements in different segments of the market. Motivated by advertisements in niche magazines, Esteban et al. (2001) allows a monopolist to decide an optimal targeting strategy in a context where heterogeneous consumers have different reservation prices. Firms are able to target their advertisement to consumer groups with different valuations who have self-selected to reading different magazines by sending messages to magazines with varying degrees of focused readership. They show that a monopolist is likely to use specialized media to concentrate ads on individuals who are willing to pay more for the product, thereby increasing its market power by making demand less elastic and decreasing advertising costs by reducing wasted ads.

Iyer et al. (2005) expand the insights of Esteban et al. (2001) by introducing targeting under competition. In their model, firms either face loyal consumers or shoppers. Loyal consumers will buy from the firm as long as the price is below their reservation price. Remaining consumers are indifferent between the firms and will buy the product with the lowest price. They show that by targeting ads to loyal consumers who are willing to pay a higher price, firms increase their profits by eliminating "wasted" advertising and reducing demand elasticity by increasing differentiation in the market. Firms' profit gains remain whether or not firms can price discriminate.

Esteves and Resende (2016) allow targeted advertising to play two roles in a duopolistic market. Targeted advertisements provide information to consumers and are a mechanism for direct price discrimination. They compare two advertising strategies: mass advertising with no price discrimination and targeted advertising with price discrimination. Unlike Iyer et al. (2005), they find that if the goods are imperfect substitutes, then firms will focus more attention on their rival's market as long as advertising costs are low enough. They conclude that when targeted advertising serves as a mechanism for price discrimination, firms increase their profits at the expense of consumer surplus.

As technology improvements have enabled firms to more precisely target advertisements to segments of the market, firms have benefited from an increased ability to erode at competition by reducing advertisements to shoppers and reducing cost by focusing advertisements to those most likely to purchase. A consistent assumption in this strand of literature is that consumers still only learn about the existence of the products through the informative advertisement. There is no outside information available to customers outside the advertisements provided by the firm. This ignores low cost search options available to customers on the internet. Additionally, it ignores the cost of information congestion which may impede customers' ability to internalize the messages sent to them even when they are targeted. I will argue that in the context of information saturation, a new role of targeted advertising is to cut through the noise and better attract the attention of customers who have limited capacity to process all of the information they are receiving.

Economics of Competition for Attention

One of the defining aspects of the Information Age is that consumers are overloaded with information and have a scarcity of attention. Van Zandt (2004) and Anderson and de Palma (2009) both highlight how customers have a limited ability to process advertisements sent to them because of information overload. Anderson and de Palma (2009) argue that the externality of sending more messages can be compared to the externality of more cars entering an already busy highway. Attention economics was first identified by Herbert A. Simon, who famously stated, " ...in an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients." (Simon, 1971, p. 40).

In a novel approach, Anderson and de Palma (2009) incorporate aspects of search into their model of information congestion by accounting for the fact that looking through more messages is costly to the customer by incorporating an examination cost function. The likelihood of examining a message is decreasing in the number of firms sending messages, and they find that firms do not internalize the externality cost of over-sending messages. As a result, they argue for a tax on messages to help advertisers internalize this cost.

Anderson and de Palma (2012) adapt Butters (1977) to incorporate the consumers' lack of attention to advertisements. Whereas consumers in Butters' model only receive a subset of advertisements, consumers in Anderson and de Palma's model only process a subset of messages received, reflecting an overabundance of messages rather than a dearth of them. Consumers in their model register a fixed number of messages ($\phi < N$), where N is the total number of messages received per customer. They show how firms' competition for attention affects price competition for those firms in that higher attention spans lower prices. Their model cleverly adapts the classic Butters model to account for critical attention costs facing customers.

Anderson and de Palma (2009, 2012) address a new problem facing consumers by returning to the search dynamic at the heart of informative advertising models. While consumers in the Butters (1977) model faced search costs deriving from challenges to finding information, consumers in Anderson and de Palma (2009, 2012) face attention costs in processing the information sent to them. With an abundance of information, consumers must now figure out ways to process the beneficial information and filter out the useless information. While the authors do not explicitly model bounded rationality, their model addresses many of the issues faced by consumers in a bounded rational context. I will be utilizing the bounded rationality framework to model the consumers in my model, accounting for their lack of attention as they process advertising messages from the firm.

One area for expanded research mentioned by Anderson and de Palma (2012) is to endogenize the attention space by allowing consumers to equate the marginal benefit to the marginal cost of an additional advertisement. A way consumers could do this is to assess how relevant the ad is to them. This accounts for the next generation of targeted advertisements in which the level of targeting reflects the personalization to the consumers' tastes in order to *attract attention.* Therefore, higher quality messages offer greater marginal benefit to the customer, increasing the likelihood of them using attention to register the message.

3.2.4 Endogenizing Attention

Eliaz and Spiegler (2011) endogenize attention in their model of consideration set formation by developing a model of bounded rationality in which customers choose products within a limited consideration scope. They model consumers' primitives to choice with the $\succ$-ordering of their preferences and a consideration function, ϕ. The endogenization of attention occurs through the consideration function, ϕ, which is a function of the marketing strategy, M and the status quo product customers are randomly assigned, x^s.

By modeling the consideration function, Eliaz and Spiegler (2011) unpack important primitives of consumer behavior. In most economics models, consumers are considered to be fully rational, generally implying unbounded capacity to internalize and optimize on information given them. Even when customers are limited in their information, they are assumed to have perfect ability to draw Bayesian inferences based on correct understanding of the market. By providing a framework for analyzing consideration functions, Eliaz and Spiegler offer an alternative hypothesis of how consumers respond to information in the form of advertisements. They suggest that consideration of the marketing tool depends on the marketing tool itself (in the sense that more costly advertisements are more likely to be considered). Additionally, they assume that products are only evaluated if they are considered in the first stage. This framework limits the consideration set to the status quo product if a new product is not considered and to the status quo product and the new product if the new one is considered.

The novel approach has significant implications for analyzing customers' purchase path more broadly. I extend their model to account for customers' grocery purchase path. Customers have a status quo purchase path which takes them to various grocery stores and different departments within the grocery stores they visit. Marketing tools are used by the retailer to attract customers' attention. If a customer considers a product from a marketing tool, she will potentially change her purchase path in order to evaluate the product. Changing the purchase path opens the customer to localized marketing tools in the aisles which can further expand her consideration set. I also model the differential impact personalized advertisements have relative to generic advertisements, arguing that personalized advertisements are more likely to be considered. I am the first to evaluate the purchase path of customers within the context of limited consideration sets.

Evaluation of advertising in the context of consideration sets differs significantly from persuasive and information advertising. While persuasive advertising assumes that the advertisement imputes utility to the consumption of the product being advertised, the value of the advertisement in this context is in how effective it is at causing a customer to consider the product being advertised, independent of the evaluation of the product. Additionally, the advertisements in the consideration set framework can have an element of informativeness, but this the context of the information transmission is different. Informative advertising models assume that advertisements are informative if a customer has not previously received a message for a product whereas in this framework, thereby expanding consideration to the advertised product. Whereas within the context of my model, an advertisement does not need to be for a new product in order to expand the consideration set; the customer merely needs to not have the product in her status quo shopping list. Consideration of the product is also dependent on the personalization or relevance of the advertisement to the recipient.

The work by Eliaz and Spiegler (2011) is part of a small but growing literature on bounded rationality. As early as the mid-1950's, Herbert A. Simon began to question some of the basic premises of rationality foundational to the way economists typically analyze consumer decision making. One of the central arguments Simon made in his early works (1955,1956) is that consumers do not know all of their alternatives. This concern has only grown as consumers are increasingly faced with more options than they can process. An overabundance of information arguably impedes consumers ability to effectively optimize over all possible choices.

Rubinstein (1998) delved deeper into choice theory by modeling knowledge, limited memory, and choosing what to know. By modeling knowledge formation, Rubinstein

(1998) provides a framework for analyzing how agents consider information. Spiegler (2011) applies these bounded rationality models to the context of market interactions in industrial organization.

A number of experimental psychologists have done extensive work in showing that decision makers consistently deviate from the rational choice model (see Hogarth and Reder, 1987). While significant strides have been made toward questioning and empirically testing premises used to model rational choice, there is certainly not a consistent framework. It is the aim of this paper to propose a context in which limited consideration sets reasonably model consumer behavior in the context of consumers' grocery purchase path.

3.2.5 Marketing Literature on Targeted Advertising

Targeted advertisement research has primarily focused on estimating direct response (Arora et al., 2008). Multiple studies have shown that advertisements matched to customer preferences are more effective in engaging response and increasing sales of the advertised product. Targeted advertisements not only increase effectiveness, they also decrease cost by reducing wasted advertisements (Iyer et al. 2005). As shopping moves online, stores are increasingly tailoring their communication with customers. Many studies show that personalization of emails and websites each increase customer engagement (Ansari, A. and C. Mela, 2003 and Hauser, J. R., G. L. Urban, G. Liberali, and M. Braun, 2009).

Increased targeting is not always correlated with increased response. Lambrecht and Tucker (2013) show that dynamic retargeted advertisements are generally less effective than generic retargeted advertisements unless customers have spent more time developing their preferences. Tucker (2011) also finds that while matching ad content to the website and personalizing advertisement content independently increases effectiveness, combining these strategies decrease engagement.

While it is important to understand the direct impact of targeted advertisements, my research shows that a related metric tells a more complete story. The spillover effect of targeted advertisements can drive more sales than the direct effect of the targeted advertisement, particularly for retailers with multiple products. Marketing spillover effects have traditionally been analyzed primarily in the context of brand alliances. It is well documented that when brands have an alliance, customers evaluations of brand A will spillover to their evaluations of brand B (Balachander and Ghose 2003; Baumgarth 2004; Desai and Keller 2002; Janiszewski and Van Osselaer 2000; Park, Jun, and Shocker 1996; Samu, Krishnan, and Smith 1999; Simonin and Ruth 1998; Vaidyanathan and Aggarwal 2000). Unfortunately, the direction of this impact has not been consistent. Some studies have found that the spillover effect is positive (Simonin and Ruth 1998; Washburn, Till, and Priluch 2000), while others found it was negative (Keller and Aaker 1992; Loken and John 1993; Till and Shimp 1998). Tseng (2010) and Raghubir (2004) both found a negative spillover affect associated with gift promotions. Customers tend to discount the gifts' value, which leads them to discount the value of the product category with the gift. The mechanism of spillover effects is assessed in Erdem and Sun (2002), who find that advertisements reduce the uncertainty of related brands.

As marketers increase their attention to and assessment of multi-channel advertisements, researchers have begun to analyze the spillover effects across channels. Rutz, O. J. and R. E. Bucklin (2011) find that asymmetric spillover effects with generic and branded search activity. While generic search activity positively affects branded search activity via increased awareness, branded search does not affect generic search. Joo et al. (2014) examine the impact television advertisements have on online search behavior. They find that TV ads lead to increased related online search and increased branded searches.

Dias et al. (2008) and Venkatesan and Farris (2012) separately identify spillover effects of targeted advertisements in grocery stores. Dias et al. (2008) finds that when an online grocery store recommends products based on customers' purchase history, sales of recommended items and their categories increase. Additionally, Venkatesan and Farris (2012) use the same data as this paper to identify what they term an exposure effect to customized coupons. They distinguish this indirect spillover effect on propensity to come and sales with the direct redemption effect. I extend the literature of the spillover effects of targeted advertisements in grocery stores by identifying increasing spillover returns to promotions and decreasing spillover returns to rewards. I also show that the primary channel through which the in-store spillover is realized is in department of the promoted products. Finally, I use machine learning methods to control for targeting endogeneity to ensure estimated effects are not biased.

While generally understood to be a price discrimination tool to attract customers by lowering price (Narasimhan,1984; Bester and Petrakis, 1996; Anderson, Baik, and Larson, 2015), traditional coupons have also been found to have spillover effects by increasing sales while customers are in the store and by inducing repurchase of the advertised product. Heilman, Nakamoto, and Rao (2002) show that in-store surprise coupons increase sales by increasing unplanned purchases. Nevo and Wolfram (2002) also find evidence that coupons can induce repurchase of products. This also relates to an extensive body of research on in-store marketing (SK Hui, Y Huang, J Suher, JJ Inman, 2005; Breugelmans and Campo, 2011; Inman et al., 2009; Hui et al., 2013; and Chandon et al., 2009).

This paper also contributes to the growing literature of the impact targeted advertisements have on consumer preference development. Simonson (2005) finds that a consumer's stage of preference development may significantly affect the effectiveness of ad content. In particular, advertisements that convey high-level characteristics are more effective when customer have a broad idea of what they want while advertisements that focus on specific products are more effective when consumers have narrowly construed preferences. Lambrecht and Tucker (2013) use this framework to test how targeting response varies depending on the customer's state. Fong (2012) examines this connection by looking at how targeting affects the state of the customer. He finds that consumer search decreases when consumers receive targeted offers. This paper contributes by identifying the differential impact rewards have on inducing spillover sales relative to promotions. Promotions increase awareness of relevant products and increase the consideration set of the consumer in the department while rewards do not exhibit this effect.

3.3 Industry Description

3.3.1 Grocery Industry

The grocery industry in the United States is large with sales totaling $649.1 billion in 2015 according to the leading trade journal, *Progressive Grocer*. The figure accounts for the sales of the 38,015 stores which annual sales of $2 million or more and is a 1.7 percent increase over the previous year. Profit margins are slim for grocery retailers at 1.7 percent after tax in 2015 (FMI). Americans spend 5.5 percent of their disposable income on food at home with an average of 1.6 trips made to the grocery store each week and $31.92 sales per customer transaction. Traditionally, shopping trips averaging $15 a basket are considered immediate-need driven, $51 a basket are considered fill-in, $98 a basket are considered weekly, and those at $242 are considered stock-up (Nielsen, 2011).

Since the turn of the 20th century, traditional grocery stores have been the primary destination for household grocery shopping needs (see Turow's *The Aisles Have Eyes* for a thorough description of the evolution from merchants to retailers). Offering one-stop shopping, the traditional supermarket consolidated grocery shopping for produce, meat, dairy, and consumer packaged goods under one roof. Retailers competed for the best location with the best assortment of products to serve the greatest number in their weekly shopping needs. This market terrain is quickly shifting as customers are diversifying their tastes and have more shopping options across channels. We are observing the reverse trends of a century ago when the supermarkets overtook the local butcher, produce grocer, and home delivery of milk.

According to the *Food Marketing Institute* report, "U.S. Grocery Shopping Trends, 2016", the grocery industry is becoming more fragmented by channels, making it harder for stores to attract and keep customers. The evolving needs of customers are leading them to seek a broader set of less traditional channels and not claim one store as their primary grocery retailer as seen in Table 3.1 ("Surviving the Brave New World of Food Retailing, 2017"). New niche entrants into the market are undermining the dominant position held by traditional supermarkets, responding to changing customer tastes.

Since 2005, the number of households identifying the traditional supermarket as their primary store has decreased from 67 percent to 49 percent whereas those claiming no primary store have increased from negligible to 7 percent in the same time period. Table 3.2 shows the evolution of the percent of households identifying various channels as their primary store over the past decade. As Table 3.3 from "Surviving the Brave New World of Food Retailing, 2017" shows, non-perishable purchases (or center aisle products) are particularly fragmented across channels, with a sizable fraction of respondents (28 percent) doing some of their non-perishable shopping online instead of in a brick-and-mortar store. While most respondents (79 percent) do at least some of their perishable good shopping in traditional supermarkets, a surprising 21 percent of households do not shop for perishables in these stores. In response to the changing dynamics in the grocery retail industry, a significant percentage of retailer executives view digital media and digital marketing as important in their marketing and advertising (see Table 3.4, Progressive Grocer's 2016 Annual Report of the Grocery Industry).

Table 3.1: Number of Supermarkets Shopped

Number of Supermarkets	Percent of Respondents
6+	11%
5	9%
4	17%
3	25%
2	23%
More than 1	**86%**
1	14%

Table 3.2: Primary Grocery Channels

Grocery Channel	2005	2010	2016
Traditional Supermarket	67%	56%	49%
Supercenter	22%	27%	25%
Warehouse	7%	6%	5%
Discount	2%	2%	2%
Limited Assortment	1%	7%	7%
Organic Specialty	1%	2%	3%
No Primary Store	-	-	7%

In order to fend off the increase competition from smaller specialty retailers (like Trader Joe's and Aldi Group), traditional supermarkets are becoming increasingly consolidated with larger companies acquiring local and regional chains. For example, global retail giants, Ahold and Delhaize Group merged in July 2016 and Kroger acquired Roundy's brands in December 2015 (Euromonitor, "Grocery Retailers in the US"). Wal-Mart continues to lead grocery retailers with 26 percent of the market share, reflecting its dominance in the Supercenter channel (at 83 percent). However, this channel is saturated with little growth in urban areas. One of the biggest growth retailers outside of acquisitions was Sprouts Farmers Market, a local-focused retailer focused on locally-grown, fresh perishables. Kroger is the biggest player in the tra-

Table 3.3: Channels Used for Different Grocery Needs:
Percentage of Respondents Shopping in Each Channel

Grocery Channel	Perishables	Non-Perishables
Traditional Supermarket	79%	56%
Supercenter	54%	65%
Warehouse	38%	37%
Discount	13%	38%
Convenience Store	16%	16%
Organic Specialty	33%	13%
Drug Store	12%	41%
Online	4%	28%

Table 3.4: Retailer Marketing/Advertising:
Percentage Rating Each as Extremely or Very Important

Marketing Tool	Percent of Respondents
In-Store Signage/Digital Media	62.7%
Digital Marketing	42.6%
Newspaper Inserts	42.4%
Mobile Marketing	35.8%
Direct Mail (Circulars)	32.4%
Newspaper Ads	29.9%
TV Advertising	20.9%
Radio Advertising	14.7%
Custom Magazines	9.1%

ditional supermarket channel at 25.3 percent. Aldi is the leader in Discount stores with 60 percent share and 7-Eleven is the leader in convenience stores at 30 percent share. Table 3.5 summarizes the market shares for the top grocery retailers.

Table 3.5: Market Share for Grocery Retailers

Grocery Retailer	**2012**	**2016**
Wal-Mart Stores Inc	25.5%	26.3%
Kroger Co	7.6%	10.2%
Albertson's Inc	0.4%	5.4%
Ahold Delhaize	-	4.0%
Publix Super Markets Inc	2.9%	3.3%
HE Butt Grocery Co	1.9%	2.4%
Meijer Inc	1.5%	1.8%
Whole Foods Market Inc	1.2%	1.6%
Target Corp	1.5%	1.4%
Trader Joe's Co	1.0%	1.3%
Seven & I Holdings Co Ltd	1.1%	1.1%
Giant Eagle Inc	0.9%	1.1%
Bi-Lo Inc	1.0%	1.0%
Aldi Group	0.8%	1.0%
Hy-Vee Inc	0.7%	1.0%

Whereas in the past traditional grocery retailers needed to focus on getting customers into the store to do their primary grocery shopping, now grocery retailers need to more deeply understand customers' individual needs and how they shop both within the store, but also across different channels. Customers today have substantially more options of how they can shop for their food needs. Customers can order items and have them delivered at their home, they can have them delivered at the store for pick-up, and they can also order ready-to-cook gourmet meals through Blue Apron and Plated. A deeper understanding of each customer's overall purchase paths will enable retailers to compete in this quickly changing market.

3.3.2 Targeting in Grocery Industry

In order to better compete for consumer attention and wallet share in a field of increasingly varied options, some retailers are beginning to re-evaluate how they can reposition themselves in customers' purchase path. According to a call to action in the report, "Surviving the Brave New World of Food Retailing, 2017," retailers can stay relevant only by understanding the consumer's journey well before they place an item in the cart: "It includes the lifestyle triggers, preferences and priorities that precede and influence the eventual shopping list; the realities of work routines and home logistics; and the changing social context of how meals are prepared, shared, and enjoyed." Getting to that level of understanding customers requires substantial tracking of customers' life patterns. While maintaining traditional advertising as outlined in Table 3.4, grocery retailers are quickly moving toward more tracking in order to better position themselves.

While grocers have had access to rich purchase data since adopting the scanner beginning in the 1980s, grocers have been reluctant to use the information to analyze customer-level data. By the mid-1990s, Catalina Marketing began collecting purchase information and offered discount coupons at the checkout of retailers. Customers would receive a printed coupon at the register for a product they didn't try, based on the purchases made in the recent weeks leading up to the visit. While retailers saw value to collecting data on customers, the vast volume was overwhelming for many retailers. Additionally, in response to pressures from Wal-Mart, retailers were focused more on increasing the efficiency of operations. Grocers mainly focused on providing in-store frequent-shopper rewards and were slow to develop systems to link the information to purchase behavior.

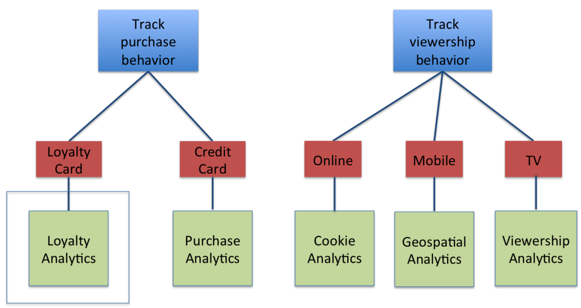
As ability to analyze and store data increased, Catalina and other similar firms have incorporated longitudinal data to observe patterns in purchase history. An industry leader responsible for transforming marketers' approach to customer analytics, Dunnhumby began using consumer data to understand deeper consumer insights from longitudinal data. Instead of relying on the most recent purchases, Dunnhumby analyzed patterns over the past year. Richer models enabled the company to send coupons relevant to the specific tastes of each household. Dunnhumby pioneered customer first marketing that prioritizes the revealed preferences of customers in order to foster organic growth. The *Scan It* system introduced by Ahold USA's Stop & Shop stores allows customers to link a handheld device to their loyalty card and scan items while traveling through the store. It offers quicker check outs and real time personalized offers while also tracking customers' purchase paths at the store.

Customized coupons allow companies to tailor messages to particular consumers. For example, someone who regularly purchases health food products may receive coupon offers on items like soymilk and granola bars rather than potato chips. Consumers benefit by saving money without searching through pages of coupons. This encourages redemption rates. Between 2005 and 2009, Kroger redemption rates increased from 2 percent to 24 percent. By articulating the proposed value of the goods to specific consumers, the grocery stores with customized coupons decrease the transaction cost for the consumer.

With the rapid decline in costs for analyzing data and the changing competitive landscape, brick-and-mortar retailers are increasingly seeing their need to collect and identify patterns in customer shopping behavior beyond the point of sale. With the surge in use of mobile devices, retailers are able to move beyond the collection of scanner data to tracking customer movements within the store. According the Turow's *The Aisles Have Eyes*, "[N]early all [retail executives] agreed that the basic imperatives of identifying and following shoppers to and through the physical store, collecting enormous amounts of information about them without their awareness, and personalizing messages to them along the way is, whether through one technology or another, a new requirement of retailing" (2017).

This movement toward tracking is in conjunction with broader trends toward tracking. With the acceleration of the online and mobile platforms and the tracking tools associated with it, personalized advertisements have gained a lot of attention. Particular focus has been placed on tracking user behavior in order to more appropriately place advertisements to households with relevant tastes for a brand. Figure 3.1 summarizes important tracking tools used by marketers personalizing advertisements.

Figure 3.1: Targeting Overview



Over the past five years, retailers have quickly increased the degree to which they track customers. While in 2013, the ten largest retailers agreed which engaged in foot traffic tracking agreed to a 'Mobile Tracking Code of Conduct' that required stores to post conspicuous signs when using tracking technology, today, store are much more conspicuous and invasive in their tracking. Retailers now track customers on their smartphones via Wi-Fi, Bluetooth, and cell tower connections either with their own retailer apps or through third party tracking companies like Shopkick, inMarket, and xAd which track location within the store through Wi-Fi or Bluetooth and outside the store with cell-tower pings. Often customers are tracked with their information bought and sold without their awareness. Each piece is helping retailers and marketers build a more comprehensive understanding of customers' journey.

The new generation of tracking is widely celebrated by analytical firms which predict that soon retailers will be able to anticipate the needs of customers. As a 2013 Gartner information technology report claimed, "Surveilling people across their digital and physical worlds would lead to such in-depth profiling that businesses would create intimate models of an individual's behaviors and dispositions regarding their consumption patterns and references, and would ultimately enable firms to anticipate customers' needs and actions." (*The Aisles Have Eyes*, pg. 150). The new goal for retailers is to understand customers through tracking in order for them to better address customer needs so that they can compete against an increasingly diverse set of competitors.

3.4 Theoretical Model

The foundational model of consideration set formation within the context of competitive marketing strategies was first developed by Eliaz and Spiegler (2011). They expand the limited attention model set up by Anderson and de Palma (2012) by making the consideration of a product, ϕ, conditional on the marketing strategy, M.

I expand the Eliaz and Spiegler (2011) model in two ways. First, I account for the *impact* of the consideration of a product on the purchase path of the customer. When a customer considers a product in a grocery setting, in order to examine the product, she changes her purchase path if the product is off their current path. Changing her purchase path can expand her consideration set of products to include the products along that new path.

Second, I identify the personalization mechanism through which marketing tools can attract consideration. Personalization is a form of highly targeted advertising. Targeted advertising refers to the proximity of the set of advertised products to the preferences of the household receiving the advertisement (in the spirit of the *match-*

products-to-buyers effect). While generic advertisements (or mass advertisements) are based on the preferences of the average customer, personalized advertisements are relevant to the customer in that they reflect the preferences of the individual receiving the advertisement. In my extension of the model, the personalized advertisements are more likely to be considered than the generic advertisements.

