Push-Me Pull-You: Comparative Advertising in the OTC Analgesics Industry*

Simon P. Anderson,[†] Federico Ciliberto,[‡] Jura Liaukonyte,[§] and Regis Renault.[¶]

Incomplete and Preliminary Draft

Jan 2008: this Version: October 2009

Abstract

We investigate how firms strategically use self-promoting and comparative advertising to push up own brand perception along with pulling down the brand images of targeted rivals. To this purpose, we first watch individual video files of all TV advertisements in the US OTC analgesics industry for the 2001-2005 time period to code the content of each ad and organize it into a unique and novel dataset. Then, we develop a simple model of targeting advertising, which we use to derive the advertising first order conditions that predict oligopoly equilibrium relations between advertising levels (for different types of advertising) and market shares.

With regard to self-promotion advertising we find: i) higher market shares are associated with higher non-comparative advertising, with an elasticity of self-promoting advertising expenditures to shares estimated to be between 1 and 1.5; ii) outgoing attacks are half as powerful as direct non-comparative ads in raising own perceived quality; iii) every dollar spent by its competitors on incoming attacks has a statistically and economically strong effect on the perceived quality of the attacked brand.

With regard to comparative advertising we find: i) firms have a greater incentive to attack larger firms, and this incentive is increasing in the share of the attacker, with the elasticities of comparative advertising expenditures to own market shares and to market shares of the attacked firm equal to 1; ii) firms carry attacks on their competitors jointly.

Keywords: Comparative Advertising, persuasive advertising, targeted advertising, analgesics.

1 Introduction

This paper investigates how firms strategically use self-promoting and comparative advertising to push up own brand perception along with pulling down the brand images of targeted rivals.[1] While non-comparative

[1]The Pushmi-Pullyu is a fictitious two-headed llama befriended by Dr Doolittle. The heads are pointed in different directions. When one pushes forward, it pulls the other end back from its preferred direction.

advertising involves only positive promotion, a comparative advertisement, by comparing one's own product in favorable light relative to a rival, has both a positive promotion component (in common with non-comparative advertising) and an indirect effect through denigrating a rival. Denigration can be per se advantageous insofar as consumers who switch from the demeaned product are picked up by the denigrating firm. However, they may also be picked up by other rival firms. This logic indicates a possible free-rider situation in the provision of comparative advertising against any particular rival, but it also indicates an equilibrium at which each firm's positive promotion (through both comparative and non-comparative channels) is devalued by others' comparative advertising. In this paper we propose a simple model of targeting advertising to determine who should do more of what kind of advertising against whom, and then use a novel dataset from the Over-The-Counter (OTC) analgesics industry in the US to look for whether those relationships are actually there and how large they are.

Our push-pull model is based on a discrete choice approach to demand, in which firms' perceived qualities are shifted by advertising. The way in which advertising enters the model is most simply thought of as persuasive advertising that shifts demand up.[2] Promoting one's own product increases demand directly, whether through non-comparative advertising or comparative advertising, while denigrating a rival helps a firm indirectly by decreasing perceived rival quality.[3] By hurting the rival product directly, some consumers are diverted, and the comparative advertiser succeeds in attracting some portion of those consumers.

We use our simple model to derive the advertising first order conditions that predict oligopoly equilibrium relations between advertising levels (for different types of advertising