3.4.1 Basic Model (Eliaz and Spiegler, 2011)

Below I outline the basic framework developed by Eliaz and Spiegler (2011). Each firm produces one product (exclusive to the firm), x, from the set of products, $\mathcal{X}$. Firms choose a marketing strategy, M, for each product, x from the set of marketing strategies, $\mathcal{M}$. Each marketing strategy is a vector of marketing tools, m. Following them, I denote product-marketing pairs to be (x, M). Homogeneous customers each are randomly assigned a status quo product, x^s and decide whether they want to consider a new product, x^n. Consideration is a function of the marketing strategy the customer receives related to the product, M^n.

In the first stage, customers construct a consideration set which can take on two elements $\{x^s, x^n\}$ if the customer considers the option meaning $\phi(x^s, M^n) = 1$ or one value $\{x^s\}$ if $\phi(x^s, M^n) = 0$.[1] In the second stage, the customer chooses the product which maximizes her $\succ$-ordered preferences. They interpret the linear order $\succ$ as the customer's "true" preferences over $\mathcal{X}$. Preferences are linear in the sense that they are monotonically increasing up to the $\succ$-maximal option. These tastes are stable and not affected by marketing. Therefore, if a customer always considered all of her options, then her revealed preferences would be rationalized by $\succ$. The consideration function ϕ reflects a customer's willingness or ability to consider alternative purchase paths. A customer with consideration function, ϕ', may be more "open" to considering

[1]The framework precludes customers from going out a searching for products themselves. Therefore customers' consideration of new products is based solely on the advertisements given by firms.

alternative purchase paths than another customer with consideration function, ϕ if $\phi(x^s, M^n) = 1$ implies that $\phi'(x^s, M^n) = 1$. In general, more costly marketing tools, m are more effective.

The consideration function can violate transitivity in the sense that a consumer may have preferences over products given by $x'' \succ x' \succ x$ but her consideration functions may yield, $\phi(x, M') = 1$, $\phi(x', M'') = 1$ and $\phi(x, M'') = 0$. This would imply that she may choose $(x', M') \succ (x, M)$ and $(x'', M'') \succ (x', M')$, but she may not choose $(x'', M'') \not\succ (x, M)$ because the advertisement, M'' is not effective against the product, x. However, marketing cannot reverse consumers' revealed preferences over products. Therefore, if one marketing tool, M' associated with product x' is effective against product x, then we cannot observe product x beating x' with another marketing strategy, M''. The takeaway is that marketing can affect the perception of a customer's feasible set, but it cannot alter her preferences over her consideration set. This model has a status quo bias in the sense that customers will choose the status quo (x^s, M^s) over a new option, (x^n, M^n) either when $x^s \succ x^n$ or when $\phi(x^s, M^n) = 0$ even if $x^n \succ x^s$.[2]

The key takeaway from Eliaz and Spiegler (2011) are that customers use the consideration function ϕ to decide whether to evaluate new products. The consideration function extends the limited attention of Anderson and de Palma (2012), making attention contingent on the advertising message received. The authors do not specify what causes a customer to consider an advertisement apart from assuming that more costly advertisements are more likely to be considered than less costly ones.

[2]This model provides a lot of insight into the primitives of consumer choice, but the specificity at this level abstracts from other important features of consumer choice including prices. Incorporating prices into this choice framework is an area for future research.

3.4.2 New Model with Extensions

I generalize Eliaz and Spiegler's model in two dimensions to fit the grocery retail setting. Expanding the model better enables me to reflect the complex consideration set and decision making structures in the grocery setting. This is useful for analysis of purchase behavior more generally since households often face a complex decision tree when making purchase decisions on their consumer journey.

Consideration Set Expansion through Purchase Paths

First, I account for the impact that the marketing strategy, M, has on expanding a customer's consideration set within the store.[3] The grocery retailer has a wide range of products they would like the customer to purchase. However, customers are generally focused on a limited subset of all the products available, following a status quo purchase path for their basic products, P^s. In order for a store to change customers' overall purchase behavior, they need to influence their purchase path. This will expand the set of products a customer considers. When a customer decides to evaluate a product, x, in an advertisement, M, she will adjust her purchase path to P^n in order to examine the product (when the product is not on her current path).

I define a set of purchase paths, $\mathcal{P}$, which define the geospatial routes at three levels. The first level is the discrete choice of whether to visit a store, 0 or 1. The second level is the set of departments the customer walks through within the store, conditional on coming to the store. And the third level is the set of products the customer purchases within each department through which she walks. Various marketing tools affect different levels of the purchase path. For example, some may be more

[3]Marketers have a deep literature on consideration sets (see Shocker et al., 1991 and Andrews and Srinivasan, 1995). Conceptually, the concept is similar in that consideration sets are defined to be an intermediary between awareness sets and choice sets. In practice, this is generally applied to a setting of category and brand choice using nested logit-type analysis to evaluate the options available to the customer, where consideration and evaluation are based on known utilities about the product.

focused on bringing customers to the store, others may be more focused on getting customers to consider products once they are in the department, and others transcend all three levels. Consideration of the products in targeted and generic mailers sent to the customer's home is assumed to occur when the customer evaluates the advertisements at home. Therefore, this advertisement can affect her choice of coming to the store. Since evaluation of the product requires the customer to visit the product at the store, consideration of the product can also affect her overall sales at the store and in the department of the advertised product by causing her to potentially change her path in the store and be exposed to more marketing tools. In-store displays cannot influence a customer's choice of coming to the store, but they can influence a customer's path within the store. Localized marketing tools like retail discounts cannot influence the store visit or path within the store, but they can influence the customer's consideration set once she is in the department.

Customers have an initial status quo purchase path, P^s, with a propensity to visit the store and a customer-specific path within the store. Each status quo purchase path is associated with a status quo set of products, $X_{P^s} \in \mathcal{X}$. This can be thought of as a shopping list. A single retailer, j may not have all of the products on a customer's status quo shopping list such that $X_{P^s} \not\subset \cup_{D_j} X_{d_j}$ where $\cup_{D_j} X_{d_j}$ is the union of all the separate department products, X_{d_j} for the full set of departments, D_j within store j. In that case, a customer may visit multiple stores to complete her shopping list. Define the intersection of the shopping list and the products in store j, $X_{P^s} \cap \cup_{D_j} X_{d_j}$ to be set of products store j sells which are on the customer's shopping list. Assume that the customer orders stores based on which store has the most products on the list. Call the store with the most products on her list her preferred store. Remaining items on the list are taken care of at the next best store and so on. Therefore, the preferred store receives the most business from the customer.

Once a customer is in store j, she determines the set of departments to walk through based on her shopping list for that store such that $X_{P^s} \cap \cup_{D_j} X_{d_j} \in \cup_{A_j} X_{d_j}$ where $A_j \in D_j$ is a subset of departments within the store. Ordering of departments is primarily based on minimizing the steps traveled in the store. Once a customer is in a department, $d_j \in A_j$, she collects the products on her list that are in the department, $X_{P^s} \cap X_{d_j}$.

At each point along the status quo purchase path, a vector of marketing tools, M can be used to attempt to expand the customer's consideration set. For example, a marketing tool, m may highlight a product x which is not in the customer's status quo shopping list. If the marketing tool is effective and she considers the product, then she must go to the physical location of the product in order to finally decide whether to purchase the product. Therefore, while the product is not on her shopping list, it is added to a set of products she is considering, X_n which is then added to her consideration set, $\{X_{P^s} \cup X_n\}$. The new set of products can influence her purchase path in a number of ways.

A marketing tool sent to a customer's home may affect a customer at three levels. In the simplest form, if the product is in her preferred store and in one of the departments along her path, consideration of the product does little to change her purchase path. Next, consider a case where the product is in her preferred store but in a department outside her list. Consideration of the product then causes her to add a department visit at the store so that her new set of departments at the store equals B_j, where $A_j \in B_j$. Finally, consider a case where a second best retailer sends an advertisement. The consideration of a new product from this store would cause the customer to prioritize this store over the other and could shift the order of stores, thereby increasing her purchase path at the new preferred retailer because it no longer just serves to sell the remaining products to the customer.

Once customers are in a department at a store, the retailer uses marketing tools such as retailer discounts to attract the attention of customers. These tools cannot serve to shift the purchase path of customers, but can attract attention from customers as they are evaluating the new products in their consideration set. Evaluation of the products in her consideration set for the department, d_j, $X_{n_{d_j}} \cap X_{d_j}$ requires the customer to compare with surrounding products. For example, if a customer is evaluating yogurt in the Dairy department, she will evaluate the other yogurt manufacturers and choose one according to her $\succ$-preferences. In the process of examination, the customer's awareness of other surrounding products increases, especially when those products are associated with other marketing tools. Therefore, evaluation of products can further increase the customer's consideration set.

The customer then visits the next best store (according to how many products on her remaining shopping list can be purchased there) and repeats the process until she has completed her shopping list. The customer always purchases the products on her shopping list. Therefore, the marketing tools always have the effect of potentially expanding a customer's purchase behavior but never detracting from it. This is distinct from the model set up by Eliaz and Spiegler (2011) in that newly offered products compete with the status quo product and the customer always chooses one between the two. It also differs in that I analyze the impact of consideration of a product on the entire purchase path of the customer.

Note that path to purchase models are well established in traditional marketing models (Shankar, 2011). Key stages in this process include awareness, search, evaluation, store visit, and product choice (Baik, Venkatesan, Farris, 2014). This model draws insights from this framework, but differs in its application. Specifically, in this framework, the limited consideration set focuses analysis of products to the status quo shopping list and products considered through marketing tools. The focus of my model is on how consideration of a product outside of the status quo path can influence the customer's consideration set and basket size. This is distinct from the traditional customer path to purchase framework.

Personalized Advertising

The second way I extend Eliaz and Spiegler (2011) is that I allow marketing tools to vary by their personalization to the customer. Personalization is defined to be the proximity of the set of advertised products to the preferences of the household receiving the advertisement (in the spirit of the *match-products-to-buyers effect*). Therefore, customers in this framework are by nature heterogeneous.

I assume that marketing tools which are personalized, m_p are more likely to be considered by households. The motivation behind this assumption is that customers are overwhelmed by advertisements and information given to them by the full marketing strategies of all firms, $\cup_j M_j$. Therefore, they do not have the capacity nor desire to consider every advertisement they receive. Advertisements that are more closely aligned to their preferences are more likely to be noticed and considered than advertisements intended for a broader audience.

While search costs are not explicitly modeled here, the personalization of advertisements addresses the nature of search costs in that it acknowledges that it takes time and energy to find new options and consider those that are relevant. There are too many products available in the store or even within each department for a customer to effectively search through and evaluate each available product. By personalizing advertisements, the retailer does this work for the customer and facilitates her search process.

For each marketing strategy vector, M, personalized marketing tools, m_p are more likely to be considered than generic marketing tools, m_g. Therefore, we propose that for each marketing tool, $E[\phi(P^s, m_p)] > E[\phi(P^s, m_g)]$.

This definition differs from the assumptions in Eliaz and Spiegler (2011) in that personalization is not necessarily correlated with the cost of the advertisement. Additionally, while Eliaz and Spiegler (2011) describe ways in which their model can be applied to targeted advertising, there are some important differences. First, because customers in their model are homogenous, targeting is based on the status quo product each customer is randomly assigned which has nothing to do with their underlying preferences. Because marketing tools are in relation to the status quo product, their effectiveness in influencing consideration relies on the quality of the initial product relative to the quality of the marketing tool.

In my framework, households are heterogeneous with distinct preferences for different products. Since the marketing tool is not meant to compete with the initial set of products in their shopping list, targeting is in relation to the household's underlying preferences and not their initial shopping list (though the shopping list likely reflects the household's preferences). This subtle difference is important when assessing the role of personalization in my model versus targeting in their model. The more closely aligned an advertisement is to the underlying preferences of the household, the more likely the advertisement is to be considered, regardless of the contents of the initial shopping list or even the cost of the advertisement.

In this sense, personalization in my model more closely aligns with informative targeted advertising models. However, the difference from these models is that consideration of advertisements in this model is a primitive to the evaluation of the product being advertised. Since personalized advertisements are more likely to be considered, this may yield a similar effect as advertisements sent to loyal customers. However, we do not observe the same competitive effects for contested customers because advertisements that are not closely aligned with a customer's preferences are less likely to be considered.

3.4.3 Advertisements and Consumer Purchase Path Overview

To outline the steps of the model, consider

- **Status Quo Path**: Each customer begins with a status quo purchase path, P^s which is the result of the determination of which store offer the most products on her status quo shopping list and which departments in the store carry the products for which she is looking.
  - The customer first visits her preferred store which has the greatest number or products in her list, $X_{P^s} \cap \cup_{D_j} X_{d_j}$.
  - Then the customer chooses her departments within the store which carry the products she plans to buy at the store, $X_{P^s} \cap \cup_{D_j} X_{d_j} \in \cup_{A_j} X_{d_j}$ where $A_j \in D_j$.
  - Then the customer chooses products in the department, $X_{P^s} \cap X_{d_j}$.
- **Advertisement**: Firms send advertisements, M^n for products, x^n to customers.
- **Consideration Function**: The customer uses her consideration function, ϕ to determine if she will consider the product in the marketing tool. If $\phi(P^s, M^n) = 1$, then she will evaluate the product by going to its location in the store.
  - **Personalization**: Personalized messages are messages more closely aligned with the customer's underlying preferences and are more likely to be considered than generic messages, $E[\phi(P^s, m_p)] > E[\phi(P^s, m_g)]$
- **Consideration and Purchase Path**: Consideration of the product on the marketing tool can affect the customer's purchase path in different ways.
  - *No Consideration*: If a customer does not consider the product, x^n (i.e., $\phi = 0$), then there is no change to the customer's purchase path, P^s.
  - *Consideration*: If a customer does consider the product ($\phi = 1$), then there is potentially a change to the customer's purchase path. Change

to the purchase path depends on whether the product is currently on the shopping list.

* *Consideration, No Path Change*: If a product is already in the shopping list, there is no change to the consideration set or the purchase path.
* *Consideration, Path Change*: If the product is not in the shopping list, then the purchase path changes. Change depends on how far the product is from the status quo purchase path.
    - **Same store, same department**: If the product is along the purchase path, the only impact to the consideration set is in whether to purchase the product when in the department.
    - **Same store, new department**: If the product brings the customer to a new department, the impact of the advertisement is to expand the consideration set to the advertised products and the set of products potentially added to the consideration set while the customer is in the department (for example, through retail discounts for products to which the customer would not otherwise be exposed).
    - **Different store, same department**: If the product brings the customer to a new store with the same department, it could induce her to do her primary shopping at this store rather than her other preferred store.
    - **Different store, new department**: If the product brings the customer to a new store with a new department, it could expand her shopping as in **Same store, new department**, but in the new store.

- **Impact on Consideration Set and Overall Purchases**:
  - *No Path Change*: If the customer's path does not change, we expect little impact on the overall purchases.
  - *Path Change*: If the customer considers a product x^n in an advertisement, we expect to see an expansion of overall purchases relative to the magnitude of the change in path and consideration set. For example, we expect exposure to a new department within the store would increase sales within the department because it also exposes the customer to new marketing tools within the department which can further expand her consideration set.
- After shopping at her preferred store, the customer completes the rest of her shopping list at the next best store and so on until her shopping list is complete.
- In the beginning of the new period, the customer returns to the status quo purchasing path, P^s.

Store Choice

In order to understand the purchase path, p in detail, consider first the choice of whether to visit the store. Each week, each household i decides whether to visit a store j. Choice of whether to visit a store depends on a set of time-invariant household-store match characteristics, F_{ij} and household characteristics, w_i, which are represented within M^s (since M^s is the cumulative learned value of the current status quo path for a recurring customer) and a set of marketing strategies employed by the store, M^n. For marketing strategies meant to increase the likelihood of bringing households to the store, focus will be made on addressing how the store satisfies household-store match characteristics. I will outline the store-level and household-level marketing strategies employed by the retailers in Section 3.4.4.

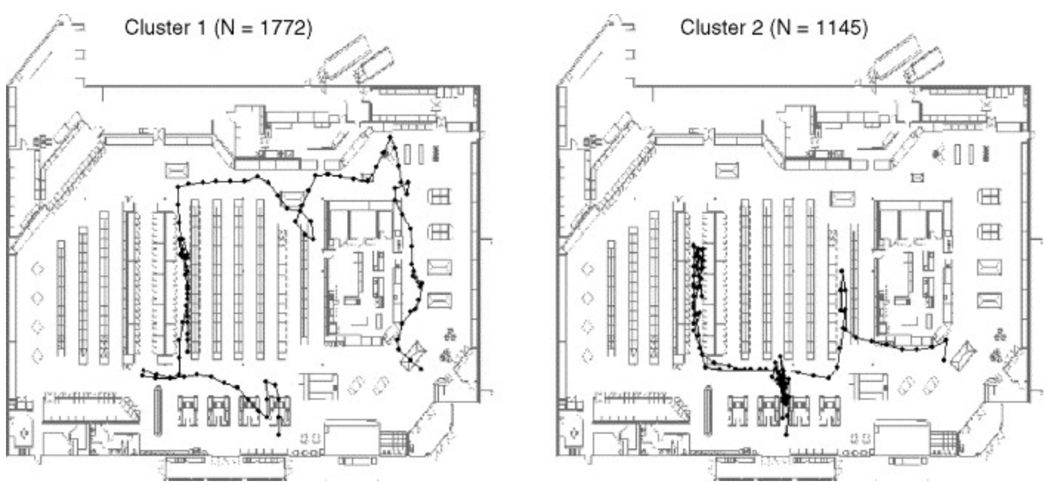
Let household-store match characteristics, F_{ij} be a function of a number of characteristics. First, the distance the household must travel to the store, $dist_{ij}$ has been shown to influence store choice (Huff, 1966; Achabal, Gorr, and Mahajan, 1982; Donthu and Rust, 1989; Ghosh and Craig, 1983). Second, availability of products relevant to the customer, $variety_{ij}$, is also an important variable affecting store choice (Kumar and Leone, 1988; Messinger and Narasimhan, 1997; Kahn and Wansink, 2004; Jacoby and Mazursky, 1984). Thirdly, store attractiveness (including customer service, plenty of parking, lighting, number of employees, number of checkouts), $attract_{ij}$, can induce some customers to drive to a store further away (Mehrabian and Russell, 1974; Baker, Parasuraman, Grewal, and Voss, 2002). Finally, average prices of products relevant to the customer, $price_{ij}$, is a factor in store choice (Bell and Lattin, 1997; Mulhern and Leone, 1990; Ho, Tang, and Bell, 1997). The combination of these time-invariant characteristics suggests that the household-store match characteristics can be described by the function, $F_{ij} = F(dist_{ij}, variety_{ij}, attract_{ij}, price_{ij})$.

Next, let household characteristics, w_i, be a function of demographic characteristics that might influence shopping behavior. First, household income, $income_i$, can affect a household's willingness to spend more at the store (Kalyanam and Putler, 1997; Sampson and Tigert, 1992; Hoch, et al., 1995). Second, whether the individual is married or single, $married_i$ can also determine the type of food purchased (Zeithaml, 1985). Third, the number in household, num_i, certainly has an impact on expected basket size (Arnold, 1997). Finally, age of the household, age_i, can also affect the number of times a household visits the store (Crask and Reynolds, 1978). The combination of these time-invariant household characteristics can be described such that $w_i = w(income_i, married_i, num_i, age_i)$. Note that these characteristics will likely describe the shopping patterns of overall grocery shopping but they may not be store-specific.

Purchase Path Choice and Product Choice

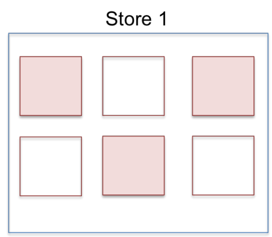
Once the customer has chosen a store, customers must decide on their shopping path within the store. This is the second part of the two-part purchase path, p. Using radio-frequency identification (RFID) tracking, Larson, Bradlow, and Fader (2005) and Hui and Bradlow (2012) find that most customers stay around the perimeter of the store and visit only a few aisles during their visit (see Figure 3.2). At the same time, the longer a customer is in the store and the more distance they travel in the store, the more likely they are to increase their unplanned purchases (Huang, Hui, Inman, Suher, 2013). Therefore, the more effectively stores attract a customer into a new department, the more likely they are to increase spending in the store.

Figure 3.2: Customer Paths around the Perimeter of Store
Larson, Bradlow, Fader (2005)



The customer's evaluation of their current path depends on a set of household-department match characteristics, K_{id} including preference for fresh produce, meats, and dairy versus preference for packaged goods, preference to have one-stop shopping trips and purchase items such as general medicine or household goods, and preference for freshly made products such as the salad bar or grocery bakery. The store's marketing strategies, M^n affecting the within store path are meant to address these concerns.

Figure 3.3: Conditional on their store choice, they choose which departments to walk through.

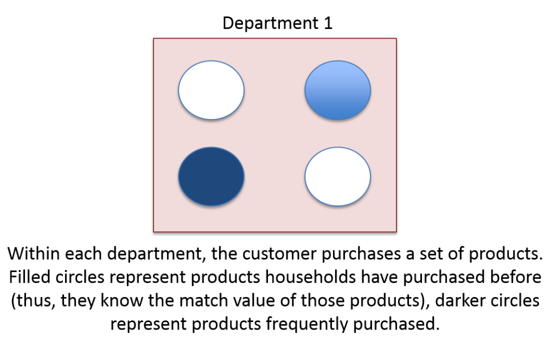


Once in the store, the customer chooses which aisles to visit.
Filled boxes represent visited aisles.

Households are susceptible to retailer marketing strategies, M^n for new purchase paths, p^n which can divert them from their status quo purchase path, P^s. In fact, on average a large portion of purchases are subject to in store decisions. According to the Point of Purchase Advertising Institute, only 24 percent of grocery purchases are specifically planned with the remainder affected by in-store decision making (POPAI, 2014). Inman, Winer, and Ferraro (2009) have shown that most grocery purchases are unplanned at the category level, giving room for in store cues.

Given a purchase path, customers evaluate products according to the linear ordering. Most literature on unplanned purchases is limited to surveys of customers before and after they enter the store (Beatty and Ferrell, 1998; Bell, Corsten, and Knox, 2011; Bucklin and Lattin, 1991; Inman, Winer, and Ferraro, 2009; Park, Iyer, and Smith, 1989). Huang, Hui, Inman, Suher (2013) shed more light into point-of-purchase drivers of unplanned consideration and purchase by tracking shoppers with video while in the store. They observe that unplanned considerations are more likely to turn to purchase when the shopper spends more time in consideration, engages in more product touches, views fewer product shelf displays, stands closer to the shelf, references external information, and interacts with store staff. Because people often consider a number of options before deciding on an option, product consideration is one of the more important factors which sway consumers decisions in the store (Roberts and Lattin, 1991), accounting for up to 70 percent of the variance in a choice (Hauser and Wernerfelt, 1989). Figure 3.4 demonstrates product choice conditional on department choice.

Figure 3.4: Conditional on a department choice, they choose to purchase certain products.



3.4.4 Firm Marketing Strategy Tools

Assume that the marketing strategy M is a collection of separable marketing tools which can be used to try to influence customers. In this model, I differentiate between store-level and household-level marketing tools. Store-level tools include displays and mailers and utilize aggregated information collected by the retailer at the store-level.

Displays highlight select products in the front of the store, the back of the store, on end-aisles, within aisles, or at check out in order to try to divert a customer off of her status quo path. In-store displays are an important marketing mechanism used by manufacturers to direct customers' attention (Nelson and Ellison, 2005). The goal of these displays is to increase consideration and purchase of the advertised products and bring them into aisles they may not have originally planned to visit.

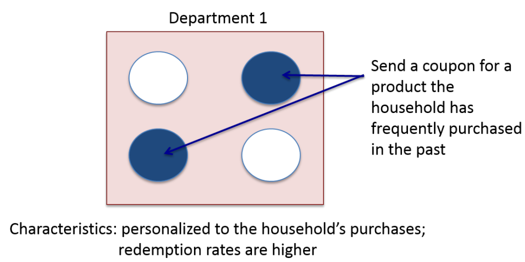
Mailers are store-level advertisements sent to households highlighting select products which are on sale or are still at regular price. The purpose of the marketing strategy is to attract customers' attention and motivate them to consider shifting their path to evaluate products highlighted in the advertisement. Mailers are also meant to remind customers of their household-store match value and encourage them to visit the store.

Personalized marketing utilize purchase patterns at the household-level. With data analytics, the store is able to track the purchase path of each household and identify strategies which may divert them off their path. There are two main types of marketing tools I evaluate: reward and promotion campaigns. Evaluation of these two targeted marketing tools is the key goal of this paper.

The purpose of the **reward campaign** is to improve to improve the lifelong value of a customer to the store by rewarding customers for their loyalty with coupons highly relevant to the customer. The campaign offers discounts for products the customer typically purchases, lowering the cost for the customer to visit the store, and highlighting the household-store match value. Since the campaign highlights products on the household's current purchase path, its purpose is more in increasing the likelihood of coming to the store in a given week rather than trying to influence its purchase path within the store.

Rewards have been shown to increase sales for retailers (Lewis, 2004; Dréze and Hoch, 1998). This positive response may be a result of gratitude for receiving the coupons best suited for them (Palmateer et al., 2009). Venkatesan and Farris (2012) give evidence that this gratitude may be shown through an exposure effect to customized coupons. Kumar and Leone (1998) also show that retailer coupons can increase the likelihood of shoppers choosing one store over another because of the decreased cost of shopping at that retailer. Since reward coupons are personalized, this decreased cost is likely amplified since the reward coupons are matched to customer preferences. Figure 3.5 demonstrates how reward campaigns send coupons for frequently purchased items.

Figure 3.5: Role of Reward Campaigns



The second household-level marketing tool is the **promotional campaign**. The purpose of the promotional campaign is like the displays and mailers in that it is intended to change the purchase path of the households. However, instead of using store-level data in the market strategy, the store uses the household-level purchase history to send messages relevant to the household. Unlike the reward campaigns, the purpose of the promotional campaigns is not to discount items the household frequently purchases. Rather, it is to highlight products relevant to the household, which are outside the typical purchase path, primarily through discounts. For example, for a household that tends purchase products from the perimeter of the store, a promotional campaign may feature a product in one of the center aisles in store which can be used with a produce or dairy product. If the household decides to search for the new product, $\phi(P^s, M_p^n) = 1$ and $p^n \succ P^s$, then the store has succeeded in moving the customer off her status quo path. While looking for this new product, the household evaluates a new set of products, deciding if they are worth purchasing.

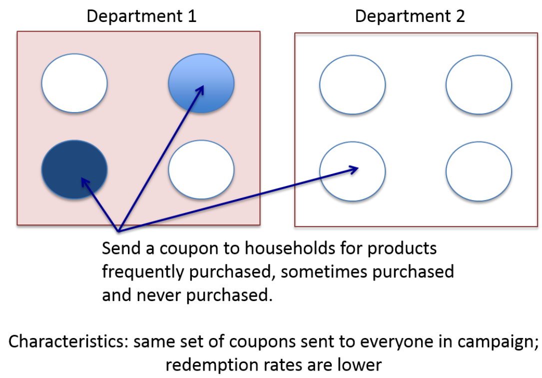
Promotions signal information about relevant products within the store. Since retailers carry a number of *experience goods* which must be consumed in order for households to determine their quality, signals of product relevance within the store can increase the indirect utility of the store to the household through the match-products-to-buyers effect (Nelson, 1974; Meurer and Stahl, 1994; Anderson and Renault, 2006). Relevance can maintain the returns to advertising when retailers have an opportunity to send targeted advertisements about products that the household is more likely to purchase. Further, targeted promotions may be instrumental in directing the focus of shoppers particularly when they have abstract shopping goals (Bell, Corsten, and Knox, 2011).

Promotion campaigns also serve to direct customers to departments highlighted on the coupon. Because coupons include products less frequently purchased, this can encourage a customer to go to a department outside their normal shopping path, enabling them to consider a product they otherwise would not have considered. Since reward coupons only discount items frequently purchased, they do not have the same effect on directing department sales. Figure 3.6 demonstrates how promotion campaigns send coupons for a mix of frequently purchased items, moderately purchased items, rarely purchased items.

3.5 Data Description

I have access to a unique dataset which contains the complete purchase history for 2,500 households in 582 store locations over a two-year (102 week) period from a single unidentified retailer. A subset of these households received either reward or promotional coupon campaigns which were targeted to them based on their purchase history. The data was collected prior to 2008, so it is not affected by the overall increase in coupon usage during the recession.

Figure 3.6: Role of Promotion Campaigns



The novel characteristic of the dataset is that beginning in week 33 of the first year, the retailer began reward and promotion coupon campaigns to a subset of its customers. This sample is a subset from a longer period of test and control samples the retailer employs in order to test the effectiveness of their campaigns. The beginning weeks of the sample serve as a control period during which no households receive any targeted campaigns. I utilize analysis of purchase information during this period to predict targeted campaign receipt during the test period.

During the test period, a total of 5 reward campaigns and 25 promotion campaigns were mailed to households in the form of paper coupons. Table 3.6 outlines the distribution of campaign recipients over the course of the test period. A subset of the households received no campaigns (37 percent), and a majority of the households received a combination of campaigns during the course of the test period (40 percent). There is significant variation in the number of coupon campaigns received. Table 3.7 breaks down the receipt of campaigns even further by showing the percent of households receiving the k^{th} reward or promotion campaign given at least one receipt of the other campaign. Although there were up to 25 promotion campaigns, the maximum number of promotion campaigns a household received in the sample was twelve. The maximum number of reward campaigns was five.

Table 3.8 outlines the weeks during which each campaign ran during the test period. Note that Campaign 26 is the first campaign starting at week 33 and Campaign 24 is the last campaign ending in week 102. The average campaign runs 6.1 weeks. The shortest campaign is 4 weeks and the longest is 23 weeks. Four weeks is the most common length of campaigns. The table also identifies which campaigns are reward and promotional campaigns: campaigns 8,13, 18, 26, and 30 are reward campaigns and the remaining campaigns are promotional. Finally, the table outlines the number of households receiving each type of campaign. Reward campaigns are sent to a wider set of households with the average number of recipients being 796 while the average number of recipients for the promotional campaigns is 129.

The dataset includes the point of sale data for each household over the two year period at each of the stores in the sample. Purchase information includes characteristics of product including its department, brand (whether private or national label), manufacturer, commodity description, sub-commodity description, the size of the product, and unique product identifier. I outline the number of manufacturers, brands, commodities, sub-commodities, and products in Table 3.9. I also indicate whether each department is coded as containing perishable products. I observe the quantity purchased, the final price paid and discounts provided (retailer, other coupon discounts, and match discounts), and the day and time of the sale at which store in the sample. The dataset also lists which households receive targeted campaigns including the reward and promotion campaigns and non-targeted campaigns including store-mailers. I observe which products are highlighted in the mailers and which products are highlighted on in-store displays. Finally, I observe which households redeemed the reward and promotion coupons.

Table 3.6: Distribution of Coupon Receipt

Coupon Receipt	Number of Households	% of Household
No Coupons	916	37%
Only Personalized Reward	508	20%
Only Targeted Promotion	71	3%
Both Coupons	1,005	40%
Total Households	2,500	100%

Table 3.7: Distribution of Coupon Receipt by Number Received

Number of Coupons Received	Number of Households	% of Households
0 Promotion Coupon (range 1-5 Reward)	508	20%
1 Promotion Coupon (range 1-5 Reward)	293	12%
2 Promotion Coupon (range 1-5 Reward)	205	8%
3 Promotion Coupon (range 1-5 Reward)	158	6%
4 Promotion Coupon (range 1-5 Reward)	114	5%
5 Promotion Coupon (range 1-5 Reward)	90	4%
6 Promotion Coupon (range 1-5 Reward)	50	2%
7 Promotion Coupon (range 1-5 Reward)	43	2%
8 Promotion Coupon (range 2-5 Reward)	28	1%
9 Promotion Coupon (range 2-5 Reward)	11	0%
10 Promotion Coupon (range 2-5 Reward)	8	0%
11 Promotion Coupon (range 2-5 Reward)	3	0%
12 Promotion Coupon (range 4-5 Reward)	2	0%
Total Receiving Both Coupons	1,005	40%
0 Reward Coupon (range 1-3 Promotion)	71	3%
1 Reward Coupon (range 1-7 Promotion)	117	5%
2 Reward Coupon (range 1-11 Promotion)	187	7%
3 Reward Coupon (range 1-11 Promotion)	462	18%
4 Reward Coupon (range 1-12 Promotion)	126	5%
5 Reward Coupon (range 1-12 Promotion)	113	5%
Total Receiving Both Coupons	1,005	40%

Table 3.8: Weeks of Campaigns

Campaign Type	**Campaign**	**Start week**	**End week**	**Num. of weeks**	**Number of HH**
Promotional	1	50	55	5	13
Promotional	2	51	55	4	48
Promotional	3	52	60	8	12
Promotional	4	54	58	4	81
Promotional	5	55	59	4	166
Promotional	6	57	61	4	65
Promotional	7	58	62	4	198
Reward	8	60	66	6	1,076
Promotional	9	63	67	4	176
Promotional	10	67	71	4	123
Promotional	11	69	75	6	214
Promotional	12	69	73	4	170
Reward	13	73	79	6	1,077
Promotional	14	77	86	9	224
Promotional	15	79	102	23	17
Promotional	16	81	85	4	188
Promotional	17	83	87	4	202
Reward	18	85	92	7	1,133
Promotional	19	87	91	4	130
Promotional	20	89	99	10	244
Promotional	21	90	94	4	65
Promotional	22	90	94	4	276
Promotional	23	93	98	5	183
Promotional	24	95	102	7	100
Promotional	25	95	99	4	187
Reward	26	33	38	5	332
Promotional	27	35	44	9	12
Promotional	28	38	46	8	17
Promotional	29	41	48	7	118
Reward	30	47	53	6	361

Table 3.9: Department Product Description

Department	Total Manufacturer	Total Brand	Total Commodity	Total Sub-Commodity	Total Product	Perishable
AUTOMOTIVE	2	2	2	2	2	0
CHARITABLE CONT	2	1	1	2	2	0
CHEF SHOPPE	8	1	1	5	14	1
CNTRL/STORE SUP	4	2	1	3	4	0
COSMETICS	173	2	5	99	3,011	0
COUP/STR						
MFG	14	1	1	1	39	0
DAIRY DELI	2	1	2	3	3	1
DELI	822	2	13	109	2,354	1
DELI/SNACK BAR	2	1	1	2	2	1
DRUG GM	1,985	2	91	894	31,529	0
ELECT						
PLUMBING	1	1	1	1	1	0
FLORAL	197	2	11	104	938	0
FROZEN GROCERY	4	1	2	13	23	0
GARDEN CENTER	20	2	2	30	128	0
GM MERCH EXP	3	1	2	2	3	0
GRO BAKERY	1	1	1	1	2	1
GROCERY	1,527	2	94	736	39,021	0
HBC	1	1	1	1	1	0
HOUSEWARES	1	1	1	1	1	0
KIOSK-GAS	1	1	2	1	16	0
MEAT	1,033	2	15	161	2,544	1
MEAT-PCKGD	160	2	10	105	2,427	1
MEAT-WHSE	1	1	1	1	1	0
MISC SALES TRAN	13	2	1	13	88	0
MISC. TRANS.	302	2	1	43	490	0
NUTRITION	257	2	25	158	2,914	0
PASTRY	144	2	9	78	2,149	1
PHARMACY SUPPLY	1	1	1	1	1	0
PHOTO	2	1	1	2	2	0
PORK	1	1	1	1	1	1
POSTAL CENTER	2	1	2	3	3	0
PROD-WHS SALES	1	1	1	2	2	0
PRODUCE	252	2	32	202	3,118	1
RESTAURANT	8	1	3	13	102	0
RX	9	2	2	8	9	0
SALAD BAR	16	2	2	7	48	1
SEAFOOD	219	2	4	15	369	1
SEAFOOD-PCKGD	76	2	5	32	563	0
SPIRITS	49	1	1	36	377	0
TOYS	2	1	1	2	3	0
TRAVEL						
LEISUR	4	1	4	7	28	0
VIDEO	2	1	1	1	2	0
VIDEO RENTAL	1	1	1	1	3	0

In Subsection 3.6.1, I outline my strategy for predicting the receipt of the targeted coupon campaigns. Summary statistics for the predictors are outlined in Table 3.10. Note that the variables, Reward Camp Running and Promotion Camp Running are indicator variables for the weeks in which campaigns are on according to Table 3.8. Variables, Daily, Twice Weekly, One & Half Weekly, Weekly, Biweekly, Infrequent, Low Sales, Low Medium Sales, Medium Sales, Medium High Sales and High Sales, indicate the average sales and average frequency with which households visited the store during the control period. These *loyalty* variables are important predictors for receipt of the campaigns. The definitions of these variables are in Tables 3.18 and 3.19. The retailer observes the responsiveness in terms of average change in sales of households to previous campaigns with the variables, Δ Spend in L.Reward Camp and Δ Spend in L.Promotion Camp. Finally, the retailer matches household purchase behavior in the control period to products in each campaign through variables including Sum[% Purch in Dept], Sum[% Purch of Manu], Sum[% Purch of Brand], Sum[% Purch of Commodity], Sum[% Purch of Sub-commodity], Sum[% Purch of Manu-Sub-comm], Sum[% Purch of Man-Comm], Sum[% Purch of Man-Dept], Sum[% Purch of Brand-Sub-comm], Sum[% Purch of Brand-Comm], and Sum[% Purch of Brand-Dept]. Store and department level predictors are equal apart from the indicator variable, Promotion Camp Running in Dept, which also identifies which departments are included in the promotional campaigns.

Table 3.11 provides the summary statistics for the main estimation equations outlined in Section 3.6. We observe that the average store level weekly sales is \$31.65 with a high standard deviation of \$56.43. The average visits to the store for households over the course of the sample is 0.49 with high variation of a 0.5 standard deviation. We also observe that department sales vary even more (both across households, over time, and across departments), with average weekly department sales of \$0.74 and a standard deviation of \$6.09.

Table 3.10: Campaign Predictor Summary Statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Reward Camp Running	0.34	0.47	0	1
Promotion Camp Running	0.65	0.48	0	1
Daily	0.08	0.27	0	1
Twice Weekly	0.22	0.41	0	1
Weekly	0.13	0.34	0	1
Oneout	0.19	0.39	0	1
Biweekly	0.24	0.43	0	1
Infrequent	0.14	0.35	0	1
Low Sales	0.46	0.5	0	1
Low Medium Sales	0.22	0.41	0	1
Medium Sales	0.2	0.4	0	1
Medium High Sales	0.09	0.29	0	1
High Sales	0.03	0.18	0	1
Sum[% Purch in Dept]	16.84	41.97	0	366.33
Sum[% Purch of Manu]	9.89	26	0	200.95
Sum[% Purch of Brand]	18.2	45.46	0	413.86
Sum[% Purch of Commodity]	10.39	27.75	0	179.41
Sum[% Purch of Sub-commodity]	7.18	19.9	0	164.71
Sum[% Purch of Manu-Sub-comm]	5.15	15.2	0	133.33
Sum[% Purch of Man-Comm]	6.25	17.8	0	133.33
Sum[% Purch of Man-Dept]	8.82	23.88	0	166.67
Sum[% Purch of Brand-Sub-comm]	6.31	17.84	0	164.71
Sum[% Purch of Brand-Comm]	9.06	24.58	0	179.41
Sum[% Purch of Brand-Dept]	15.55	38.72	0	327.04
Δ Spend in L.Reward Camp	0.75	7.09	-100.93	188.81
Δ Spend in L.Promotion Camp	0.62	6.58	-75.67	207.59
N		254,490		
Promotion Camp Running in Dept	0.07	0.25	0	1
N		10,943,070		

Department promotion predictions use same predictors except for department level running variable.

Note that the average number of weeks in which households receive reward coupons is slightly higher (at 0.11) than the average number of weeks in which households receive promotional coupons (at 0.07). These are much smaller than the average weeks in which households receive store-level mailers (0.69). The average for departments to be in mailers is 0.26 while the average for departments to be in displays is 0.2. Table 3.13 outlines the average departments highlighted in mailers and displays by department. We can also see the average discounts from retail discounts, other coupon discounts, and match coupon discounts relative to the average discounts for the reward and promotional coupons (see Table 3.12). This table outlines the average discount per targeted coupon redeemed and the average weekly discounts applied with the reward and promotional campaigns. We do not observe the discounts given on each coupon sent, therefore, the redemptions give an imperfect indication of the discounts offered in each targeted campaign.

Redemption rates are higher for reward coupons than for promotion coupons as seen in Table 3.14. This is understandable since rewards are coupons for products that the customer purchases frequently. We see that only one percent of promotional coupons and three percent of reward coupons are redeemed. Households receiving promotion campaigns receive a pamphlet with a set of coupons ranging in number between 1 and 34 whereas the average number of coupons sent as reward coupons is 16.4. I outline the number redemptions for each campaign in Table 3.15 and the number of redemptions by campaign and department in Table 3.16.

Reward campaign coupons are individualized to each household in the campaign. I unfortunately do not observe the coupons received at the household level for reward campaigns, but I am able to observe the campaign level products. Given the number of households in each reward campaign, this hinders my analysis of reward coupons beyond the store level. However, at the campaign level, I am able to observe some characteristics of the reward campaigns relative to the promotion campaigns. For

Table 3.11: Main Equation Summary Statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Store Weekly Sales	31.65	56.43	0	1,283.33
Visited Store	0.49	0.5	0	1
Reward Coupon	0.11	0.32	0	1
Promotion Coupon	0.07	0.25	0	1
Reward Redemption	0	0.06	0	1
Promotion Redemption	0	0.03	0	1
Retail Discounts	-5.49	11.1	-456.26	0
Other Coupon Discounts	-0.17	1.12	-79.43	0
Match Coupon Discounts	-0.03	0.26	-20.8	0
Mailer	0.69	0.46	0	1
Reward Coupon Residual	0	0.05	-0.88	0.83
Promotion Coupon Residual	0	0.1	-1.93	1.13
Season 1	0.24	0.42	0	1
Season 2	0.24	0.42	0	1
Season 3	0.24	0.42	0	1
Season 4	0.24	0.42	0	1
L.[% Basket Perishable Goods] - Ave[% Basket Perishable Goods]	-1.15	18.31	-71.2	96
N		254,490		
Department Weekly Sales	0.74	6.09	0	876.02
Promotion Department Coupon	0.01	0.08	0	4
Reward Department Redemption	0	0.02	0	1
Promotion Department Redemptions	0	0.01	0	1
Department Retail Coupon	-0.13	1.34	-455.26	0
Department Other Coupon	0	0.12	-55.93	0
Department Match Coupon Discounts	0	0.04	-20.8	0
Department Mailer	0.26	0.44	0	1
% Mailer in Department	2.06	9.79	0	86.18
Department Display	0.2	0.4	0	1
Weeks Since Last Visit	3.36	7.57	0	94
Promoted Department Coupon Residual	0	0.05	-1.74	2.07
Department Dummies	0.02	0.15	0	1
N		10,943,070		

Table 3.12: Coupon Discounts

Coupon Discount	Mean	St. Dev.	Min	Max
Coupon Discount	-0.97	0.54	0.08	6.00
Total Weekly Discount	-1.70	1.54	0.18	19.20

Table 3.13: Summary Mailer and Display by Department

Location of product in the store	Mean Mailer Department	Mean Display Department
AUTOMOTIVE	0.00	0.00
CHARITABLE CONT	0.00	0.01
CHEF SHOPPE	0.04	0.00
CNTRL/STORE SUP	0.00	0.00
COSMETICS	0.69	0.67
COUP/STR & MFG	0.24	0.13
DAIRY DELI	0.00	0.00
DELI	0.88	0.64
DELI/SNACK BAR	0.00	0.00
DRUG GM	0.88	0.86
ELECT &PLUMBING	0.00	0.00
FLORAL	0.81	0.26
FROZEN GROCERY	0.10	0.00
GARDEN CENTER	0.16	0.01
GM MERCH EXP	0.00	0.00
GRO BAKERY	0.00	0.00
GROCERY	0.88	0.86
HBC	0.00	0.00
HOUSEWARES	0.00	0.00
KIOSK-GAS	0.00	0.00
MEAT	0.88	0.38
MEAT-PCKGD	0.88	0.86
MEAT-WHSE	0.00	0.00
MISC SALES TRAN	0.01	0.06
MISC. TRANS.	0.06	0.37
NUTRITION	0.87	0.84
PASTRY	0.88	0.57
PHARMACY SUPPLY	0.01	0.00
PHOTO	0.01	0.03
PORK	0.00	0.00
POSTAL CENTER	0.02	0.03
PROD-WHS SALES	0.00	0.00
PRODUCE	0.88	0.77
RESTAURANT	0.01	0.03
RX	0.00	0.01
SALAD BAR	0.05	0.00
SEAFOOD	0.74	0.15
SEAFOOD-PCKGD	0.88	0.74
SPIRITS	0.12	0.40
TOYS	0.00	0.00
TRAVEL & LEISUR	0.27	0.05
VIDEO	0.00	0.00
VIDEO RENTAL	0.00	0.00
Total	0.26	0.20

Table 3.14: Coupon Redemptions

Coupon Type	Coupons Sent	Total Coupon Redemptions	Number of Campaigns
Reward	63,664	1,791	5
Promotion	53,617	527	25

Table 3.15: Coupon Redemptions by Campaign

Campaign	Redemptions	Redeeming Households
1	1	1
2	5	2
3	2	2
4	11	6
5	13	8
6	1	1
7	7	5
8	372	158
9	43	20
10	15	10
11	8	6
12	26	11
13	629	196
14	34	18
15	2	2
16	43	19
17	45	18
18	653	214
19	29	15
20	33	20
21	5	4
22	47	17
23	60	23
24	10	7
25	61	24
26	73	31
27	1	1
28	1	1
29	24	13
30	64	36

Table 3.16: Coupon Redemptions by Campaign Type & Department

Campaign Type	Department	Redemptions in Department
Promotional	COSMETICS	18
Promotional	COUP/STR MFG	10
Promotional	DELI	1
Promotional	DRUG GM	145
Promotional	GROCERY	357
Promotional	MEAT-PCKGD	10
Promotional	MISC. TRANS.	10
Promotional	NUTRITION	13
Promotional	PASTRY	1
Promotional	PRODUCE	5
Promotional	SEAFOOD-PCKGD	4
Reward	CHEF SHOPPE	98
Reward	COSMETICS	35
Reward	COUP/STR MFG	115
Reward	DAIRY DELI	41
Reward	DELI	313
Reward	DRUG GM	338
Reward	FLORAL	98
Reward	FROZEN GROCERY	98
Reward	GARDEN CENTER	98
Reward	GM MERCH EXP	152
Reward	GRO BAKERY	98
Reward	GROCERY	1564
Reward	HBC	31
Reward	MEAT	329
Reward	MEAT-PCKGD	371
Reward	MISC SALES TRAN	208
Reward	MISC. TRANS.	381
Reward	NUTRITION	285
Reward	PASTRY	148
Reward	PHARMACY SUPPLY	98
Reward	PHOTO	41
Reward	PORK	121
Reward	PRODUCE	257
Reward	RX	31
Reward	SALAD BAR	177
Reward	SEAFOOD	168
Reward	SEAFOOD-PCKGD	312
Reward	TRAVEL LEISUR	98

promotional campaigns, a smaller set of households receive the same set of coupons, and I observe each product sent in promotional campaigns.

Reward campaigns on average cover a broader set of products, manufacturers, and departments reflecting the individualization of each reward campaign to the households' preferences. As Table 3.17 shows, promotional campaigns are concentrated in the Grocery and Drug GM departments while the reward campaigns are more evenly distributed. While coupons in both types of campaigns feature national manufacture brands more than private labels, more of the promotion coupons (90.5 percent) are for national brands than personal reward coupons (86 percent). Ninety-three percent of promotion coupons were for one department, with the maximum range up to four departments in one coupon (2 percent) while 75 percent of reward coupons were for one department, with the maximum number of departments covered in one coupon going up to nineteen (5 percent).

Table 3.17: Department distribution of Coupons by Type

Department	**Reward**	**Promotion**
Grocery	28.0	75.8
Drug GM	16.7	15.0
Meat	14.1	-
Produce	12.8	0.4
Meat Packaged	12.6	0.5

Both types of campaigns include a message of appreciation for the customer's loyalty with the coupon packet, stating that the coupons were specially chosen for the household. The reward message emphasizes that the "exclusive offers" help the customer buy more of what they like best while the promotion message suggests that the coupons fit the shopping pattern and includes more product information and recipes using the promoted items. Since reward coupons do not convey new information, the coupons are simple in form. Figures 3.7 3.8 show an example of a reward coupon cover and a subset of coupons from a reward campaign.

Figure 3.7: Reward Cover

Figure 3.8: Reward Coupons

Since promotion campaigns convey new information in line with the revealed preferences of the customers, the coupons are more colorful and exhibit more product information. Figures 3.9 and 3.10 show an example of a promotion coupon cover and inset with descriptions, pictures, and text about the product. Table **??** demonstrates a representative example of the purchase history of households for coupons in a typical promotion campaign. Here household 93 receives coupons for 9 product commodities in campaign 1, and I outline the number of times the household purchased these products during the control period. The pattern of frequent, moderate, and low purchases is common across households and promotion campaigns.

Figure 3.9: Promotion Cover

3.6 Empirical Strategy

The main estimation goal is to identify the comparative effectiveness of personalized and generic marketing tools on the purchase behavior of customers. According to our theoretical model, personalized advertisements are more likely to be considered by the customer, with $E[\phi(P^s, m_p)] > E[\phi(P^s, m_g)]$. If a customer does not

Figure 3.10: Promotion Coupons

consider an advertisement, there is no effect on the purchase path. However, if the consumer considers a campaign, there is potential that she will change her purchase path and expand her consideration set. The expansion of her consideration set has a non-negative expected change to her expected purchases. I outline the impact of consideration of advertisements below. I test the effectiveness of each marketing tool by estimating the average change in purchase behavior induced by a set of marketing tools at a single retailer.

- *Consideration, No Path Change*: If a product is already in the shopping list, there is no change to the consideration set or the purchase path.

- *Consideration, Path Change*: If the product is not in the shopping list, then the purchase path changes. Change depends on how far the product is from the status quo purchase path.

– **Same store, same department**: If the product is along the purchase path, the only impact to the consideration set is in whether to purchase the product when in the department.

– **Same store, new department**: If the product brings the customer to a new department, the impact of the advertisement is to expand the consideration set to the advertised products and the set of products potentially added to the consideration set while the customer is in the department (for example, through retail discounts for products to which the customer would not otherwise be exposed).

– **Different store, same department**: If the product brings the customer to a new store with the same department, it could induce her to do her primary shopping at this store rather than her other preferred store.

– **Different store, new department**: If the product brings the customer to a new store with a new department, it could expand her shopping as in **Same store, new department**, but in the new store.

Recall that reward campaigns are individualized to households and offer discounts for products the customer already purchases. This implies that even if customers consider the advertisements, there is little expected impact on the store purchases, unless the campaign induces the customers to switch the ordering of the stores they visit. As such, the main channel through which we expect to observe a change in customer purchase paths is through **Different store, same department**, which would yield an average increase in likelihood of a customer coming to the store. If the increased likelihood also changed the ordering of store visits (i.e., changed the second preferred store to be the preferred store), we may also observe an increase in store level sales because the customer is now doing her main shopping at the new preferred store. Because on average we expect reward campaigns to fall in the category, **Same**

store, same department, so we expect to see limited change in sales conditional on coming to the store and limited change in sales at the department level.

Promotional campaigns are also targeted to customers, but they include coupons for products the household does not regularly purchase. Therefore, we expect promotional campaigns to have the largest impact on changing the purchase path within the store. Thus, promotional campaigns are most likely aligned with **Same store, new department** and **Different store, new department**. Recall that if advertisements induce a change in the customer's purchase path to a new department, this will on average increase the expected consideration set for the customer and increase the expected sales both within the department and at the store level, even when a customer does not necessarily redeem the coupon. The **Different store, new department** effect also would also yield an increase in average likelihood of customers coming to the store in the weeks of the campaign in the same way that **Different store, same department** would. The only difference is that we would expect to see an even greater increase in store sales and department sales conditional on coming to the store for the **Different store, new department** case.

Mailers and display marketing tools are generic marketing tools which we expect to have little impact on purchase behavior because they are unlikely to be considered. Within the department, localized marketing tools such as retailer discounts are expected to have a greater impact once customers are in the department. We expect the localized marketing tools to not alter customer's purchase path but increase sales along the purchase path (either status quo or new). Therefore we would expect to see greater impact of localized marketing tools in the new departments visited during campaigns which alter the customer's purchase path.

Therefore, the model predicts that the promotion campaigns will be most effective, followed by the reward, display, and mailer tools. In order to test this prediction, I estimate the effect of each marketing tool on store sales in the weeks of the campaigns for each household. I will first estimate the average effect on overall sales at the store in order to get a big picture of how each of the tools affect customer behavior. Next, I will decompose the effect of each tool on bringing customers to the store and change in sales conditional on them coming to the store. Finally, I will directly estimate the effect of the marketing tools on changing the purchase path of customers within the store by estimating the impact each campaign has on the department level sales.

Since the targeted marketing tools are sent to households with specific characteristics, we will estimate biased coefficients if we do not control for the endogeneity of the variables. In particular, the store tends to send reward and promotion campaigns to households which exhibit higher loyalty characteristics including higher average sales and higher likelihood of coming to the store in a given week. Without accounting for this endogeneity, we would expect upwardly biased results for these campaigns. I use the control function technique to account for the endogeneity of these marketing tools. Since the display and mailer campaigns are chosen at the store level, it is reasonable to assume that these campaigns are independent of the individual household characteristics.

Before describing the empirical approach in more detail, I will define the key variables of interest. Let the indicator variable, v_{it} capture whether household i visits a grocery store in week t where $v_{it} = 1$ if the latent variable, $v_{it}^* > 0$. The amount i spends at the store (i.e., the basket size) is denoted by the latent continuous variable S_{it}^*, where the realized spending conditional on coming to the store equals $S_{it} = S_{it}^* v_{it}$.

There are two types of targeted coupons, $\mathbf{C}_{it}$: reward campaigns, R_{it} and promotion campaigns, P_{it}. Let $R_{it} = 1$ if household i received a reward campaign in week t. Similarly, let $P_{it} = 1$ if a household received a promotion campaign in a given week t. Additionally, for store level marketing tools: mailers, m_{jt} and displays, d_{jt}, let each equal 1 if store j sends a mailer or exhibits a display in week t.

3.6.1 Endogenous Targeted Campaign Predictors

Before outlining the estimation strategy for the effectiveness of the different marketing tools used by the retailer, I will outline the variables used to estimate the receipt of each of the targeted coupon campaigns, $\mathbf{C}_{it}$.

During the test period of the sample, there are 30 separate campaigns, $k = 1, ..., 30$, which can be categorized as either a reward, R_{it}, or promotion, P_{it}, campaign. For the purposes of this section, I will denote each campaign type by the variable, $m \in \{r, p\}$ in order to specify variables pertaining to one campaign type or the other. I will simultaneously estimate two control functions for each household i in each week t. The exogenous predictors for these campaigns can be categorized into four types.

First, I account for household store loyalty characteristics, $\mathbf{loyalty}_i$, a (1×11) vector of dummy variables, denoted as $\mathbf{W_{it}}$ in Table 3.20. Variables included in this vector capture average frequency of coming to the store in a given week and average spending when customers come to the store during the control period in which no households received any campaigns. In conversations, the retailer identified these variables as important variables in deciding which households get campaigns. Variables include $Daily_i$, $TwiceWeekly_i$, $One\&HalfWeekly_i$, $Weekly_i$, $Biweekly_i$, $Infrequent_i$, $LowSales_i$, $Low-MediumSales_i$, $MediumSales_i$, $Medium-HighSales_i$ and $HighSales_i$. The definitions for each of these variables are listed in Table 3.18 and 3.19. These variables are household-specific, and time-, campaign-, and campaign type-invariant.

Second, I account for the retailer learning about households based on their average responsiveness to previous campaigns of each type m with the (1×2) vector, $\mathbf{learning}_{i,m_{\overline{k}_{-t}}}$, where k_{-t} reflects the previous campaigns, $\overline{k}_{-t}$ reflects the average

Table 3.18: Loyalty Categories: Average Weekly Sales

Average Weekly Sales Cut Off	**Label**
$\leq$ \$13	Low Sales
\$13–\$25	Low-Medium Sales
\$25–\$50	Medium Sales
\$50–\$90	Medium-High Sales
$\geq$ \$90	High Sales

Table 3.19: Loyalty Categories: Average Days Between Visits

Average Days Between Visits Cut Off	**Label**
$\leq$ 3 Days	Daily
3–7 Days	Twice Weekly
7–10 Days	Weekly
10–16 Days	One & Half Weekly
16–35 Days	Bi-Weekly
$\geq$ 35 Days	Infrequent

across these previous campaigns, and $m_{\overline{k}_{-t}}$ reflects the average across these previous campaigns for each campaign type, m. The $\mathbf{learning}_{i,m_{\overline{k}_{-t}}}$ variables are denoted as $\mathbf{Z_{it}}$ in Table 3.20. I measure the average change in total basket size during the weeks of previous campaigns for each type, m. I predict that if the household increased spending during the average four weeks of the previous campaigns, this would indicate the household is a good candidate for future campaigns. The total basket size is net of coupon discounts. These variables are household-, time-, campaign-, and campaign-type-specific.

Third, I account for the match between the campaign product j promoted and household products purchased in the control period with the (1×11) vector, $\mathbf{match}_{i,m_{k_j}}$, where k_j reflects each product j within each campaign k. Variable $\mathbf{match}_{i,m_{k_j}}$ is captured as $\mathbf{Z_{it}}$ in Table 3.20. The match is based on purchase behavior for each household i during the control period. For example, Sum[% Purch of Commodity]$_{i,m_{k_j}}$ captures the percentage of purchases made by household i of each commodity in campaign k_j during the control period. Campaign 6 only has coupons for yogurt, so Sum[% Purch of Commodity]$_{i,m_{k_j}}$ accounts for the percent of yogurt purchases relative to all commodity purchases during the control period. However, campaign 19 has five separate commodities: baby foods, diapers & disposables, film & camera products, infant care products, and infant formula. For campaign 19, Sum[% Purch of Commodity]$_{i,m_{k_j}}$ accounts for the sum of percentages of each commodity in campaign 19 purchased by i during the control period. I also account for the percentage of household purchases from the manufacturers highlighted in campaign k with Sum[% Purch of Manu]$_{i,m_{k_j}}$, the percentage of household purchases from a manufacturer for the specific commodities in the campaign with Sum[% Purch of Man-Comm]$_{i,m_{k_j}}$, and the manufacturer for the specific sub-commodity with Sum[% Purch of Manu-Sub-comm]$_{i,m_{k_j}}$. The other variables include Sum[% Purch in Dept]$_{i,m_{k_j}}$, Sum[% Purch of Brand]$_{i,m_{k_j}}$, Sum[% Purch of Sub-commodity]$_{i,m_{k_j}}$, Sum[% Purch of Man-Dept]$_{i,m_{k_j}}$, Sum[% Purch of Brand-Dept]$_{i,m_{k_j}}$, Sum[% Purch of Brand-Comm]$_{i,m_{k_j}}$, Sum[% Purch of Brand-Sub-comm]$_{i,m_{k_j}}$. These variables are household-, campaign-, and campaign-type-specific at the product level. There is no time-varying characteristic in this variable since match occurs with control period purchase data. Since each campaign offers discounts to a different set of products, the match analysis occurs at the campaign level for each coupon type.

Fourth, I limit prediction of campaign receipt to the weeks of the campaigns of each type, $\mathbf{CampaignRunning_{m_k}}$, denoted as $\mathbf{R_t}$ in Table 3.20. Since the 30 campaigns run during pre-specified weeks of the test period, I utilize the observed weeks of each campaign so I can better predict which households receive which campaigns. The variable *RewardRunning* captures the weeks during which Reward Campaigns are running and *PromotionRunning* captures the weeks during which Promotion Campaigns are running. This indicator variable is multiplied by each variable in the prediction regression so that I isolation prediction estimation only to the weeks predetermined by the retailer prior to the start of the sample. The motivation behind this is the manufacturers and the retailer engage in detailed planning to come up with the timing and content of campaigns months in advance. Once product set and weeks of the campaign are set, the retailer identifies the households who will be matched with the campaigns.

Finally, when estimating the predictions, I include all other exogenous variables in the prediction. These variables include, $\mathbf{X_{1it}}$ and $\mathbf{X_{2it}}$ as outlined in Table 3.20.

I use linear probability to predict selection of household i into campaign type m by household i. Since I use the predicted residuals to control for endogeneity in the estimation of coupon impact on sales, use of linear probability considerably simplifies analysis. For each campaign type m, I regress

$$\mathbf{C}_{it}|\tilde{\mathbf{Z}}_{it} = \mathbf{R_t} \cdot [\mathbf{W}_{it}, \mathbf{Z}_{it}]\mathbf{\Gamma}_1 + [\mathbf{X}_{1it}, \mathbf{X}_{2it}]\mathbf{\Gamma}_2 + \boldsymbol{\epsilon}_{it} \equiv \tilde{\mathbf{Z}}_{it}\mathbf{\Gamma} + \boldsymbol{\epsilon}_{it} \tag{3.1}$$

3.6.2 Regression of Expected Sales

The first step in assessing the effect of each marketing campaign on sales is to get an overall view of how the coupons affect expected sales at the store. Targeted coupons sent to households serve as a form of advertisement to the household for the store affecting the probability a customer visits the store and the basket size conditional on coming to the store. I model the expected sales for household i in a given week t *unconditional* on coming to the store using a linear regression

$$S_{it} = \beta_R R_{it} + \beta_P P_{it} + \beta_m m_{jt} + \mathbf{loyalty}_i \boldsymbol{\beta}_{loyal} + \mathbf{s}\boldsymbol{\beta}_s + u_{1it}. \tag{3.2}$$

I do not include a variable to account for displays at the store level because the stores in the dataset have displays for each of the weeks of the sample. I include a set of household loyalty characteristics, $\mathbf{loyalty}_i$ to capture the average basket size and likelihood of coming to the store in a given week for each household. Since seasonality affects grocery store purchase patterns, I account for the seasons of the year with for indicator variables, s_1, s_2, s_3, s_4.

Control Function Strategy for Linear Regression Model

Let $\mathbf{C}_{it} = \{R_{it}, P_{it}\}$ be the 1×2 vector of the endogenous coupon receipt variables modeled in Subsection 3.6.1. Following Wooldridge (2007), we use the model outlined above in Equation 3.1:

$$\mathbf{C}_{it}|\tilde{\mathbf{Z}}_{it} = \tilde{\mathbf{Z}}_{it}\mathbf{\Gamma} + \boldsymbol{\epsilon}_{it}$$

where $\tilde{\mathbf{Z}}_{it}$ is exogenous in the sense that it satisfies orthogonality conditions with u_{1it}. As in the case of two-stage least squares, the linear projection of $\mathbf{C}_{it}$ onto the exogenous variables enables us to isolate the part of the coupon model which is correlated with the error, u_{1it}, from Equation 3.2. The endogeneity of $\mathbf{C}_{it}$ implies that it does not satisfy the zero covariance condition, $E(\mathbf{C}'_{it}u_{1it}) \neq \mathbf{0}$. This means that u_{1it} is correlated with $\boldsymbol{\epsilon}_{it} = (\epsilon_{itR}, \epsilon_{itP})$. We can write the linear projection of u_{1it} on $\boldsymbol{\epsilon}_{it}$ as

$$u_{1it} = \boldsymbol{\epsilon}_{it}\boldsymbol{\rho} + e_{it}, \tag{3.3}$$

where $\boldsymbol{\epsilon}_{it}$ is a 1×2 vector, $\boldsymbol{\rho}$ is a 2×1 vector, and e_{it} is a scalar and $\boldsymbol{\rho} = E(\boldsymbol{\epsilon}'_{it}\boldsymbol{\epsilon}_{it})^{-1}E(\boldsymbol{\epsilon}'_{it}u_{1it})$. Since $\tilde{\mathbf{Z}}_{it}$ is uncorrelated with u_{1it} and $\boldsymbol{\epsilon}_{it}$, we have $E(\boldsymbol{\epsilon}'_{it}e_{it}) = \mathbf{0}$. Plugging in equation 3.3 into equation 3.2 gives us

$$S_{it} = \beta_R R_{it} + \beta_P P_{it} + \beta_m m_{jt} + \mathbf{loyalty}_i\boldsymbol{\beta}_{loyal} + \mathbf{s}\boldsymbol{\beta}_s + \rho_R\epsilon_{itR} + \rho_P\epsilon_{itP} + e_{it} \tag{3.4}$$

where $\boldsymbol{\epsilon}_{it}$ can be seen as an explanatory variable in the equation. Since $\mathbf{C}_{it}$ is a linear function of both $\tilde{\mathbf{Z}}_{it}$ and $\boldsymbol{\epsilon}_{it}$, this implies that e_{it} is also uncorrelated with $\mathbf{C}_{it}$.

Since we do not observe $\boldsymbol{\epsilon}_{it}$, we must estimate it using the first stage regression of $\mathbf{C}_{it}$ onto $\tilde{\mathbf{Z}}_{it}$ in equation 3.1. We then recover the residual, $\hat{\boldsymbol{\epsilon}}_{it} = \mathbf{C}_{it} - \tilde{\mathbf{Z}}_{it}\hat{\boldsymbol{\Gamma}}$. Substituting the residual into 3.4 gives us

$$S_{it} = \beta_R R_{it} + \beta_P P_{it} + \beta_m m_{jt} + \mathbf{loyalty}_i \boldsymbol{\beta}_{loyal} + \mathbf{s}\boldsymbol{\beta}_s + \rho_R \hat{\epsilon}_{itR} + \rho_P \hat{\epsilon}_{itP} + e_{it} \quad (3.5)$$

where for each observation, $error_{it} = e_{it} + \boldsymbol{\rho}\tilde{\mathbf{Z}}_{it}[\hat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}]$, which depends on the sampling error in $\hat{\boldsymbol{\Gamma}}$ unless $\boldsymbol{\rho} = \mathbf{0}$. This approach gives consistent control estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$. Here the $\hat{\epsilon}_{itR}$ controls for the endogeneity of $\mathbf{C}_{it}$ in 3.2, although with some sampling error since $\hat{\boldsymbol{\Gamma}} \neq \boldsymbol{\Gamma}$.

3.6.3 Decomposing Store Level Estimates

When assessing the effectiveness of marketing tools in the retail setting, it is important to disentangle the tools' effectiveness in bringing customers to the store and inducing them to purchase more once they are in the store. The most suitable econometric model for this analysis is the approach developed by Heckman (1976), otherwise known as the Type 2 Tobit Model (see Amemiya, 1985 and Cameron and Trivedi, 2005). This model is particularly useful since the endogenous targeted coupons affect both the likelihood of coming to the store and the household spending conditional on coming to the store. Once in the store, other factors also affect spending, including redemption of coupons, displays, and other store discounts. This model provides the flexibility needed to control for the endogeneity of the targeted coupon receipt using control functions and to account for the different factors which can affect store spending.

Recall that the purchase path, P, of each household includes both the path to the store and the path conditional on coming to the store. The Type 2 Tobit Model enables us to identify the effectiveness of each campaign in impacting each component of the consumer purchase path. Here, let v_{it} be a discrete variable equal to 1 when household I decides to go to the store in week t. The factors affecting whether the household goes to the store are described by the underlying latent variable, v^*_{it}. Then, let S_{it} be the positive and continuous variable equal to the total weekly purchases made by each household i in week t. The factors influencing S_{it} are captured by the latent variable S^*_{it} when household i comes to the store and $v_{it} = 1$.

The variables affecting household purchase behavior and coupon receipt are listed in Table 3.20. Importantly, receipt of reward and promotion campaigns affect both the likelihood of a household coming to the store and their purchases once in the store. The variables used to predict households' receipt of these endogenous campaigns are listed in the third column. I assume mailer coupons similarly affect both the likelihood of a household coming to the store and the purchases conditional on being in the store. Displays, on the other hand, only affect purchases conditional on being in the store because the household cannot see them if they are not in the store.

The theoretical model predicts that marketing tools will have a differential impact on customers' purchase path. In particular, some marketing tools may induce a customer to change the ordering of their store visits in order to evaluate a product. In the case of a (**Different store, same department**) or (**Different store, new department**), we expect to observe an increased average likelihood of customers coming to the store when they receive the campaigns inducing this shift. The model predicts that personalized campaigns are more likely to induce this change in store ordering than generic campaigns because personalized campaigns are more likely to be considered. Changing the ordering of stores also yields a predicted increase in the expected sales of the store which has increased it's ordering because the preferred store receives a bigger share of the customer's shopping list than the second best store. Finally, in the case of a (**Same store, new department**), we expect to observe an increase in overall sales at the store because a customer with an expanded

Table 3.20: Variables Affecting Store Purchase Behavior

Variables	Come to Store, v_{it}^*	Purchase Conditional on Coming to Store , S_{it}^*	Campaign Receipt, $\mathbf{C_{it}}$
$\mathbf{C_{it}}$	Reward Campaign	Reward Campaign	
$\mathbf{C_{it}}$	Promotion Campaign	Promotion Campaign	
$\mathbf{X_{1it}}$	Mailer	Mailer	Mailer
$\mathbf{X_{1it}}$	Season (1×4)	Season (1×4)	Season (1×4)
$\mathbf{W_{it}}$	Loyalty Characteristics (1×11)	Loyalty Characteristics (1×11)	Loyalty Characteristics (1×11)
$\mathbf{X_{2it}}$			
$\mathbf{X_{2it}}$		Reward Redemption	Reward Redemption
$\mathbf{X_{2it}}$		Promotion Redemption	Promotion Redemption
$\mathbf{X_{2it}}$		Retail Discounts	Retail Discounts
$\mathbf{X_{2it}}$		Other Coupon Discounts	Other Coupon Discounts
$\mathbf{X_{2it}}$		Match Discounts	Match Discounts
Q_{it}	L.[% Basket Perishable Goods]- Ave.[% Basket Perishable Goods]		
$\mathbf{Z_{it}}$			Learning from Rewards
$\mathbf{Z_{it}}$			Learning from Promotions
$\mathbf{Z_{it}}$			Match Variables (1×11)
$\mathbf{R_t}$			Indicator for weeks of Reward Campaigns
$\mathbf{R_t}$			Indicator for week of Promotion Campaigns

consideration set will have a greater expected increase in their overall spending at the store. The identification of the specific impact of sales within the new departments is estimated in Subsection 3.6.5

The model for decomposing the effect of the marketing tools at the store level is given by

$$S_{it}^* = \mathbf{W}_{it}\boldsymbol{\beta}_1 + \mathbf{X}_{1it}\boldsymbol{\beta}_2 + \mathbf{X}_{2it}\boldsymbol{\beta}_3 + \mathbf{C}_{it}\boldsymbol{\beta}_4 + u_{it} \equiv \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} + u_{it} \tag{3.6}$$

$$v_{it}^* = \mathbf{W}_{it}\boldsymbol{\gamma}_1 + \mathbf{X}_{1it}\boldsymbol{\gamma}_2 + \mathbf{C}_{it}\boldsymbol{\gamma}_3 + Q_{it}\gamma_4 + \mu_{it} \equiv \tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \mu_{it} \tag{3.7}$$

$$\mathbf{C}_{it} = \mathbf{R_t} \cdot [\mathbf{W}_{it}, \mathbf{Z}_{it}]\boldsymbol{\Gamma}_1 + [\mathbf{X}_{1it}, \mathbf{X}_{2it}]\boldsymbol{\Gamma}_2 + \boldsymbol{\epsilon}_{it} \equiv \tilde{\mathbf{Z}}_{it}\boldsymbol{\Gamma} + \boldsymbol{\epsilon}_{it} \tag{3.8}$$

$$v_{it} = 1(v_{it}^* > 0) \tag{3.9}$$

$$S_{it} = S_{it}^* v_{it} \tag{3.10}$$

where $i = 1, ..., n$ indexes households and $t = 1, ..., T$ indicates weeks during the sample. The first equation (weekly sales) is the main equation, where the latent dependent-variable S_{it}^* is related to $\mathbf{W}_{it}$, a (1×11) vector of exogenous variables of loyalty characteristics, $\mathbf{X}_{1it}$, a (1×5)-vector of exogenous explanatory variables including whether a household received a store-level mailer in the week and variables indicating the season, $\mathbf{X}_{2it}$, a (1×6) vector of exogenous variables including whether there are products on display at the store, and whether the household redeemed reward coupons or promotion coupons, and whether the household received retail discounts, other coupon discounts, or match discounts, and to $\mathbf{C}_{it}$, a (1×2)-vector of the endogenous variables, receipt of reward and promotion campaigns, in both the main and selection equations. I assume that no households come to the store without buying because of fixed cost associated with coming to the store.

The second equation is the selection equation, where the latent variable v_{it}^* is related to $\mathbf{W}_{it}$, to $\mathbf{C}_{it}$ and to Q_{it},[4] an exogenous variable which only appears in the selection equation. Here Q_{it} captures the difference between the percent of the previous visit was perishable goods versus the average percent of a household's basket is perishable goods. If a household wasn't able to buy a lot of perishables relative to their average amount, they are more likely to come back sooner. The average

[4]Estimates are robust to exclusion of this variable.

helps to weigh this variable more heavily toward those who tend to purchase more perishables. Note that Q_{it} in the selection equation satisfies the exclusion restriction, which is desirable to include to avoid multicolliniarity problems (Cameron and Trivedi 16.5).

In Equation (3.8), it is assumed that the endogenous variables, the reward and promotion campaigns, can be explained by exogenous instrumental variables $\mathbf{Z}_{it}(1 \times 13)$, which include learning variables (how much households changed their spending during previous campaigns) and match variables (what percent of households' baskets included products in the campaign of interest). Additionally, loyalty characteristics, $\mathbf{W}_{it}$, are important variables which helps predict the receipt of the targeted coupon campaigns. Finally, I interact each term by an indicator variable, $\mathbf{R_t}$, which equals unity if a reward or promotion campaign is running. I assume this variable is exogenous to the model in that the manufacturer and retailers need to negotiate the timing and content of campaigns prior to determining which households will participate in which campaigns. I explain more about these variables in Subsection 3.6.1, but these instrumental variables satisfy the exclusion restrictions for control functions. Equations (3.6), (3.7), (3.9), and (3.10) represent the Type 2 Tobit sample selection framework without endogeneity. Including equation (3.8) allows for endogenous variables in the main and selection equations that are correlated with the error terms u_{it} and μ_{it}. I categorize each of the variables according to whether they are exogenous or endogenous in Table (3.21).

For each household i and week t, u_{it}, μ_{it}, and $\boldsymbol{\epsilon_{it}}$ are independent of $\tilde{\mathbf{X}}_{1it}$, $\tilde{\mathbf{X}}_{2it}$, $\tilde{\mathbf{W}}_{it}$, and $\tilde{\mathbf{Z}}_{it}$. It is assumed that the vector of error terms $(u_{it}, \mu_{it}, \boldsymbol{\epsilon_{it}})'$ is distributed jointly normal according to

Table 3.21: List of Variables

Exogenous	Endogenous
$\mathbf{X}_{it}$	S_{it}^*
$\mathbf{W}_{it}$	v_{it}^*
Q_{it}	$\mathbf{C}_{it}$
$\mathbf{Z}_{it}$	
$\mathbf{R_t}$	

$$\begin{pmatrix} u_{it} \\ \mu_{it} \\ \boldsymbol{\epsilon_{it}} \end{pmatrix} \sim N \left(\mathbf{0}, \begin{bmatrix} \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_\mu \\ \rho\sigma_u\sigma_\mu & \sigma_\mu^2 \end{pmatrix} & \mathbf{\Omega}' \\ \mathbf{\Omega}_{(2\times 2)} & \mathbf{\Sigma}_{(2\times 2)} \end{bmatrix} \right) \tag{3.11}$$

The covariance matrix of errors consists of four parts. The upper left part captures the covariance and variance for the errors of the main and selection equations, respectively, where σ_u^2 and σ_μ^2 denote the variances of u_{it} and μ_{it}, and ρ denotes the correlation coefficient. This error structure of this part of the matrix is classic to the Type 2 Tobit model and is also used in the Heckman selection model. Without endogeneity, estimation would be solely based on this covariance matrix. The potential presence of endogeneity is accounted for by the (2×2)-matrix $\mathbf{\Omega}$, which captures the influence of unobserved factors which jointly affect the dependent variables in equations (3.6) and (3.7) and the endogenous explanatory variables. This implies that there is no endogeneity if and only if $\mathbf{\Omega}$ is equal to the null matrix. Finally, the error terms for the endogenous explanatory variance have covariance matrix $\mathbf{\Sigma}$. I assume that the distribution of $\boldsymbol{\epsilon_{it}}$ is normal, which is best suited for continuous endogenous variables, $\mathbf{C}_{it}$. While the receipt of reward and promotion coupons is discrete, for simplicity, I use linear probability to estimate these endogenous variables.

3.6.4 Estimation, Interpretation, and Testing for Endogeneity

In order to estimated the parameters from equations (3.6)-(3.10), I use the limited information maximum likelihood (LIML) method first introduced by Smith and Blundell (1986) and Rivers and Vuong (1988). Also known as the control function method, the LIML approach provides the flexibility needed to account for endogeneity in both the main and selection equations. Schwiebert (2015) is the first to derive the Maximum Likelihood estimation for the Type 2 Tobit model with endogenous covariates in both the main and selection equations. The procedure follows the standard two step procedure for estimation: first estimate the reduced form Equation (3.8) by OLS and obtain the residuals $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{it}}$ and second insert these residuals into the log-likelihood function. Schwiebert (2015) also derives the full information maximum likelihood (FIML) for the same model. This model is fully efficient, but when the number of observations and or covariates is large (as is the case in my model), this method is quite time-consuming. Given this constraint, I use the LIML approach. I will derive the log-likelihood function below.

In order to build the likelihood function which accounts for the effect of the endogenous variable errors, we must find the conditional distribution of $(u_{it}, \mu_{it})'$ given $\boldsymbol{\epsilon}_{\boldsymbol{it}} = \boldsymbol{e}$. The multivariate normal conditional distribution is given by

$$\begin{bmatrix} u_{it} \\ \mu_{it} \end{bmatrix} \Bigg| \, [\boldsymbol{\epsilon}_{\boldsymbol{it}} = \boldsymbol{e}] \sim N\left(\boldsymbol{\Omega}'\boldsymbol{\Sigma}^{-1}[\boldsymbol{\epsilon}_{\boldsymbol{it}} = \boldsymbol{e}]', \mathbf{B}\right) \tag{3.12}$$

where

$$\mathbf{B} \equiv \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \equiv \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_\mu \\ \rho\sigma_u\sigma_\mu & \sigma_\mu^2 \end{pmatrix} - \boldsymbol{\Omega}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega} \tag{3.13}$$

Since the scale of the dependent variable is not observed with Probit estimation, we cannot separately identify the coefficients and each element of the covariance matrix. Therefore, without loss of generality, we normalize B_{22} to 1 as is standard for Heckman estimation.[5] We can then define the normalized conditional distribution as

$$\boldsymbol{\Gamma} \equiv \begin{pmatrix} \tilde{\sigma}^2 & \tilde{\rho}\tilde{\sigma} \\ \tilde{\rho}\tilde{\sigma} & 1 \end{pmatrix} \equiv \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_\mu \\ \rho\sigma_u\sigma_\mu & \sigma_\mu^2 \end{pmatrix} - \boldsymbol{\Omega}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega} \tag{3.14}$$

To simplify notation, define

$$\boldsymbol{\Psi} \equiv \begin{pmatrix} \boldsymbol{\psi}_{11} \\ (1 \times 2) \\ \boldsymbol{\psi}_{21} \\ (1 \times 2) \end{pmatrix} \equiv \boldsymbol{\Omega}'\boldsymbol{\Sigma}^{-1} \tag{3.15}$$

Therefore equation (3.12) can be rewritten as

$$\begin{bmatrix} u_{it} \\ \mu_{it} \end{bmatrix} \Bigg| [\boldsymbol{\epsilon_{it}} = \boldsymbol{e}] \sim N\left(\begin{bmatrix} \boldsymbol{\psi}_{11}[\boldsymbol{\epsilon_{it}} = \boldsymbol{e}]' \\ \boldsymbol{\psi}_{21}[\boldsymbol{\epsilon_{it}} = \boldsymbol{e}]' \end{bmatrix}, \begin{pmatrix} \tilde{\sigma}^2 & \tilde{\rho}\tilde{\sigma} \\ \tilde{\rho}\tilde{\sigma} & 1 \end{pmatrix} \right) \tag{3.16}$$

which resembles the (unconditional) joint error distribution of the sample selection model without endogeneity (except for the nonzero means). Finding the conditional distribution allows us to rewrite the main and selection equations as

$$S_{it}^*|\boldsymbol{\epsilon_{it}} = \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} + \boldsymbol{\psi}_{11}\boldsymbol{\epsilon_{it}}' + \boldsymbol{\xi_{1it}} \tag{3.17}$$

$$v_{it}^*|\boldsymbol{\epsilon_{it}} = \tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\boldsymbol{\epsilon_{it}}' + \boldsymbol{\xi_{2it}} \tag{3.18}$$

Where

$$\boldsymbol{\xi_{it}}|\boldsymbol{\epsilon_{it}} \sim N\left(0, \begin{pmatrix} \tilde{\sigma}^2 & \tilde{\rho}\tilde{\sigma} \\ \tilde{\rho}\tilde{\sigma} & 1 \end{pmatrix}\right) \tag{3.19}$$

[5] Note that with this normalization, we take into account two levels by which the final coefficients otherwise would need to be scaled. The first is with respect to the variance of μ, σ_μ. This variable is typically normalized to one in probit estimation. The second is with respect to the correlation between μ and $\boldsymbol{\epsilon_{it}}$, captured by $\boldsymbol{\Omega}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}$. In Rivers and Vuong (1988) and Wooldridge (2002) 15.7.2, we see that control function probit estimates are scaled by this correlation value. By normalizing $B_{22} = 1$, we account for both of these scaled parameters in our final estimation.

The probability of household i not coming to the store in week t conditional on $\tilde{\mathbf{X}}_{it}, \tilde{\mathbf{W}}$, and $\boldsymbol{\epsilon_{it}}$ can be written as

$$\begin{aligned} P(v_{it}^* \leq 0 | \boldsymbol{\epsilon_{it}}) &= P(\boldsymbol{\xi_{2it}} \leq -\tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} - \boldsymbol{\psi}_{21}\boldsymbol{\epsilon_{it}}') \\ &= 1 - \Phi(\tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\boldsymbol{\epsilon_{it}}') \end{aligned}$$

Additionally, using Amemiya (1985, pp. 385-386), the probability of household i coming to the store and spending $S_{it}^* = S_{it}$ with $v_{it}^* > 0$ in week t conditional on $\tilde{\mathbf{X}}_{it}, \tilde{\mathbf{W}}$, and $\boldsymbol{\epsilon_{it}}$ can be written as

$$\begin{aligned} & f(S_{it} | v_{it}^* > 0 | \boldsymbol{\epsilon_{it}}) \cdot P(v_{it}^* > 0 | \boldsymbol{\epsilon_{it}}) \\ &= \int_0^\infty f(v_{it}^*, S_{it} | \boldsymbol{\epsilon_{it}}) dv_{it}^* \\ &= \int_0^\infty f(v_{it}^* | S_{it}, \boldsymbol{\epsilon_{it}}) \cdot f(S_{it} | \boldsymbol{\epsilon_{it}}) dv_{it}^* \\ &= \int_0^\infty f(v_{it}^* | S_{it}, \boldsymbol{\epsilon_{it}}) dv_{it}^* \cdot f(S_{it} | \boldsymbol{\epsilon_{it}}) \\ &= \Phi\left(\frac{\tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\boldsymbol{\epsilon_{it}}' + \frac{\tilde{\rho}}{\tilde{\sigma}}(S_{it} - \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} - \boldsymbol{\psi}_{11}\boldsymbol{\epsilon_{it}}')}{\sqrt{1 - \tilde{\rho}^2}}\right) \cdot \frac{1}{\tilde{\sigma}}\phi\left(\frac{S_{it} - \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} - \boldsymbol{\psi}_{11}\boldsymbol{\epsilon_{it}}'}{\tilde{\sigma}}\right) \end{aligned}$$

In the first line, $f(S_{it} | v_{it}^* > 0, \tilde{\mathbf{X}}_{it}, \tilde{\mathbf{W}}_{it}, \boldsymbol{\epsilon_{it}})$ is the conditional density of S_{it} given $v_{it}^* > 0$. Given that we are conditioning this all on the case when a household comes to the store, $v_{it}^* > 0$, and since the errors are distributed joint normally, we can rewrite the first line to be the joint density, $f(v_{it}^*, S_{it} | \tilde{\mathbf{X}}_{it}, \tilde{\mathbf{W}}_{it}, \boldsymbol{\epsilon_{it}})$, where we integrate observations where $v_{it}^* > 0$ in the second line. Then, in the third line, since the joint density is the product of the conditional density and the marginal density, we can replace the joint density with $f(v_{it}^*, S_{it} | \tilde{\mathbf{X}}_{it}, \tilde{\mathbf{W}}_{it}, \boldsymbol{\epsilon_{it}}) = f(v_{it}^* | S_{it}, \tilde{\mathbf{X}}_{it}, \tilde{\mathbf{W}}_{it}, \boldsymbol{\epsilon_{it}}) \cdot f(S_{it} | \tilde{\mathbf{X}}_{it}, \tilde{\mathbf{W}}_{it}, \boldsymbol{\epsilon_{it}})$. Since $f(S_{it} | \tilde{\mathbf{X}}_{it}, \tilde{\mathbf{W}}_{it}, \boldsymbol{\epsilon_{it}})$ is not dependent on v_{it}^*, we can remove it from the integral.

The conditional distribution $f(v_{it}^*|S_{it}, \tilde{\mathbf{X}}_{it}, \tilde{\mathbf{W}}_{it}, \boldsymbol{\epsilon_{it}})$, is distributed normal with mean $\tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\boldsymbol{\epsilon_{it}}' + \frac{\tilde{\rho}}{\tilde{\sigma}}(S_{it} - \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} - \boldsymbol{\psi}_{11}\boldsymbol{\epsilon_{it}}')$ and variance equal to $(1 - \tilde{\rho}^2)$. The mean and variance is based on well-know conditional normal density formulas. Thus, in the last line, we can write the conditional cumulative distribution function for $f(v_{it}^*|S_{it}, \tilde{\mathbf{X}}_{it}, \tilde{\mathbf{W}}_{it}, \boldsymbol{\epsilon_{it}})$ within $\Phi(\cdot)$ and the probability density function for $f(S_{it}|\tilde{\mathbf{X}}_{it}, \tilde{\mathbf{W}}_{it}, \boldsymbol{\epsilon_{it}})$ within $\phi(\cdot)$.

The likelihood of the model conditional on $\boldsymbol{\epsilon_{it}}$ is equal to

$$\boldsymbol{L}_{it}(\boldsymbol{\theta}) = \prod_i \prod_t P(v_{it}^* \leq 0|\boldsymbol{\epsilon_{it}})^{1-v_{it}}[f(S_{it}|v_{it}^* > 0|\boldsymbol{\epsilon_{it}}) \cdot P(v_{it}^* > 0|\boldsymbol{\epsilon_{it}})]^{v_{it}} \tag{3.20}$$

Thus, the log likelihood function conditional on $\boldsymbol{\epsilon_{it}}$ can be written as

$$\begin{aligned}\boldsymbol{l}_{it}(\boldsymbol{\theta}) = \sum_i \sum_t &(1 - v_{it})[\log(1 - \Phi(\tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\boldsymbol{\epsilon_{it}}'))] \\ &+ (v_{it})[\log(\Phi((1-\tilde{\rho}^2)^{-1/2}[\tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\boldsymbol{\epsilon_{it}}' + \tilde{\rho}\tilde{\sigma}^{-1}(S_{it} - \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} - \boldsymbol{\psi}_{11}\boldsymbol{\epsilon_{it}}')]) \\ &+ \phi(\tilde{\sigma}^{-1}(S_{it} - \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} - \boldsymbol{\psi}_{11}\boldsymbol{\epsilon_{it}}'))) - \log\tilde{\sigma}],\end{aligned} \tag{3.21}$$

Inserting the estimated residuals for $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{it}}$ from the Equation (3.5) as covariates, the log likelihood function becomes

$$\begin{aligned}\boldsymbol{l}_{it}(\boldsymbol{\theta}) = \sum_i \sum_t &(1 - v_{it})[\log 1 - \Phi(\tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{it}}')] \\ &+ (v_{it})[\log\Phi((1-\tilde{\rho}^2)^{-1/2}[\tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{it}}' + \tilde{\rho}\tilde{\sigma}^{-1}(S_{it} - \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} - \boldsymbol{\psi}_{11}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{it}}')]) \\ &+ \phi(\tilde{\sigma}^{-1}(S_{it} - \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} - \boldsymbol{\psi}_{11}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{it}}')) - \log\tilde{\sigma}],\end{aligned} \tag{3.22}$$

and is then maximized over $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}', \boldsymbol{\gamma}', \tilde{\rho}, \tilde{\sigma}, \boldsymbol{\psi}_{11}, \boldsymbol{\psi}_{21})'$.

Note that the log-likelihood function is the same as for the Type 2 Tobit selection model without endogenous covariates, with additional covariates $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{it}}$ (see Amemiya, 1985).

3.6.5 Department Sales

In the grocery store setting, customers face a number options. In fact, we argue that there are too many options for her to fully evaluate. Instead, she must choose a purchase path within the store. The purchase path is the walking path across different departments within the store. Departments can include Produce, Meat, Dairy, Drug / General Medicine, and other center aisles. I fully expand on this list in the Data Section (3.5).

As outlined in the theoretical model, I assume that the household has a *status quo* path which she automatically takes to buy her groceries in the store. However, marketing tools employed by the retailer can induce her to change her path within the store (**Same store, new department**), where marketing tools include the reward and promotion campaigns, mailers, and displays. The estimation goal in this section is to test the impact of the various marketing tools on changing the purchase path of customers.

The theoretical prediction is that if a marketing tool is considered and advertises a product outside of the customer's status quo purchase path, it will have a greater expected increase on the customer's consideration set and thus the customer's expected basket size. In this section, we empirically test these predictions by evaluating the differential impact of the various marketing tools on customers' average spending in the departments of advertised products. We specifically evaluate the department sales impact of promotion campaign, reward campaigns, mailers, and displays. We also identify the impact of localized marketing tools like retail discounts, other coupon discounts, and match discounts.

One limitation in the dataset is that I only observe reward products sent at the campaign level instead of at the individual level. Because individuals receiving reward campaigns are sent individualized coupons, this limits my ability to identify the specific impact of reward campaigns at the department level. However, from what we know of the reward campaigns, we presume they have little effect on changing the direction of the purchase path since they offer discounts on products the household currently purchases. Additionally, from results at the store level we find that there is little overall effect of the reward coupons in changing purchase behavior. The other marketing tools offer much more information which can be used to expand the consideration set for the consumers, and we focus on the impact of these other tools in this section.

When a customer decides to change their path either because of a display or in search of a product in the promotion campaign or mailer, my model predicts that she considers products along her new path. As she walks through the new path, she evaluates the products according to her linear ordering of preferences. The fact that she is on this new path opens her up to more opportunities to spend in the department of the product being advertised, increasing her likelihood of purchasing within that department. I call this change in sale the spillover effect of the advertised product. I am the first to estimate this spillover effect in the context of the departments in the store.

The model for decomposing the effect of the marketing tools at the store level is given by

Table 3.22: Variables Affecting Department Purchase Behavior

Variables	Come to Store, v_{it}^*	Purchase Conditional on Coming to Store , D_{it}^*	Department Promotional Campaign Receipt, $\mathbf{C_{idt}}$
$\mathbf{C_{it}}$	Reward Campaign		
$\mathbf{C_{it}}$	Promotion Campaign		
$\mathbf{C_{dit}}$	Reward Campaign	Reward Campaign	
$\mathbf{C_{dit}}$		Dept Promotion Campaign	
$\mathbf{X_{1it}}$	Season (1×4)		Season (1×4)
$\mathbf{X_{1it}}$	Mailer		Mailer
$\mathbf{X_{1dit}}$		Dept Mailer	Dept Mailer
$\mathbf{X_{1dit}}$		Season (1×4)	
$\mathbf{W_{it}}$	Loyalty Characteristics (1×11)	Loyalty Characteristics (1×11)	Loyalty Characteristics (1×11)
$\mathbf{X_{2dit}}$		Dept Display	Dept Display
$\mathbf{X_{2dit}}$		Dept Reward Redemption	Dept Reward Redemption
$\mathbf{X_{2dit}}$		Dept Promotion Redemption	Dept Promotion Redemption
$\mathbf{X_{2dit}}$		Dept Retail Discounts	Dept Retail Discounts
$\mathbf{X_{2dit}}$		Dept Other Coupon Discounts	Dept Other Coupon Discounts
$\mathbf{X_{2dit}}$		Dept Match Discounts	Dept Match Discounts
Q_{it}	L.[% Basket Perishable Goods]- Ave.[% Basket Perishable Goods]		
$\mathbf{Z_{it}}$			Learning from Rewards
$\mathbf{Z_{it}}$			Learning from Promotions
$\mathbf{Z_{it}}$			Match Variables (1×11)
$\mathbf{R_t}$			Indicator for weeks of Reward Campaigns
$\mathbf{R_t}$			Indicator for week of Promotion Campaigns
R_{dt}			Indicator for week of Promotion Campaigns in given Dept

$$D^*_{dit} = \mathbf{W}_{it}\boldsymbol{\beta}_1 + \mathbf{X}_{1dit}\boldsymbol{\beta}_2 + \mathbf{X}_{2dit}\boldsymbol{\beta}_3 + \mathbf{C}_{dit}\boldsymbol{\beta}_4 + u_{dit} \equiv \tilde{\mathbf{X}}_{dit}\boldsymbol{\beta} + u_{it} \tag{3.23}$$

$$v^*_{it} = \mathbf{W}_{it}\boldsymbol{\gamma}_1 + \mathbf{X}_{1it}\boldsymbol{\gamma}_2 + \mathbf{C}_{it}\boldsymbol{\gamma}_3 + Q_{it}\gamma_4 + \mu_{it} \equiv \tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \mu_{it} \tag{3.24}$$

$$\mathbf{C}_{dit} = R_{dt} \cdot [\mathbf{W}_{it}, \mathbf{Z}_{it}]\boldsymbol{\Lambda}_1 + [\mathbf{X}_{1it}, \mathbf{X}_{1dit}, \mathbf{X}_{2dit}]\boldsymbol{\Lambda}_2 + \boldsymbol{\epsilon}_{1it} \equiv \tilde{\mathbf{Z}}_{dit}\boldsymbol{\Gamma} + \boldsymbol{\epsilon}_{1it} \tag{3.25}$$

$$\mathbf{C}_{it} = \mathbf{R_t} \cdot [\mathbf{W}_{it}, \mathbf{Z}_{it}]\boldsymbol{\Gamma}_1 + [\mathbf{X}_{1it}, \mathbf{X}_{2it}]\boldsymbol{\Gamma}_2 + \boldsymbol{\epsilon}_{it} \equiv \tilde{\mathbf{Z}}_{it}\boldsymbol{\Gamma} + \boldsymbol{\epsilon}_{it} \tag{3.26}$$

$$v_{it} = 1(v^*_{it} > 0) \tag{3.27}$$

$$D_{it} = D^*_{dit}v_{it} \tag{3.28}$$

where $i = 1, ..., n$ indexes households and $t = 1, ..., T$ indicates weeks during the sample. The first equation (department weekly sales) is the main equation of interest in this section, where the latent dependent-variable D^*_{dit} is related to $\mathbf{W}_{it}$, a (1×11) vector of exogenous variables of loyalty characteristics, $\mathbf{X}_{1dit}$, a (1×5)-vector of exogenous explanatory variables including whether a household received a mailer highlighting products in the relevant department in the week and variables indicating the season, $\mathbf{X}_{2dit}$, a (1×6) vector of exogenous variables including whether there are products on display in a department, and whether the household redeemed reward coupons or promotion coupons in the department, and whether the household received retail discounts, other coupon discounts, or match discounts in the department, and to $\mathbf{C}_{dit}$, a (1×2)-vector of the endogenous variables, receipt of reward at the store-level and promotion campaigns at the department-level. Since I am unable to observe the department-level products for the reward campaign, I need to use a store-level variable indicating the receipt of a reward campaign at the store-level.

The second equation, 3.24, is the selection equation which is equal to Equation 3.7 in the store-level analysis. Similarly, the prediction of the store-level endogenous coupon receipt in Equation 3.26 used to predict a household coming to the store is the same as Equation 3.8 above. Equation 3.25 predicts the receipt of a promotional campaign at the department level. It is similarly explained by exogenous instrumental variables $\mathbf{Z}_{it}(1 \times 13)$, which include learning variables (how much households changed their spending during previous campaigns) and match variables (what percent of households' baskets included products in the campaign of interest). As before, loyalty characteristics included in $\mathbf{W}_{it}$ also help to predict the receipt of campaigns. I additionally restrict predict to the weeks in which campaigns are running in specific departments with $\mathbf{R_{dt}}$. For example, if a promotion campaign is running with coupons with discounts in Dairy and Produce, these departments would have unity for this variable during the weeks of the campaign and the other departments would have zero.

As before, the model represents the Type 2 Tobit sample selection framework using control functions and analysis of error structure generally follows from above.

For each household i and week t, u_{it}, μ_{it}, $\boldsymbol{\epsilon_{1it}}$, and $\boldsymbol{\epsilon_{2it}}$ are independent of $\tilde{\mathbf{X}}_{1it}$, $\tilde{\mathbf{X}}_{1dit}$, $\tilde{\mathbf{X}}_{2dit}$ $\tilde{\mathbf{W}}_{it}$, and $\tilde{\mathbf{Z}}_{it}$. It is assumed that the vector of error terms $(u_{it}, \mu_{it}, \boldsymbol{\epsilon_{1it}}, \boldsymbol{\epsilon_{2it}})'$ is distributed jointly normal according to

$$\begin{pmatrix} u_{it} \\ \mu_{it} \\ \boldsymbol{\epsilon_{1it}} \\ \boldsymbol{\epsilon_{2it}} \end{pmatrix} \sim N \left(\mathbf{0}, \begin{bmatrix} \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_\mu \\ \rho\sigma_u\sigma_\mu & \sigma_\mu^2 \end{pmatrix} & \boldsymbol{\Omega}' \\ \boldsymbol{\Omega}_{(4\times 2)} & \boldsymbol{\Sigma}_{(4\times 4)}. \end{bmatrix} \right) \tag{3.29}$$

We follow the same procedure as in the store-level, accounting for the additional errors, $\boldsymbol{\epsilon_{2it}}$. From Equation 3.12, we now have

$$\begin{bmatrix} u_{it} \\ \mu_{it} \end{bmatrix} \Bigg| [\boldsymbol{\epsilon_{1it}} = \boldsymbol{e_1}, \boldsymbol{\epsilon_{2it}} = \boldsymbol{e_2}] \sim N \left(\boldsymbol{\Omega}'\boldsymbol{\Sigma}^{-1}[\boldsymbol{\epsilon_{1it}} = \boldsymbol{e_1}, \boldsymbol{\epsilon_{2it}} = \boldsymbol{e_2}]', \mathbf{B} \right) \tag{3.30}$$

where $\mathbf{B}$ equals the same equation as in Equation 3.13 and $\boldsymbol{\Gamma}$ equals the same as in Equation 3.14. Now redefining Equation 3.15 to equal

$$\boldsymbol{\Psi} \equiv \begin{pmatrix} \underset{(1\times 2)}{\boldsymbol{\psi}_{11}} & \underset{(1\times 2)}{\boldsymbol{\psi}_{12}} \\ \underset{(1\times 2)}{\boldsymbol{\psi}_{21}} & \underset{(1\times 2)}{\boldsymbol{\psi}_{22}} \end{pmatrix} \equiv \boldsymbol{\Omega}'\boldsymbol{\Sigma}^{-1} \tag{3.31}$$

gives us

$$\begin{bmatrix} u_{it} \\ \mu_{it} \end{bmatrix} \Bigg| [\boldsymbol{\epsilon_{1it}} = \boldsymbol{e_1}, \boldsymbol{\epsilon_{2it}} = \boldsymbol{e_2}] \sim N\left(\begin{bmatrix} \boldsymbol{\psi}_{11}[\boldsymbol{\epsilon_{1it}} = \boldsymbol{e_1}]' + \boldsymbol{\psi}_{12}[\boldsymbol{\epsilon_{2it}} = \boldsymbol{e_2}]' \\ \boldsymbol{\psi}_{21}[\boldsymbol{\epsilon_{1it}} = \boldsymbol{e_1}]' + \boldsymbol{\psi}_{22}[\boldsymbol{\epsilon_{2it}} = \boldsymbol{e_2}]' \end{bmatrix}, \begin{pmatrix} \tilde{\sigma}^2 & \tilde{\rho}\tilde{\sigma} \\ \tilde{\rho}\tilde{\sigma} & 1 \end{pmatrix} \right) \tag{3.32}$$

Therefore, we can rewrite the main and selection equations for the department sales to equal

$$D^*_{dit}|\boldsymbol{\epsilon_{1it}}, \boldsymbol{\epsilon_{2it}} = \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} + \boldsymbol{\psi}_{11}\boldsymbol{\epsilon_{1it}}' + \boldsymbol{\psi}_{12}\boldsymbol{\epsilon_{2it}}' + \boldsymbol{\xi_{1it}} \tag{3.33}$$

$$v^*_{it}|\boldsymbol{\epsilon_{1it}}, \boldsymbol{\epsilon_{2it}} = \tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\boldsymbol{\epsilon_{1it}}' + \boldsymbol{\psi}_{22}\boldsymbol{\epsilon_{2it}}' + \boldsymbol{\xi_{2it}} \tag{3.34}$$

As in the store sales analysis, I first estimate Equations 3.25 and 3.26 to obtain the estimated residuals, $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{1it}}$ and $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{2it}}$ and then insert them into Equations 3.33 and 3.34 as covariates. However, this time, I estimate the covariates, $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}', \boldsymbol{\gamma}', \tilde{\rho}, \tilde{\sigma}, \boldsymbol{\psi}_{11}, \boldsymbol{\psi}_{12}, \boldsymbol{\psi}_{21}, \boldsymbol{\psi}_{22})'$ using Heckman's two-step procedure. While the LIML approach is more efficient, the two-step Heckman correction approach has practical advantages especially when estimating over a large dataset with many covariates (see Cameron and Trivedi, 2005, Section 16.5.4). In particular, it is much easier to get convergence of estimates using the two-step Heckman procedure than with maximum likelihood estimation. One thing to note in the two-step Heckman correction procedure is that I am implicitly assuming that the covariance across the departments is zero. This may be another reason why the LIML approach at the department level fails to converge.

In order to do this, I use Amemiya (1985, pg. 386) to adapt the Type 2 Tobit model to generalize the Heckman selection correction approach. In order to get something similar to the standard Heckman model of $y_i = x_i'\beta + \lambda(x_i'\beta) + \epsilon_i$ for i such that $y_i > 0$, where $\epsilon_i = y_i - E(y_i|y_i > 0)$, we need to evaluate $E(D^*_{dit}|v^*_{it} > 0)$.

In order to derive this expression, we use

$$D^*_{dit}|\boldsymbol{\epsilon_{1it}}, \boldsymbol{\epsilon_{2it}} = \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} + \boldsymbol{\psi}_{11}\boldsymbol{\epsilon_{1it}}' + \boldsymbol{\psi}_{12}\boldsymbol{\epsilon_{2it}}' + \frac{\tilde{\rho}}{\tilde{\sigma}}(v^*_{it} - \tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} - \boldsymbol{\psi}_{21}\boldsymbol{\epsilon_{1it}}' - \boldsymbol{\psi}_{22}\boldsymbol{\epsilon_{2it}}') + \zeta_{it} \tag{3.35}$$

where $\zeta_{it} \sim N(0, 1 - \tilde{\rho}^2)$ and is independent of v^*_{it}. Therefore, we can write

$$D^*_{dit}|\boldsymbol{\epsilon_{1it}}, \boldsymbol{\epsilon_{2it}} = \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} + \boldsymbol{\psi}_{11}\boldsymbol{\epsilon_{1it}}' + \boldsymbol{\psi}_{12}\boldsymbol{\epsilon_{2it}}' + \frac{\tilde{\rho}}{\tilde{\sigma}}[\lambda(\tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\boldsymbol{\epsilon_{1it}}' + \boldsymbol{\psi}_{22}\boldsymbol{\epsilon_{2it}}')] + \iota_{it} \tag{3.36}$$

where $\lambda(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}$ is the inverse mills ratio. Finally, using the residual estimates from Equations 3.25 and 3.26, we can consistently estimate the two-stage Heckman selection correction equation

$$D^*_{dit}|\boldsymbol{\hat{\epsilon}_{1it}}, \boldsymbol{\hat{\epsilon}_{2it}} = \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} + \boldsymbol{\psi}_{11}\boldsymbol{\hat{\epsilon}_{1it}}' + \boldsymbol{\psi}_{12}\boldsymbol{\hat{\epsilon}_{2it}}' + \frac{\tilde{\rho}}{\tilde{\sigma}}[\lambda(\tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\boldsymbol{\hat{\epsilon}_{1it}}' + \boldsymbol{\psi}_{22}\boldsymbol{\hat{\epsilon}_{2it}}')] + \iota_{it} \tag{3.37}$$

I estimate the coefficients in the department sales model according to the following steps

- Estimate endogenous Equations 3.25 and 3.26. Obtain estimated residuals, $\boldsymbol{\hat{\epsilon}_{1it}}$ and $\boldsymbol{\hat{\epsilon}_{2it}}$.
- Estimate two-stage Heckman correction, where the first stage probit estimation yields estimates for $\boldsymbol{\hat{\gamma}}, \boldsymbol{\hat{\psi}}_{21}, \boldsymbol{\hat{\psi}}_{22}$.
- Estimate Equation 3.37 using the estimates from the first stage.

3.6.6 Bootstrap Algorithm

Since the second stage of the LIML method imputes estimated unobservable regressors from the first stage, we must adjust the asymptotic variance to account for the fact that these estimated residuals are not randomly sampled from the population. It would be ideal to analytically derive the asymptotic distribution through methods such as Murphy and Topel (1985) and Newey and McFadden (1994). However, adjustment in this setting is complicated by the fact that we have two separate endogenous variables, *promotion* and *reward* coupons which cannot be predicted with a single maximum likelihood estimator. While it is certainly possible to consistently estimate the main equation coefficients using the control function methods outlined above, we need to employ bootstrapping methods to adjust for the imputed estimated regressors. We use the algorithm outlined by Cameron and Trivedi (2005).

- Given data $\boldsymbol{w}_1, ..., \boldsymbol{w}_N$, draw a bootstrap sample of size N and denote this new sample, $\boldsymbol{w}_1^*, ..., \boldsymbol{w}_N^*$.
- Calculate the standard error, $s_{\hat{\theta}^*}$, of the estimate $\hat{\theta}^*$. Here $\hat{\theta}^*$ and $s_{\hat{\theta}^*}$ are calculated in the usual way but using the new bootstrap sample rather than the original sample.
- Repeat the first two steps B independent times, where B is a large number, obtaining B bootstrap replications of the standard error.
- Use these B bootstrap replications to obtain a bootstrapped version of the statistic.

The bootstrap estimate of calculates variance from the B bootstrap replications $\hat{\theta}_1^*, ..., \hat{\theta}_B^*$.

$$s^2_{\hat{\theta},Boot} = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^*_b - \bar{\hat{\theta}}^*)^2, \tag{3.38}$$

where

$$\bar{\hat{\theta}}^* = B^{-1}\sum_{b=1}^{B}\hat{\theta}^*_b. \tag{3.39}$$

Given the panel structure of the data, we must account for the dependence of observations within households across time by clustering at the household level. The cluster bootstrap samples the clusters with replacement such that if there are C clusters, then the bootstrap resample has C clusters. This may mean that the number of observations $N = \sum_{c=1}^{C} N_c$ may vary across bootstrap resamples, but this does not pose a problem. Consistency of the bootstrap estimate of the standard error of $\hat{\theta}$ depends on the smoothness of the distribution of the data and variance, consistency of the estimation for the empirical distribution of the data, and independence of the clusters.

3.7 Results

In this section we provide the estimation results for Equations 3.1, 3.5, 3.22, 3.25, and 3.37.

3.7.1 Predicting Coupon Receipt

In order to account for the endogeneity of the receipt of the targeted coupon campaigns, we must predict the likelihood of a household receiving these campaigns. Recall that the model for receipt of the campaigns of each type can be summarized by 3.1:

$$\mathbf{C}_{it} = \mathbf{R_t} \cdot [\mathbf{W}_{it}, \mathbf{Z}_{it}]\mathbf{\Gamma}_1 + [\mathbf{X}_{1it}, \mathbf{X}_{2it}]\mathbf{\Gamma}_2 + \boldsymbol{\epsilon}_{it} \equiv \tilde{\mathbf{Z}}_{it}\mathbf{\Gamma} + \boldsymbol{\epsilon}_{it}$$

where $\mathbf{C}_{it}$ is a (1×2) vector for the receipt of reward campaigns, R_{it} and promotion campaigns, P_{it}. For each campaign type, $\mathbf{R_t}$ is a (1×2) vector which captures whether a campaign of each type is running or not in a given week, $\mathbf{W}_{it}$ is a (1×11) vector of household-specific loyalty characteristics, $\mathbf{Z}_{it}$ is a (1×13) vector of learning and match variables, and $\mathbf{X}_{1it}$, and $\mathbf{X}_{2it}$ are the exogenous variables from the main estimation equations.

Recall that the stated intention of reward campaigns is to build loyalty for the store's best customers by giving discounts to customers that are most loyal, therefore we expect to see positive coefficients on the variables indicating that households spend more and come more frequently on average. Promotion campaigns do not have the same stated goal and instead are intended to expand the consideration set of households by advertising coupons relevant to them. Note that in order to avoid multicollinearity we exclude the lowest levels of loyalty, Low Sales ($\leq$ \$13), and Infrequent Visits ($\geq$ 35 days), so each coefficient reflects the difference between the variable in question and the respective excluded dummy variable.

Additionally, since we expect the retailer to learn from household responsiveness to previous campaigns and favor households that increase their spending during past campaigns, we expect the coefficient on the learning variables, "Δ Spend in L.Reward Camp" and "Δ Spend in L.Promotion Camp" to be non-negative.

Finally, we expect a level of matching of household previous purchases with the products in the campaigns. One caveat here is that I observe the product list for each campaign at the campaign level rather than at the household level. While this is not a problem for promotion campaigns since each household in promotion campaign receives the same set of products, this is a hindrance to my prediction for reward campaigns since each household in a reward campaign receives their own individualized set of coupons. Therefore, the matching variables for the reward campaigns are less precise than they would otherwise be if I had access to the complete list of household-specific reward coupons. Since manufacturers are involved in designing the campaigns and part of the payment of the redemption, we expect there to be a positive coefficient on variables relating to the purchase of manufacturer products.

The two equations are simultaneously estimated with linear probability and the estimates are reported in Tables 3.23 and 3.24. Each regression is highly predictive with an $R^2 = 0.9767$ for the reward campaigns and $R^2 = 0.8418$ for the promotion campaigns, indicating that I have identified the main components determining their targeting algorithm. Note that the predictor variables are interacted by the variable, R_t to indicate when campaigns of each type are running, therefore, during weeks in which no campaigns are running, the prediction is forced to zero. Concentrating the estimation to the weeks of the campaigns significantly increases the predictive power.

The results give us more insight into the focus of each campaign type. The reward campaigns are focused more on customers who came to the store more frequently during the control period. Specifically, being a daily customer increased the likelihood of receiving a reward campaign by 3 percent relative to being an infrequent customer and being a twice weekly customer increased likelihood by 2.3 percent relative to being an infrequent customer. Recall that the omitted variables for the frequency dummies is Infrequent customers, so all estimates in this category are in relation to this variable. Customers that came on average weekly and bi-weekly during the control period were about a third less likely to receive reward campaigns than daily customers.

On the other hand, for promotion campaigns we actually find that customers who on average came daily are less likely (-0.4 percent) to receive promotion campaigns than those who come infrequently. All other frequency loyalty characteristics are not significant except for the bi-weekly customers who are slightly more likely to receive promotion campaigns than the infrequent customers.

For average spending loyalty variables, we find the opposite story. In the prediction

Table 3.23: Reward Campaign Predictions

Variable	**Coefficient** (Std. Err.)
Equation 1 : Reward	
Reward Camp Running	-0.001*** (0.000)
(Reward Camp Running) * Daily	0.030*** (0.001)
(Reward Camp Running) * Twice Weekly	0.023*** (0.001)
(Reward Camp Running) * Weekly	0.010*** (0.001)
(Reward Camp Running) * Oneout	0.009*** (0.001)
(Reward Camp Running) * Biweekly	0.001 (0.001)
(Reward Camp Running) * Low-Medium Sales	-0.001* (0.000)
(Reward Camp Running) * Medium Sales	-0.004*** (0.001)
(Reward Camp Running) * Medium-High Sales	-0.003*** (0.001)
(Reward Camp Running) * High Sales	-0.002* (0.001)
(Reward Camp Running) * Sum[% Purch in Dept]	-0.004*** (0.000)
(Reward Camp Running) * Sum[% Purch of Manu]	0.004*** (0.000)
(Reward Camp Running) * Sum[% Purch of Brand]	-0.004*** (0.000)
(Reward Camp Running) * Sum[% Purch of Commodity]	0.012*** (0.000)
(Reward Camp Running) * Sum[% Purch of Sub-commodity]	0.001*** (0.000)
(Reward Camp Running) * Sum[% Purch of Manu-Sub-comm]	-0.012*** (0.000)
(Reward Camp Running) * Sum[% Purch of Man-Comm]	0.007*** (0.000)
(Reward Camp Running) * Sum[% Purch of Man-Dept]	-0.005*** (0.000)
(Reward Camp Running) * Sum[% Purch of Brand-Sub-comm]	0.002*** (0.000)
(Reward Camp Running) * Sum[% Purch of Brand-Comm]	-0.003*** (0.000)
(Reward Camp Running) * Sum[% Purch of Brand-Dept]	0.010*** (0.000)
(Reward Camp Running) * Δ Spend in L.Reward Camp	0.000*** (0.000)
Reward Redemption	0.008*** (0.002)
Promotion Redemption	-0.030*** (0.003)
Retail Discounts	0.000* (0.000)
Other Coupon Discounts	0.000 (0.000)
Match Coupon Discounts	0.000 (0.000)
Mailer	0.000** (0.000)
Season 2	-0.006*** (0.000)
Season 3	-0.005*** (0.000)
Season 4	0.001*** (0.000)
Intercept	0.003*** (0.000)
N	254,490
R-Squared	0.9767

Table 3.24: Promotion Campaign Predictions

Variable	**Coefficient** (Std. Err.)
Equation 2 : Promotions	
Promotion Camp Running	0.005*** (0.001)
(Promotion Camp Running) * Daily	-0.004*** (0.001)
(Promotion Camp Running) * Twice Weekly	-0.002 (0.001)
(Promotion Camp Running) * Weekly	0.002 (0.001)
(Promotion Camp Running) * Oneout	-0.001 (0.001)
(Promotion Camp Running) * Biweekly	0.002** (0.001)
(Promotion Camp Running) * Low-Medium Sales	0.009*** (0.001)
(Promotion Camp Running) * Medium Sales	0.018*** (0.001)
(Promotion Camp Running) * Medium-High Sales	0.015*** (0.001)
(Promotion Camp Running) * High Sales	-0.009*** (0.002)
(Promotion Camp Running) * Sum[% Purch in Dept]	0.008*** (0.000)
(Promotion Camp Running) * Sum[% Purch of Manu]	-0.012*** (0.000)
(Promotion Camp Running) * Sum[% Purch of Brand]	0.013*** (0.000)
(Promotion Camp Running) * Sum[% Purch of Commodity]	-0.004*** (0.000)
(Promotion Camp Running) * Sum[% Purch of Sub-commodity]	0.000*** (0.000)
(Promotion Camp Running) * Sum[% Purch of Manu-Sub-comm]	0.008*** (0.000)
(Promotion Camp Running) * Sum[% Purch of Man-Comm]	-0.006*** (0.000)
(Promotion Camp Running) * Sum[% Purch of Man-Dept]	0.012*** (0.000)
(Promotion Camp Running) * Sum[% Purch of Brand-Sub-comm]	-0.001** (0.000)
(Promotion Camp Running) * Sum[% Purch of Brand-Comm]	-0.002*** (0.000)
(Promotion Camp Running) * Sum[% Purch of Brand-Dept]	-0.015*** (0.000)
(Promotion Camp Running) * Δ Spend in L.Promotion Camp	0.001*** (0.000)
Reward Redemption	-0.004 (0.003)
Promotion Redemption	-0.005 (0.006)
Retail Discounts	0.000 (0.000)
Other Coupon Discounts	0.000 (0.000)
Match Coupon Discounts	0.000 (0.001)
Mailer	0.000 (0.000)
Season 2	0.011*** (0.001)
Season 3	0.005*** (0.001)
Season 4	-0.002*** (0.001)
Intercept	-0.004*** (0.000)
N	254,490
R^2	0.8418

of reward campaigns, medium, medium-high, and high sales customers are actually less likely to receive reward campaigns than the low sales customers. For promotion campaigns, medium and medium-high sales customers are the most likely to receive campaigns relative to low sales, whereas low-medium sales customers are about half as likely (0.9 percent) as the medium sales (1.8 percent) customers to receive promotion campaigns but still more likely to receive the campaigns than low sales customers. High sales customers are actually less likely to receive the campaigns than the low sales customers (-0.9 percent).

While the retailer emphasizes that the reward campaigns are aimed at the most loyal customers, we observe that this is only partly true. The emphasis of the reward campaigns is on customers who come frequently to the store but who actually spend less per week on average. The promotion campaigns are aimed at customers who spend more on average at the store but who don't necessarily come more frequently. In fact, those that come the most frequently are the least likely to receive the promotion campaigns.

We predicted that the retailer learned from households' responsiveness to previous campaigns by prioritizing households that increased their spending in past campaigns. For the reward campaign, the value is positive, yet small. For the promotion campaign, the value is slightly higher at 0.001. Both are statistically significant.

Finally, we predicted that match variables calculated from the match between the customers' control period purchases and products in each campaign would be predictive. Note that we find some interesting patterns here. First, we see that for reward campaigns, a 1 percent increase in the Sum[% Purch of Commodity] increases the likelihood of a customer receiving the likelihood of a customer receiving a reward campaign by 1.2 percent. Recall that the Sum[% Purch of Commodity] is equal to the percent of purchases a household made for each of the commodities in a campaign, summed across the commodities in the campaign (see more description in Subsection

3.6.1). We also see that for reward campaigns, a 1 percent increase in the Sum[% Purch of Man-Comm] increases the likelihood of a customer receiving a reward campaign by 0.7 percent. Interestingly, we see that at the manufacturer sub-commodity level, an increase in 1 percent of the Sum[% Purch of Manu-Sub-comm] actually decreases the likelihood of a customer receiving a reward campaign by 1.2 percent. This implies that at least at the manufacturer level, the reward campaigns are more focused on commodity matches rather than sub-commodity matches. We also observe that an increase in 1 percent of the Sum[% Purch of Brand-Dept] increases the likelihood of a customer getting a reward campaign by 1 percent. Recall that for the reward campaigns, we have less precision in match predictions. We would expect with individual level information on campaign products, we would see even stronger predictions that what we observe.

For promotion campaigns, we observe some interesting dynamics for the predictive match variables. While we observe that a 1 percent increase in the Sum[% Purch of Brand] increases the probability of getting a promotion campaign by 1.3 percent, we see that a 1 percent increase in the Sum[% Purch of Brand-Dept] decreases the likelihood of receiving a promotion campaign by 1.5 percent. This implies that while an increase in the overall purchases of national or private labels at the store level increases the likelihood of getting a promotion campaign, at the department level, the opposite is true. Similarly, for manufacturers, we observe a switch in predictions at the store and department level. While a 1 percent increase in Sum[% Purch of Manu] decreases the likelihood of getting a promotion campaign by 1.2 percent, at the department level, a 1 percent increase in Sum[% Purch of Man-Dept] increases the likelihood of receiving a promotion campaign by 1.2 percent. Therefore, while the retailers decreases the likelihood of sending a promotion when a household is a frequently purchaser of a manufacturer, it increases the likelihood of sending a promotion campaign if the household is a frequent purchaser of a manufacturer in a specific department being promoted. Finally, we observe that a 1 percent increase in Sum[% Purch of Commodity] and Sum[% Purch of Man-Comm] decrease the likelihood of a household receiving a promotion campaign by 0.4 percent and 0.6 percent respectively. This is in line with the promotion campaigns highlighting new products.

Table 3.25 gives the predictions for receipt of the promotion for the department of the promoted products given in Equation 3.25:

$$\mathbf{C}_{dit} = R_{dt} \cdot [\mathbf{W}_{it}, \mathbf{Z}_{it}]\mathbf{\Lambda}_1 + [\mathbf{X}_{1it}, \mathbf{X}_{1dit}, \mathbf{X}_{2dit}]\mathbf{\Lambda}_2 + \boldsymbol{\epsilon}_{1it} \equiv \tilde{\mathbf{Z}}_{dit}\mathbf{\Gamma} + \boldsymbol{\epsilon}_{1it}$$

As in the predictions for the reward and promotion campaigns at the store level, each predictive variable is interacted with the dummy, R_{dt} which identifies the departments with promotion campaigns in the weeks of the campaigns. We find similar patterns for loyalty characteristics in that the promotion campaigns are more focused on households that spend more on average and less on households that come more frequently. Similarly we find a nonnegative effect of the learning variable on receipt of the promotion campaigns.

Perhaps most interesting, we find some variation in the effect of match variables in the receipt of the promotion campaigns at the department level. Specifically, because we are focused at the department level receipt of the promotion campaigns, we see that the percent of manufacturer purchases is now positively predictive of the campaign receipts whereas at the store level, this variable was negatively correlated with receipt of the promotion campaigns. Percent of purchases at the department, of the brand, and manufacturer-sub-commodity level are still positively correlated with receipt of the promotion campaigns.

For each of the campaign prediction regressions, we calculate the predicted residuals in order to use them as control functions in estimating the effect of the coupon on sales in the final estimations.

Table 3.25: Department Promotion Campaign Predictions

Variable	**Coefficient** (Std. Err.)
Equation 3 : Department Promotions	
Promotion Camp Running in Dept	-0.002*** (0.000)
(Promotion Camp Running in Dept) * Daily	-0.010*** (0.000)
(Promotion Camp Running in Dept) * Twice Weekly	-0.009*** (0.000)
(Promotion Camp Running in Dept) * Weekly	-0.004*** (0.000)
(Promotion Camp Running in Dept) * Oneout	-0.001*** (0.000)
(Promotion Camp Running in Dept) * Biweekly	0.001*** (0.000)
(Promotion Camp Running in Dept) * Low-Medium Sales	0.001*** (0.000)
(Promotion Camp Running in Dept) * Medium Sales	0.002*** (0.000)
(Promotion Camp Running in Dept) * Medium-High Sales	0.005*** (0.000)
(Promotion Camp Running in Dept) * High Sales	0.009*** (0.000)
(Promotion Camp Running in Dept) * Sum[% Purch in Dept]	0.005*** (0.000)
(Promotion Camp Running in Dept) * Sum[% Purch of Manu]	0.003*** (0.000)
(Promotion Camp Running in Dept) * Sum[% Purch of Brand]	0.004*** (0.000)
(Promotion Camp Running in Dept) * Sum[% Purch of Commodity]	-0.002*** (0.000)
(Promotion Camp Running in Dept) * Sum[% Purch of Sub-commodity]	-0.001*** (0.000)
(Promotion Camp Running in Dept) * Sum[% Purch of Manu-Sub-comm]	0.005*** (0.000)
(Promotion Camp Running in Dept) * Sum[% Purch of Man-Comm]	-0.001*** (0.000)
(Promotion Camp Running in Dept) * Sum[% Purch of Man-Dept]	-0.006*** (0.000)
(Promotion Camp Running in Dept) * Sum[% Purch of Brand-Sub-comm]	-0.001*** (0.000)
(Promotion Camp Running in Dept) * Sum[% Purch of Brand-Comm]	-0.005*** (0.000)
(Promotion Camp Running in Dept) * Sum[% Purch of Brand-Dept]	-0.002*** (0.000)
(Promotion Camp Running in Dept) * Δ Spend in L.Promotion Camp	0.000*** (0.000)
Reward Redemption	0.005*** (0.001)
Promotion Redemption	0.607*** (0.003)
Retail Discounts	-0.003*** (0.000)
Other Coupon Discounts	-0.001*** (0.000)
Match Coupon Discounts	-0.003*** (0.001)
Dept Mailer	-0.001*** (0.000)
Dept Display	0.002*** (0.000)
Season 2	0.001*** (0.000)
Season 3	0.001*** (0.000)
Season 4	-0.001*** (0.000)
Intercept	-0.001*** (0.000)
N	10,943,070
R^2	0.5847

3.7.2 Effect of Coupon on Sales

The purpose of this paper is to assess the differential impact of various marketing tools on sales. My model predicts that promotion campaigns will be the most effective in increasing sales at the store level because they are most likely to increase the consideration set of the customers. While reward coupons are equally likely to be considered, since the campaigns are meant to provide discounts on products the customers most likely purchased, the campaigns are not likely to change the consideration set. Finally, we expect that while non-targeted mailers and displays may expand the consideration set of customers if they are considered, the likelihood of consideration is low since they are not relevant to the household and unlikely to grab their attention.

I begin this analysis by looking at the average effect each marketing tool has on expected sales in a week. Then I decompose the effect of the marketing tools on bringing customers to the store and sales conditional on coming to the store. Finally, I examine the effect of the marketing tools on the sales of the departments being promoted.

Average Effect on Expected Sales

Recall that the model for the impact of the marketing tools on overall sales at the store, not conditional on coming to the store, is given by Equation 3.5:

$$S_{it} = \beta_R R_{it} + \beta_P P_{it} + \beta_m m_{jt} + \mathbf{loyalty}_i \boldsymbol{\beta}_{loyal} + \mathbf{s}\boldsymbol{\beta}_s + \rho_R \hat{\epsilon}_{itR} + \rho_P \hat{\epsilon}_{itP} + e_{it}$$

Estimates are reported in Tables 3.26 and 3.27. This model is helpful in giving a sense of the magnitude of the impact of each marketing tool. A downside is that it conflates the effects of each campaign on bringing customers to the store and affecting their purchase behavior once at the store. However, it is helpful to get a perspective on how effective the reward, promotional, and mailer campaigns are relative to each other. The results are given in Table 3.26. As the theoretical model predicts, the promotional campaign is significantly more effective at increasing sales than the reward campaigns or the mailer campaigns. Specifically, we see that the promotional campaigns increase sales by about four times the reward campaigns and almost five times the mailer campaigns. Promotional campaign increase sales by an average of \$24.48 in the weeks of the campaigns while reward campaigns only increase sales by about \$6.04 and mailers by an average of \$5.17 in the weeks of the campaigns.

Table 3.27 shows the estimation including the interaction terms, accounting for when households receive both the reward and promotion campaigns in a week. Recall that although 40 percent of households receive both targeted campaigns over the course of the test period, households receive both campaigns only 14 percent of the total weeks in which households receive either campaign. We see that by controlling for this, reward campaigns are more effective when they are sent alone (increasing sales on average by \$7.22) than when they are combined with promotion campaigns (\$7.22 - \$6.53 = \$0.69). We observe the same thing with promotion campaigns which increase sales by \$26.27 when sent alone and by \$26.27 - \$6.53 = \$19.74 when received together. The same patterns hold true in that promotion campaigns are much more effective overall than reward campaigns.

At first pass, this result may be surprising since redemption rates of reward coupons far surpass those of promotion coupons. Examination of the redemption rates alone may lead one to presume that reward coupons perform better than promotion coupons. Additionally, the underlying presumption in targeting is that often that the closer the targeted advertisement is to the preferences of the individual, the more likely they are to respond positively. This result suggests that the positive response via higher redemption may not be the most important factor to examine. While this result cannot identify the driver of the differential impact, it suggests that the theoretical predictions of consideration sets have merit. The estimation results set the stage for more examination at the store and the department levels.

We also see the importance of controlling for loyalty characteristics in the estimation of the coupons on sales. We see that the average sales of a household with high sales which comes bi-weekly is \$110.31 + \$2.28+ \$1.94 = \$114.53, whereas the average sales of a household with medium sales which comes weekly is \$27.81 + \$3.37 + \$1.94 = \$33.12.

Estimates are bootstrapped 400 times in order to account for the fact that the residuals for the receipt of the reward and promotion campaigns are generated regressors. The bootstraps are done with household clusters. The overall model has an $R^2 = 0.2478$.

Decomposing Impact on Coming to Store and Spending Conditional on Coming

To better understand how the different marketing tools differentially impact customer paths in relation to the theoretical predictions, we decompose the effect of each tool on coming to the store and store-level sales conditional on coming to the store. Recall that our estimating model for this decomposition is given by Equation 3.22, where we control for the endogeneity of the reward and promotion campaigns with the predicted residuals, $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{it}}$:

$$
\begin{aligned}
\boldsymbol{l}_{it}(\boldsymbol{\theta}) = \sum_i \sum_t &(1 - v_{it})[\log 1 - \Phi(\tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{it}}')] \\
&+ (v_{it})[\log \Phi((1-\tilde{\rho}^2)^{-1/2}[\tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{it}}' + \tilde{\rho}\tilde{\sigma}^{-1}(S_{it} - \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} - \boldsymbol{\psi}_{11}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{it}}')]) \\
&+ \phi(\tilde{\sigma}^{-1}(S_{it} - \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} - \boldsymbol{\psi}_{11}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{it}}')) - \log\tilde{\sigma}],
\end{aligned}
$$

Estimates for household purchase conditional on coming to the store are reported in Tables 3.28 and 3.30, and estimates for coefficients in the likelihood of coming

Table 3.26: Average Effect Estimates

Variable	**Coefficient** (Bootstrap Std. Err.)
Reward Coupon	6.041***
	(0.702)
Promotion Coupon	24.478***
	(1.748)
Mailer	5.167***
	(0.294)
Daily	3.863
	(2.265)
Twice Weekly	2.238
	(1.202)
Weekly	3.371
	(1.144)
Oneout	2.619***
	(0.968)
Biweekly	2.281***
	(0.806)
Low-Medium Sales	12.012***
	(0.714)
Medium Sales	27.809***
	(1.004)
Medium-High Sales	57.551***
	(1.985)
High Sales	110.305***
	(5.327)
Reward Coupon Residual	-11.857***
	(4.818)
Promotion Coupon Residual	-28.268***
	(3.663)
Season 2	5.799***
	(0.344)
Season 3	6.021***
	(0.379)
Season4	6.613***
	(0.346)
Intercept	1.943
	(0.717)
N	254,490
Replications	400
Replication Clusters	2,495
R^2	0.2478

Table 3.27: Average Effect Estimates with Interaction

Variable	**Coefficient** (Std. Err.)
Reward Coupon	7.216 (0.814)
Promotion Coupon	26.265 (1.809)
Reward & Promotion Coupon	-6.528 (1.470)
Mailer	5.140 (0.294)
Daily	3.768 (2.268)
Twice Weekly	2.140 (1.201)
Weekly	3.304 (1.145)
Oneout	2.571 (0.967)
Biweekly	2.267 (0.804)
Low-Medium Sales	11.954 (0.712)
Medium Sales	27.735 (1.002)
Medium-High Sales	57.478 (1.988)
High Sales	110.234 (5.327)
Reward Coupon Residual	-15.732 (4.918)
Promotion Coupon Residual	-28.194 (3.658)
Season 2	5.790 (0.345)
Season 3	6.008 (0.380)
Season 4	6.649 (0.347)
Intercept	1.938 (0.716)
N	254490
R^2	0.248
$\chi^2_{(18)}$	3770.996

to the store are reported in 3.29 and 3.31. The maximum likelihood estimation is bootstrapped 400 times in order to account for the generated regressors, $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{it}}$.

In Table 3.28, we observe that the effect on sales of receiving a promotion campaign (and not redeeming a coupon) is an average increase in sales at the store of \$3.11. This compares to the negative impact of reward campaigns without redemption which drop sales by -\$2.41 on average during the weeks of the campaign. Each of these values is statistically significant at the 1 percent level. While we do not have a statistically significant observation of the impact of redemption of promotion campaigns, the average increase in sales when customers redeem promotion campaigns is \$5.51. At the average, this implies a total increase in sales when customers redeem promotion coupons of \$8.62. For reward coupons, we see a positive and statistically significant effect (at 1 percent) of redemption on the sales conditional on coming to the store of -\$2.41 + \$9.54 = \$7.13.

In Table 3.30, we include the interaction term to account for the weeks in which households receive both campaigns (14.1 percent of the weeks in which households receive either campaign). We similarly find that promotion campaigns have an average positive effect on sales in the store once customers are at the store without redemptions (\$2.49) and with redemptions (\$2.49+ \$5.73=\$8.22). Customers receiving only reward campaigns also on average lower there sales at the grocery store in the weeks of the campaign by \$2.83 when they do not redeem coupons, although they increase sales by -\$2.83 + \$9.58 = \$6.76 when they redeem coupons. Interestingly, when households do not redeem coupons but receive both campaign during the same week, we observe a higher level of in-store sales for both the rewards and promotion campaigns (-\$2.83 + \$2.22 = -\$0.61 for rewards and \$2.49 + \$2.22 = \$4.70 for promotions). However, when households redeem coupons from both, we observe a lower increase in sales from the redemptions (\$9.58 - \$1.68 = \$7.90 for rewards and \$5.73 - \$1.68 = \$4.05 for promotions.

The in-store estimates suggest that promotional campaigns are more effective at expanding the consideration set of customers at the store level relative to the other campaigns. While we cannot identify the specific impact on the path yet, this estimation gives us a sense of the overall impact across the store. According to our model, there are three cases in which we would expect to observe an increase in sales at the store level. The first derives from the two cases are when the changing of the order of stores in **Different store, same department** or **Different store, new department** causes an increase in the expected sales of the store because the preferred store receives a bigger share of the customer's shopping list. Because stores lower in the ordering only sell the remaining items on the customer's shopping list, the higher the store is to the top of the ordering, the more likely it is to have a bigger share of the customer's overall grocery spending. Therefore, a change in ordering can cause an increase in spending overall at the store level conditional on coming to the store.

The second which can induce a change in spending at the store level is the **Same store, new department** change in path. Here, when a product in a new department is considered, we expect an increase in the overall sales at the store level because a customer with an expanded consideration set will have a greater expected increase in their overall spending at the store. We cannot identify this specific effect with the store-level estimation. However, we can identify this with the department sales estimation which we will present in the next section.

We also note that generic mailers have a negative and statistically significant (at 1 percent) average impact on sales conditional on coming to the store at -$1.76. This also aligns with our model predictions in that we don't expect customers to pay much attention to the products in the mailers in expanding their consideration set to products they otherwise would not purchase.

Finally, we note that other discounts at the store increase average sales as well.

Retail discounts increase sales by an average of \$3.25 (statistically significant at 1 percent), rivaling the effectiveness of promotional campaigns without redemption. Retail discounts are small signs at the product location indicating a discount for the customers using the retailer's loyalty card. This suggests that the product-level signage in the aisles attract the attention of customers as they are walking through the department, causing them to increase their consideration set to these products in this context. However, analysis at the department level would give more clarity in this hypothesis. We also observe that match coupon discounts on average increase sales by \$1.68 and other coupon discounts on average increase sales by \$0.79. The match discounts are not statistically different from zero at the 10 percent level and the other coupon discounts are not statistically different from zero at the 5 percent level.

We observe two levels of advertising at the store level. The reward coupons, promotion coupons, and mailers are marketing tools customers take with them from home and are meant to direct the purchase path of customers by highlighting products. We see that without redemption, the promotion campaigns are the most effective in increasing sales conditional on coming to the store. On the other hand, redemption of reward coupons, promotion coupons, retail discounts, other coupon discounts, and match coupons are marketing tools which take effect once a customer is physically evaluating a product at the store. We observe that the redemption effect for reward coupons has the greatest statistically significant impact on increasing sales. The average effect of promotional redemption is higher, but it is not statistically significant. We also see that retail discounts increase sales, indicating that they are effective at attracting the attention of customers when customers are walking through the department. We will observe these effects in more detail in our department level analysis.

In Table 3.29 and 3.31, I report the marginal effects at the mean (means are

Table 3.28: Store Sales LIML Estimates

Variable	**Coefficient** (Bootstrap Std. Err.)
Equation 1 : Weekly Sales Conditional on Coming	
Reward Coupon	-2.405***
	(0.588)
Promotion Coupon	3.105**
	(1.244)
Reward Redemption	9.544***
	(2.583)
Promotion Redemption	5.502
	(4.459)
Mailer	-1.760***
	(0.441)
Retail Discounts	-3.247***
	(0.057)
Other Coupon Discounts	-0.789*
	(0.404)
Match Coupon Discounts	-1.684
	(1.414)
Daily	-32.851***
	(3.142)
Twice Weekly	-28.141***
	(2.232)
Weekly	-21.272***
	(2.107)
Oneout	-15.381***
	(1.874)
Biweekly	-8.199***
	(1.731)
Low-Medium Sales	9.033***
	(0.967)
Medium Sales	18.290***
	(1.296)
Medium-High Sales	35.421***
	(2.132)
High Sales	76.235***
	(5.703)
Reward Coupon Residual	-14.781***
	(3.828)
Promotion Coupon Residual	-10.452***
	(2.428)
Season 2	-5.333***
	(0.448)
Season 3	-2.864***
	(0.485)
Season 4	-1.304***
	(0.452)
Intercept	47.547***
	(3.199)
N	254,490
Replications	400
Replication Clusters	2,495

reported in Table 3.11)of the marketing tools used to impact the choice to come to the store in a given week. In Table 3.29, we observe that promotional campaigns are the most effective at increasing the likelihood of customers coming to the store by a statistically significant (at the 1 percent level) average of 21.4 percent at the mean. Reward coupons also increase the likelihood of customers coming to the store by a statistically significant 10.6 percent at the mean, but the effect is about half that of the promotional campaigns. Mailers are the least effective at bringing customers to the store, increasing on average by 8.7 percent at the mean. In Table 3.31, we similarly observe that promotional campaigns are more likely to increase sales than reward campaigns (24.9 percent for promotional and 12.6 percent for reward), but the effect is lower for both during weeks when they receive both campaigns by 13.3 percent. Mailers are still the least effective at 8.6 percent.

Recall that when consideration of a product causes a customer to change the ordering of their store visits as in the cases of **Different store, same department** or **Different store, new department**, we expect to observe an increased average likelihood of customers coming to the store when they receive the campaigns inducing this shift. The theoretical model predicts that personalized campaigns are more likely to induce this change in store ordering than generic campaigns because personalized campaigns are more likely to be considered. Therefore, the differential impact of the marketing tools on affecting the likelihood of store visits is in line with the theoretical predictions.

These estimates confirm our hypothesis that the promotional campaigns are the most effective at attracting the attention and expanding the consideration set of customers receiving the advertisements. Recall that reward coupons were sent primarily to customers which came frequently whereas promotion coupons were not specifically sent to household that came with more frequency and actually targeted away from customers that came very frequently (daily). Therefore, within this context, we see that promotional campaigns are more effective at prompting customers to change their purchase path at the earliest stage of coming to the store. Reward coupons do not have as much impact at this stage because recipients on average come more already. Finally, mailers do not seem to have much impact, confirming the hypothesis that they are not as effective at attracting the attention of the recipients.

Finally, Table 3.32 show the store level estimates conditional on coming to the store, without the selection correction. The estimated effects of each marketing campaign is biased upwards, validating the correction through the Maximum Likelihood Estimation. This implies that the errors in the selection equation and the main estimation equation are in fact correlated and that it is necessary to correct for the correlation.

Department Sales

Recall that the model for the impact of marketing tools at the department level is given by Equation 3.37. Here we control for the store-level endogenous reward and promotion coupons with $\hat{\boldsymbol{\epsilon}}_{\mathbf{1}\boldsymbol{it}}$ and the department-level promotional coupon with $\hat{\boldsymbol{\epsilon}}_{\mathbf{2}\boldsymbol{it}}$.

$$D^*_{dit}|\hat{\boldsymbol{\epsilon}}_{\mathbf{1}\boldsymbol{it}}, \hat{\boldsymbol{\epsilon}}_{\mathbf{2}\boldsymbol{it}} = \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} + \boldsymbol{\psi}_{11}\hat{\boldsymbol{\epsilon}}_{\mathbf{1}\boldsymbol{it}}' + \boldsymbol{\psi}_{12}\hat{\boldsymbol{\epsilon}}_{\mathbf{2}\boldsymbol{it}}' + \frac{\tilde{\rho}}{\tilde{\sigma}}[\lambda(\tilde{\mathbf{W}}_{it}\boldsymbol{\gamma} + \boldsymbol{\psi}_{21}\hat{\boldsymbol{\epsilon}}_{\mathbf{1}\boldsymbol{it}}' + \boldsymbol{\psi}_{22}\hat{\boldsymbol{\epsilon}}_{\mathbf{2}\boldsymbol{it}}')] + \iota_{it}$$

Estimates for household purchases in the departments conditional on coming to the store are reported in Table 3.33 and 3.35, and estimates for coefficients in the likelihood of coming to the store are reported in 3.34 and 3.36. These estimates provide the clearest picture of how the marketing tools affect the purchase path for customers. We examine the same marketing tools at the department level as we did at the store level.

First, in Table 3.33, we observe that promotional campaigns have a sizable positive and statistically significant impact on department level sales in the departments of

Table 3.29: Store Sales LIML Estimates

Variable	$\partial E(v*_{it})/\partial x$ (Bootstrap Std. Err.)
Equation 2 : Likelihood of Coming to Store	
Reward Coupon	0.106***
	(0.018)
Promotion Coupon	0.214***
	(0.030)
Daily	0.494***
	(0.063)
Twice Weekly	0.393***
	(0.043)
Weekly	0.323***
	(0.040)
Oneout	0..231***
	(0.036)
Biweekly	0.130***
	(0.031)
Low-Medium Sales	0.073***
	(0.024)
Medium Sales	0.105***
	(0.030)
Medium-High Sales	0.189***
	(0.044)
High Sales	0.288***
	(0.076)
L.[% Basket Perishable Goods] - Ave[% Basket Perishable Goods]	0.0002**
	(0.000)
Mailer	0.087***
	(0.008)
Reward Coupon Residual	0.052
	(0.112)
Promoted Coupon Residual	-0.135***
	(0.069)
Season 2	0.120***
	(0.010)
Season 3	0.122***
	(0.011)
Season 4	0.113***
	(0.010)
λ	-17.118
	(1.819)
$\tilde{\rho}$	-0.413
	(0.044)
$\tilde{\sigma}$	41.446
	(1.028)
$\chi^2_{(22)}$	5,614.0
N	254,490
Replications	400
Replication Clusters	2,495

Table 3.30: Store Sales LIML Estimates with Interaction

Variable	**Coefficient** (Bootstrap Std. Err.)
Equation 1 : Weekly Sales Conditional on Coming	
Reward Coupon	-2.828*** (0.697)
Promotion Coupon	2.486* (1.303)
Reward & Promotion Coupon	2.218* (1.189)
Reward Redemption	9.583*** (2.488)
Promotion Redemption	5.734 (4.324)
Reward & Promotion Redemption	-1.680 (13.590)
Mailer	-1.747*** (0.441)
Retail Discounts	-3.247*** (0.057)
Other Coupon Discounts	-0.789* (0.404)
Match Coupon Discounts	-1.681 (1.415)
Daily	-32.802*** (3.140)
Twice Weekly	-28.093*** (2.230)
Weekly	-21.237*** (2.106)
Oneout	-15.355*** (1.874)
Biweekly	-8.191*** (1.730)
Low-Medium Sales	9.060*** (0.967)
Medium Sales	18.323*** (1.294)
Medium-High Sales	35.457*** (2.130)
High Sales	76.269*** (5.706)
Reward Coupon Residual	-13.407*** (3.879)
Promoted Coupon Residual	-10.497*** (2.430)
Season 2	-5.323*** (0.447)
Season 3	-2.853*** (0.484)
Season 4	-1.312*** (0.450)
Intercept	47.522*** (3.194)
N	254,490
Replications	400
Replication Clusters	2,495

Table 3.31: Store Sales LIML Estimates with Interaction

Variable	$\partial E(v*_{it})/\partial x$ (Bootstrap Std. Err.)
Equation 2 : Likelihood of Coming to Store	
Reward Coupon	0.126***
	(0.020)
Promotion Coupon	0.249***
	(0.032)
Reward & Promotion Coupon	-0.133***
	(0.035)
Mailer	0.086***
	(0.008)
Daily	0.493***
	(0.063)
Twice Weekly	0.391***
	(0.042)
Weekly	0.322***
	(0.040)
Oneout	0.230***
	(0.036)
Biweekly	0.130***
	(0.031)
Low-Medium Sales	0.072***
	(0.024)
Medium Sales	0.104***
	(0.030)
Medium-High Sales	0.188***
	(0.044)
High Sales	0 .287***
	(0.076)
L.[% Basket Perishable Goods] - Ave[% Basket Perishable Goods]	0.000**
	(0.000)
Reward Coupon Residual	-0.021
	(0.113)
Promoted Coupon Residual	-0.136***
	(0.070)
Season 2	0.120***
	(0.010)
Season 3	0.122***
	(0.011)
Season 4	0.114***
	(0.010)
λ	-17.103
	(1.819)
$\tilde{\rho}$	-0.413
	(0.044)
$\tilde{\sigma}$	41.442
	(1.028)
$\chi^2_{(24)}$	5,607.9
N	254,490
Replications	400
Replication Clusters	2,495

Table 3.32: Store Sales Conditional on Coming, Without Selection Correction

Variable	**Coefficient** (Std. Err.)
Reward Coupon	-0.265 (0.569)
Promotion Coupon	7.053 (1.324)
Reward Redemption	9.515 (2.583)
Promotion Redemption	5.451 (4.477)
Mailer	0.426 (0.351)
Retail Discounts	-3.220 (0.059)
Other Coupon Discounts	-0.741 (0.411)
Match Coupon Discounts	-1.815 (1.428)
Daily	-19.561 (2.665)
Twice Weekly	-16.623 (2.064)
Weekly	-11.390 (1.934)
Oneout	-7.939 (1.820)
Biweekly	-3.877 (1.752)
Low-Medium Sales	11.210 (0.927)
Medium Sales	21.201 (1.232)
Medium-High Sales	39.758 (2.146)
High Sales	81.236 (5.581)
Reward Coupon Residual	-11.987 (3.828)
Promotion Coupon Residual	-12.008 (2.513)
Season 2	-2.152 (0.355)
Season 3	0.297 (0.344)
Season 4	1.604 (0.324)
Intercept	20.131 (1.679)
N	123,812
R^2	0.645
Replications	400
Replication Clusters	2,495

the promoted products. Even when households do not redeem coupons, we see that on average sales in the department level increase by \$3.06 (statistically significant at the 1 percent level). When customers redeem the coupons, we see the increase in average sales in the department jump to \$3.06 + \$7.57 = \$10.63. This is a far greater impact than any of the other marketing tools and it is concentrated in the department of the promoted products.

In Table 3.35, we account for the weeks and departments in which a household receives both promotional campaign and a reward campaign. Overlap is very small at the department level at 1.45 percent, but it is still helpful to see the dual campaign effect at the department level. Promotion campaigns still increase sales in the department of the promoted product by \$2.88 when households only receive the promotion campaign and do not redeem a coupon. The effect is slightly higher (\$2.88+\$0.62=\$3.50) when households receive both the reward and promotion campaign. The reward campaign effect without redemption is also slightly higher when combined with promotion campaigns (-\$0.01 + \$0.62 = \$0.61). Redemptions effects are even higher when combined together. Promotion redemptions increase department sales by an additional \$7.36 when alone and by an additional \$7.36+\$3.66 = \$11.01 when combined with reward campaigns. Reward redemptions increase sales by \$1.91 when alone and by an additional \$1.91 + \$3.66 = \$5.56 when combined with the promotion redemptions (this is in addition to the receipt effect for promotions \$2.88 and rewards -\$0.01).

The estimates give the clearest empirical support for the theoretical predictions given in the case, **Same store, new department**. As the model predicts, campaigns which attract customers' attention for products outside of their shopping path will have the greatest impact on customer consideration sets and expected sales, especially along the new path the customer takes to examine the product advertised. We observe here that customers in fact increase their sales considerably in the departments of the promoted products both when they redeem and do not redeem the coupons.

Unfortunately, analysis of the impact of reward coupons at this level is hindered by data limitations. However, we are able to see the impact of reward redemptions in the department of the rewarded products and see a statistically significant increase of \$1.96. This is a much smaller impact than the promotional campaigns which is in line with the theoretical predictions for the **Same store, same department** case. Because reward campaigns give discounts for households' frequently purchased products, they are generally for products along the purchase path or even in the shopping list for the customer. Therefore, as the model predicts, there is a small impact on the consideration set and expected sales. The estimate for reward campaigns redemptions confirms the theoretical predictions for this effect.

The impact of mailers and displays is minimal with mailers increasing sales in the department of products in the mailers by an average of \$0.04 (statistically significant at the 10 percent level) and displays increasing by an average of \$0.05 for products on display (statistically significant at the 1 percent level). These estimates are in line with the **No consideration** case which is predicted by the model for generic marketing tools. Because generic mailers and in-store displays are not personalized to the customer, the likelihood of a customer considering products in the advertisements is low according to the model. Therefore, we expect that these marketing tools have little to no impact on the customer's consideration set and expected sales. This is confirmed with these small estimates.

As our model predicts, we see that product level marketing tools like retail discounts increase the sales once customers are in the aisles of the department. We see that retail discounts increase sales on average by \$2.84 (significant at the one percent level), match coupon discounts increase sales in the department by \$4.36 (significant at the five percent level), and other coupon discounts increase sales in the department by \$1.02 (significant at the one percent level). We also see that the stronger impact of these marketing tools in the department level analysis indicates that they are more effective at attracting attention and expanding consideration sets once customers are in the department than when customers are determining their purchase path at the store level.

Finally, we find the same effects of the promotion, reward, and mailers in bringing households to the store as we estimated at the store-level. Each of the Heckman two-stage estimates are bootstrapped 400 times with 2,495 household level clusters.

Table 3.37 shows the estimation of the effect marketing tools on department sales without correction for the selection of coming to the store. The estimates are not significantly different from 3.33, suggesting that the selection effect at the store level is not as significant as at the store level.

3.8 Conclusion

It is increasingly important to understand customers' preferences and personally tailor advertisements to them. Faced with heightened competition from online retailers, brick and mortar stores are trying to differentiate themselves by understanding the customer journey in their path to purchase. This paper offers a framework for understanding how customers evaluate whether to consider advertisements and the impact on customers' purchase path.

In this paper I model the impact of personalized advertisements on consumers' purchase path with the store with the context of bounded consideration sets. I allow personalized advertisements to be more likely to be considered than generic advertisements. If a considered product is outside of the customer's purchase path, she must go to the location of the product in order to evaluate it. The process of going to the product opens her to additional localized marketing for products in the department of the product, expanding her consideration set further. The model predicts that

Table 3.33: Department Sales Heckman Estimates

Variable	Coefficient (Bootstrap Std. Err.)
Equation 1 : Weekly Department Sales Conditional on Coming	
Promotion Department Coupon	3.062*** (0.202)
Reward Coupon	0.006 (0.015)
Reward Department Redemption	1.956*** (0.304)
Promotion Department Redemptions	7.574*** (1.921)
Department Retail Coupon	-2.837*** (0.057)
Department Other Coupon	-1.020** (0.421)
Department Match Coupon Discounts	-4.357*** (1.313)
Department Mailer	0.042* (0.024)
Department Display	0.046*** (0.018)
Daily	-0.452*** (0.086)
Twice Weekly	-0.384*** (0.068)
Weekly	-0.261*** (0.062)
Oneout	-0.178*** (0.053)
Biweekly	-0.085* (0.046)
Low-Medium Sales	0.296*** (0.023)
Medium Sales	0.561*** (0.030)
Medium-High Sales	1.030*** (0.052)
High Sales	2.030*** (0.132)
Promoted Department Coupon Residual	-0.283** (0.119)
Reward Coupon Residual	0.044 (0.085)
Season 2	-0.034** (0.015)
Season 3	0.009 (0.015)
Season 4	0.038*** (0.014)
Intercept	-0.271** (0.108)

Due to space constraints, department coefficients are not displayed.

Table 3.34: Department Sales Heckman Estimates

Variable	$\partial E(v*_{it})/\partial x$ (Bootstrap Std. Err.)
Equation 2 : Likelihood of Coming to Store	
Reward Coupon	0.105*** (0.017)
Promotion Coupon	0.205*** (0.027)
Daily	0.499*** (0.063)
Twice Weekly	0.400*** (0.043)
Weekly	0.328*** (0.041)
Oneout	0.234*** (0.036)
Biweekly	0.131*** (0.031)
Low-Medium Sales	0.073*** (0.025)
Medium Sales	0.102*** (0.030)
Medium-High Sales	0.176*** (0.042)
High Sales	0.232*** (0.060)
L.[% Basket Perishable Goods] - Ave[% Basket Perishable Goods]	0.000** (0.000)
Mailer	0.087*** (0.009)
Reward Coupon Residual	0.046 (0.108)
Promoted Coupon Residual	-0.121*** (0.062)
Season 2	0.121*** (0.010)
Season 3	0.123*** (0.011)
Season 4	0.114*** (0.010)
λ	0.068 (0.058)
$\tilde{\rho}$	0.013
$\tilde{\sigma}$	5.084
$\chi^2_{(65)}$	21,417.4
N	10,943,070
Replications	400
Replication Clusters	2,495

Table 3.35: Department Sales Heckman Estimates with Interaction

Variable	**Coefficient** (Bootstrap Std. Err.)
Equation 1 : Weekly Department Sales Conditional on Coming	
Promotion Department Coupon	2.878*** (0.205)
Reward Coupon	-0.009 (0.014)
Promotion Department & Reward Coupon	0.621*** (0.224)
Reward Department Redemption	1.909*** (0.291)
Promotion Department Redemption	7.356*** (1.688)
Promotion & Reward Department Redemption	3.655 (8.446)
Department Retail Coupon	-2.837*** (0.057)
Department Other Coupon	-1.018** (0.421)
Department Match Coupon Discounts	-4.350*** (1.314)
Department Mailer	0.041* (0.024)
Department Display	0.046*** (0.018)
Daily	-0.476*** (0.083)
Twice Weekly	-0.405*** (0.066)
Weekly	-0.279*** (0.061)
Oneout	-0.192*** (0.052)
Biweekly	-0.093** (0.045)
Low-Medium Sales	0.292*** (0.023)
Medium Sales	0.556*** (0.030)
Medium-High Sales	1.022*** (0.051)
High Sales	2.021*** (0.132)
Promoted Department Coupon Residual	-0.268** (0.118)
Reward Coupon Residual	0.084 (0.083)
Season 2	-0.040*** (0.014)
Season 3	0.003 (0.015)
Season 4	0.031** (0.014)
Intercept	-0.221** (0.103)

Due to space constraints, department coefficients are not displayed.

Table 3.36: Department Sales Heckman Estimates with Interaction

Variable	$\partial E(v*_{it})/\partial x$ (Bootstrap Std. Err.)
Equation 2 : Likelihood of Coming to Store	
Reward Coupon	0.125*** (0.020)
Promotion Coupon	0.240*** (0.029)
Reward & Promotion Coupon	-0.135*** (0.034)
Mailer	0.086*** (0.009)
Daily	0.497*** (0.062)
Twice Weekly	0.398*** (0.043)
Weekly	0.327*** (0.040)
Oneout	0.234*** (0.036)
Biweekly	0.131*** (0.031)
Low-Medium Sales	0.072*** (0.025)
Medium Sales	0.101*** (0.030)
Medium-High Sales	0.175*** (0.042)
High Sales	0.230*** (0.060)
L.[% Basket Perishable Goods] - Ave[% Basket Perishable Goods]	0.000** (0.000)
Reward Coupon Residual	-0.072 (0.110)
Promotion Coupon Residual	- 0.121*** (0.063)
Season 2	0.121*** (0.010)
Season 3	0.123*** (0.011)
Season 4	0.114*** (0.010)
λ	0.036 (0.054)
$\tilde{\rho}$	0.007
$\tilde{\sigma}$	5.084
$\chi^2_{(67)}$	21,496.5
N	10,943,070
Replications	400
Replication Clusters	2,495

Table 3.37: Department Sales Conditional on Coming, No Selection Correction

Variable	**Coefficient** (Bootstrap Std. Err.)
Promotion Department Coupon	3.044***
(0.200)	
Reward Coupon	-0.005
	(0.014)
Reward Department Redemption	1.957***
	(0.304)
Promotion Department Redemptions	7.584***
	(1.921)
Department Retail Coupon	-2.837***
	(0.057)
Department Other Coupon	-1.020***
	(0.421)
Department Match Coupon Discounts	-4.357***
	(1.313)
Department Mailer	0.039
	(0.024)
Department Display	0.046***
	(0.018)
Daily	-0.505***
	(0.062)
Twice Weekly	-0.429***
	(0.048)
Weekly	-0.300***
	(0.045)
Oneout	-0.208***
	(0.043)
Biweekly	-0.102***
	(0.041)
Low-Medium Sales	0.287***
	(0.021)
Medium Sales	0.549***
	(0.028)
Medium-High Sales	1.011***
	(0.049)
High Sales	2.008***
	(0.127)
Promoted Department Coupon Residual	-0.266***
	(0.118)
Reward Coupon Residual	0.046
	(0.084)
Season 2	-0.046***
	(0.008)
Season 3	-0.004
	(0.008)
Season 4	0.025***
	(0.007)
Intercept	-0.168***
	(0.041)
N	5,323,916
R^2	0.656
Replications	400
Replication Clusters	2,495

personalized advertisements that bring a customer outside her status quo purchase path will on average increase sales more than products along her purchase path.

I empirically test these predictions with point of sale data from a national grocery retailer. I find that promotional campaigns, which personalize coupons to households while directing them to new products, are most effective in that they increase the likelihood of customers coming to the store and increase sales conditional on coming to the store. Additionally, I identify an increase in sales in the department of the promoted products, validating the hypotheses of the theoretical model. I also find that reward campaigns, which personalize coupon to households for products they regularly purchase, have higher redemption rates. However, reward campaigns are not as effective at bringing customers to the store and sales conditional on coming to the store are not as high as promotional campaigns. Even when customers redeem reward coupons, the impact on department sales is significantly lower than promotional coupon redemptions in the department.

The grocery retail context provides a natural foundation from which to develop a theoretical model for customer purchase paths given the repetitive nature of grocery shopping trips and the complexity of product offerings at the store. Additionally, advancements in loyalty card analytics provides rich data from which retailers can personalize campaigns at the household level. However, the lessons drawn from the theoretical and empirical analyses are applicable to other retailing contexts.

In particular, as brick and mortar retailers strive to differentiate themselves from online retailers, many are utilizing consumer analytics companies to track customers geolocation through their smartphones. This theoretical model helps evaluate how advertisements can influence customers in the context of a complete consumer journey. The findings in this paper have considerable implications for targeted advertising in this broader context. A lot of attention is placed on retargeted advertising for products recently purchased. The theoretical model and empirical findings suggest that the impact of these campaigns on consumer purchase paths and eventual sales is likely low. On the other hand, targeted advertisements that expand the purchase path of customers can increase overall sales.

Bibliography

[1] Supermarket facts. http://www.fmi.org/research-resources/supermarket-facts.

[2] 2014 mass merchant shopper engagement study: In-store decision rate, 2014.

[3] Retail analytics market by business function. *Markets and Markets* (2014).

[4] Retailing wholesaling. *USDA's Economic Research Service* (2015).

[5] Achabal, D., Gorr, W., and Mahajan, V. Multiloc: A multiple store location decision model. *Journal of Retailing 58* (1982), 5–25.

[6] Alfieri, P. Data driven and digitally savvy: The rise of the new marketing organization. *Forbes Insights and Turn* (2015).

[7] Anderson, S., Baik, A., and Larson, N. How targeted advertising affects price competition, profits, and consumers. 2015.

[8] Anderson, S., and de Palma, A. Competition for attention in the information (overload) age. *The RAND Journal of Economics 43* (2012), 1–25.

[9] Anderson, S., and de Palma, A. . Information congestion. *The RAND Journal of Economics 40* (2009), 688–709.

[10] Anderson, S. P., and Renault, R. Advertising content. *American Economic Review 96*, 1 (2006), 93–113.

[11] Ansari, A., and Mela, C. E-customization. *Journal of Marketing Research 40*, 2 (2003), 131–146.

[12] Arnold, S. *Shopping habits at Kingston department stores: wave III: three years after WalMart's entry into Canada.* PhD thesis, Queen's University School of Business, 1997.

[13] Arora, N., Hess, J., Joshi, Y., Neslin, S., and Thomas, J. Putting one-to-one marketing to work: Personalization, customization, and choice. *Marketing Letters 19* (2008), 305–321.

[14] Bagwell, K. The economic analysis of advertising. In *Handbook of Industrial Organization* (2007), pp. 1701–1844.

[15] Baker, J., Parasuraman, A., Grewal, D., and Voss, G. B. The influence of multiple store environment cues on perceived merchandise value and patronage intentions. *Journal of Marketing 66*, 2 (2002), 120–141.

[16] Balachander, S., and Ghose, S. Reciprocal spillover effects: A strategic benefit of brand extensions. *Journal of Marketing 67*, 1 (2003), 4–13.

[17] Baumgarth, C. Evaluations of co-brands and spillover effects: Further empirical results. *Journal of Marketing Communications 10* (2004), 115–131.

[18] Beatty, S. E., and Ferrell, M. E. Impulsive buying: Modeling its precursors. *Journal of Retailing 74*, 2 (1998), 169–191.

[19] Bell, D., Corsten, D., and Knox, G. From point of purchase to path to purchase: How preshopping factors drive unplanned buying. *Journal of Marketing 75*, 1 (2011), 31–45.

[20] Bell, D., Ho, T.-H., and Tang, C. Determining where to shop: Fixed and variable costs of shopping. *Journal of Marketing Research* (1998), 352–369.

[21] Bell, D. R., and Lattin, J. M. Grocery shopping behavior and consumer response to retailer price format: Why 'large basket' shoppers prefer edlp. *Marketing Science 17*, 1 (1998), 66–88.

[22] Bester, H., and Petrakis, E. Coupons and oligopolistic price discrimination. *International Journal of Industrial Organization 14* (1996).

[23] Bester, H., and Petrakis, E. Coupons and oligopolistic price discrimination. *International Journal of Industrial Organization 14*, 2 (1996), 227–242.

[24] Bickel, P. J., Ritov, Y., and Tsybakov, A. B. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics 37*, 4 (2009), 1705–1732.

[25] Bradlow, E., and Hui, S. Bayesian multi-resolution spatial analysis with applications to marketing. 2012.

[26] Braithwaite, D. The economic effects of advertisement. *Economic Journal* (1928).

[27] Breugelmans, . ., and Campo, K. Effectiveness of in-store displays in a virtual store environment. *Journal of Retailing 87*, 1 (2011), 75–89.

[28] Butters, G. Equilibrium distributions of sales and advertising prices. *Review of Economic Studies* (1977).

[29] Chamberlin, E. *The Theory of Monopolistic Competition.* Harvard University Press, 1933.

[30] Chandon, P., Hutchinson, J., Bradlow, E., and Young, S. Does in-store marketing work?. effects of the number and position of shelf facings on

brand attention and evaluation at the point of purchase. *Journal of Marketing 73* (2009), 1–17.

[31] Chanil, D. Pg's annual report: Optimism wavering?, 2016.

[32] Chevalier, J. A., and Kashyap, A. K., and Rossi, P. E. Why don't prices rise during periods of peak demand? evidence from scanner data. *American Economic Review 93*, 1 (2003), 15–37.

[33] Columbus, L. Roundup of analytics, big data business intelligence forecasts and market estimates, 2014. *Forbes* (2014).

[34] Crask, M., and Reynolds, F. An in-depth profile of the department store shopper. *Journal of Retailing 54* (1978), 23–32.

[35] Desai, K. K., and Keller, K. L. The effects of ingredient branding strategies on host brand extendibility. *Journal of Marketing 66*, 1 (2002), 73–93.

[36] Diamond, P. A model of price adjustment. *Journal of Economic Theory* (1971).

[37] Dias, M. B., Locher, D., Li, M., El-Deredy, W., and Lisboa, P. J. The value of personalised recommender systems to e-business: a case study. In *RecSys* (New York, NY, USA, 2008), pp. 291–294.

[38] Donthu, N., and Rust, R. T. Estimating geographic customer densities using kernel density estimation. *Marketing Science 8* (1989), 191–203.

[39] Doty, C. Marketers: Consumers don't trust your ads, 2014.

[40] Drèze, X., and Hoch, S. J. Exploiting the installed base using cross-merchandising and category destination programs. *International Journal of Research in Marketing 15*, 5 (1998), 459–71.

[41] Eliaz, K., and Spiegler, R. Consideration sets and competitive marketing. *Review of Economic Studies 78* (2011), 235–262.

[42] Erdem, T., and Sun, B. An empirical investigation of spillover effects of marketing mix strategy in umbrella branding. *Journal of Marketing Research 39*, 4 (2002), 408–420.

[43] EuroMonitor. Grocery retailers in the us: Category briefing, 2017.

[44] Fikes, D. U.s. grocery shopping trends, 2016.

[45] Fong, N. M. Targeted marketing and customer search. 2012.

[46] Frank, I., and Friedman, J. A statistical view of some chemometrics regression tools (with discussion). *Technometrics 35* (1993), 109–148.

[47] Ghosh, A., and Craig, S. Formulating retail location strategy in a changing environment. *Journal of Marketing* (1983), 56–68.

[48] Goldfarb, A., and Tucker, C. Online display advertising: Targeting and obtrusiveness. *Marketing Science 30*, 3 (2011), 389 – 404.

[49] Grossman, G. M., and Shapiro, C. Informative advertising with differentiated products. *The Review of Economic Studies 51* (1984).

[50] Hale, T. The just in time consumer: How shopping trips align with economic woes.

[51] Hauser, J. R., Urban, G. L., Liberali, G., and Braun, M. Website morphing. *Marketing Science 28*, 2 (2009), 202–223.

[52] Hauser, J. R., and Wernerfelt, B. The competitive implications of relevant-set/response analysis. *Journal of Marketing Research 26*, 4 (1989), 391–405.

[53] Heckman, J. *Annals of Economic and Social Measurement*, vol. 5. National Bureau of Economic Research, Inc, 1976, ch. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models, pp. 475–492.

[54] Heilman, C. M., Nakamoto, K., and Rao, A. G. Pleasant surprises: Consumer response to unexpected in-store coupons. *Journal of Marketing Research 39*, 2 (2002), 242–252.

[55] Hoch, S., Kim, B., Montgomery, A., and Rossi, P. Determinants of store-level price elasticity. *Journal of Marketing Research 22* (1995), 17–29.

[56] Hogarth, R. M., and Reder, M. W., Eds. *Rational Choice*. University of Chicago Press Journals, 1987.

[57] Howard, R. Surviving the brave new world of food retailing: A roadmap to relevance for the future.

[58] Huff, D. L. A probabilistic analysis of consumer spatial behavior. In *Emerging Concepts in Marketing*. American Marketing Association, 1962, pp. 443–61.

[59] Hui, S., Huang, Y., Suher, J., and Inman, J. Deconstructing the “first moment of truth”: Understanding unplanned consideration and purchase conversion using in-store video tracking. *Journal of Marketing Research 50*, 4 (2013), 445–462.

[60] Imbens, G., and Wooldridge, J. Control function and related methods. In *What's New in Econometrics?* (2007).

[61] Inman, J., Winer, R., and Ferraro, R. The interplay among category characteristics, customer characteristics, and customer activities on in-store decision making. *Journal of Marketing 73*, 5 (2009), 19–29.

[62] Iyer, G., Soberman, D., and Villas-Boas, J. M. The targeting of advertising. *Marketing Science 24*, 3 (2005), 461–476.

[63] Jacoby, J. Mazursky, D. Linking brand and retailer images – do the potential risks outweigh the potential benefits? *Journal of Retailing 60*, 2 (1984), 105–122.

[64] Janiszewski, C., and van Osselaer, S. M. J. A connectionist model of brand-quality associations. *Journal of Marketing Research 37* (2000), 331– 350.

[65] John, D. R., Loken, B., and Joiner, C. The negative impact of extensions: Can flagship products be diluted? *Journal of Marketing 62*, 1 (1998), 19 –32.

[66] Jones, B., Kryczka, J., and Ménard, M. Loyalty analytics exposed: What every program manager needs to know. *PricewaterhouseCoopers* (2013).

[67] Joo, M., Wilbur, K. C., Cowgill, B., and Zhu, Y. Television advertising and online search. *Management Science 60*, 1 (2014), 56–73.

[68] Kahn, B. E., and Wansink, B. The influence of assortment structure on perceived variety and consumption quantities. *Journal of Consumer Research 30* (2004), 519 – 533.

[69] Kaldor, N. The economic aspects of advertising. *Review of Economic Studies* (1950).

[70] Kalyanam, K., and Putler, D. S. Incorporating demographic variables in brand choice models: An indivisible alternatives framework. *Marketing Science 16*, 2 (1997), 166–181.

[71] Keller, K. L., and Aaker, D. A. The effects of sequential introduction of brand extensions. *Journal of Marketing Research 29* (1992), 35–50.

[72] Kumar, V., and Leone, R. P. Measuring the effect of retail store promotions on brand and store substitution. *Journal of Marketing Research 25*, 2 (1988), 178–185.

[73] Lambrecht, A., and Tucker, C. When does retargeting work? information specificity in online advertising. *Journal of Marketing Research 50*, 5 (2013), 561–576.

[74] Lancaster, K. Socially optimal product differentiation. *The American Economic Review 65* (1975).

[75] Larson, J., Bradlow, E., and Fader, P. An exploratory look at supermarket shopping paths. *International Journal of Research in Marketing 22*, 4 (2005), 395–414.

[76] Lewis, M. The influence of loyalty programs and short-term promotions on customer retention. *Journal of Marketing Research 41*, 3 (2004), 281–292.

[77] Linden, G., Smith, B., and York, J. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Computer Society 7*, 1 (2003), 76 – 80.

[78] Lola Esteban, A. G., and Hernandez, J. Informative advertising and optimal targeting in a monopoly. *The Journal of Industrial Economics* (2001).

[79] Malthouse, E. C., . E. R. Customisation with cross-basis sub-segmentation. *Database Marketing Customer Strategy Management 14*, 1 (2006), 40–50.

[80] Mehrabian, A., and Russell, J. *An Approach to Environmental Psychology.* MIT Press, Cambridge, 1974.

[81] Messinger, P., and Narasimhan, C. A model of retail formats based on consumers' economizing on shopping time. *Marketing Science 16*, 1 (1997), 1–23.

[82] Meurer, M., and Stahl, D. O. Informative advertising and product match. *International Journal of Industrial Organization 12* (1994), 1–19.

[83] Mulhern, F. J., and Leone, R. P. Retail promotional advertising. do the number of deal items and size of deal discounts affect store performance? *Journal of Business Research 21*, 3 (1990), 179–194.

[84] Nair, H. S., Misra, H. S., Hornbuckle, H., Mishra, R., and Acharya, A. Big data and marketing analytics in gaming: Combining empirical models and field experimentation. 2013.

[85] Narasimhan, C. A price discrimination theory of coupons. *Marketing Science 3*, 2 (1984), 128–147.

[86] Neff, J. Dunnhumby: Time to ditch the demographic. *Advertising Age* (February 2013).

[87] Nelson, E., and Ellison, S. In a shift, marketers beef up ad spending inside stores. *The Wall Street Journal* (2005).

[88] Nelson, P. Information and consumer behavior. *Journal of Political Economy 78* (1970).

[89] Nelson, P. Advertising as information. *Journal of Political Economy 82* (1974).

[90] Nelson, P. Advertising as information. *Journal of Political Economy 82* (1974), 729–54.

[91] Nevo, A., and Wolfram, C. Why do manufacturers issue coupons? an empirical analysis of breakfast cereals. *RAND Journal of Economics 33*, 2 (2002), 319–339.

[92] Palmatier, R. W., Jarvis, C. B., Bechkoff, J. R., and Kardes, F. R. The role of customer gratitude in relationship marketing. *Journal of Marketing 73*, 5 (2009), 1–18.

[93] Park, C. W., Iyer, E. S., and Smith, D. C. The effects of situational factors on in-store grocery shopping behavior: The role of store environment and time available for shopping. *Journal of Consumer Research 15* (1989), 422–433.

[94] Park, C. W., Jun, S. Y., and Shocker, A. D. Composite brand alliances: An investigation of extension and feedback effects. *Journal of Marketing Research 33* (1996), 453– 466.

[95] Perrin, A. One-fifth of americans report going online 'almost constantly', 2015.

[96] Peter Breuer, L. F., and Moulton, J. Beyond the hype: Capturing value from big data and advanced analytics. *McKinsey* (2013).

[97] Raghubir, P. Free gift with purchase: Promoting or discounting the brand? *Journal of Consumer Psychology 14*, 12 (2004), 181–185.

[98] Randolph E. Bucklin, J. M. L. A two-state model of brand choice and purchase incidence. *Marketing Science 10*, 1 (1991), 24–39.

[99] Renault, R. Advertising in markets. In *Handbook of Media Economics* (2015).

[100] Rivers, D., and Vuong, H. limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics 39* (1988), 347–366.

[101] Roberts, J. H., and Lattin, J. M. Development and testing of a model of consideration set composition. *Journal of Marketing Research 28* (1991), 429–40.

[102] Rosa-Branca Esteves, J. R. Competitive targeted advertising with price discrimination. *Marketing Science* (2016).

[103] Rubinstein, A. *Modeling Bounded Rationality*. The MIT Press, 1998.

[104] Rutz, O. J., and Bucklin, R. E. From generic to branded: A model of spillover in paid search advertising. *Journal of Marketing Research 48*, 1 (2011), 87–102.

[105] Sampson, S., and Tigert, D. The impact of warehouse membership clubs: the wheel of retailing turns one more time. *International Review of Retail, Distribution Consumer Research 4*, 1 (1992), 3358.

[106] Samu, S., Krishnan, H. S., and Smith, R. E. Using advertising alliances for new product introduction: Interactions between product complimentarity and promotional strategies. *Journal of Marketing 63*, 1 (1999), 57–74.

[107] Simon, H. A. Designing organizations for an information-rich world. In *Computers, Communications, and the Public Interest* (1971).

[108] Simonin, B. L., and Ruth, J. A. Is a company known by the company it keeps? assessing the spillover effects of brand alliances on consumer brand attitudes. *Journal of Marketing Research 35* (1998), 30–42.

[109] Simonson, I. Determinants of customers' responses to customized offers: Conceptual framework and research propositions. *Journal of Marketing 69*, 1 (2005), 32–45.

[110] Smith, R. J., and Blundell, R. W. An exogeneity test for a simultaneous equation tobit model with an application to labor supply. *Econometrica 54*, 3 (1986), 679–685.

[111] Spiegler, R. *Bounded Rationality and Industrial Organization*. Oxford University Press, 2011.

[112] Stahl, D. O. Oligopolistic pricing and advertising. *Journal of Economic Theory* (1994).

[113] Stigler, G. The economics of information. *Journal of Political Economy* (1961).

[114] Tibshirani, R. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 58*, 1 (1996), 267–288.

[115] Till, B. D., and Shimp, T. A. Endorsers in advertising: The case of negative celebrity information. *Journal of Advertising 27*, 1 (1998), 67–82.

[116] Tseng, C.-H. A research of the spillover effect of gift promotion—its forming and fluctuation. *Advances in Consumer Research 37* (2010), 916.

[117] Turow, J. *The Aisles Have Eyes: How Retailers Track Your Shopping, Strip Your Privacy, and Define Your Power.* Yale University Press, 2017.

[118] Vaidyanathan, R., and Aggarwal, P. Strategic brand alliances: Implications of ingredient branding for national and private label brands. *Journal of Product and Brand Management 9*, 4 (2000), 214–228.

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

[120] Washburn, J. H., Till, B. D., and Priluck, R. *Journal of Consumer Marketing 17*, 7 (2000), 591 – 604.

[121] Zandt, T. V. Information overload in a network of targeted communication. *The RAND Journal of Economics* (2004).

[122] Zeithaml, V. The new demographics and market fragmentation. *Journal of Marketing 49* (1985), 64–75.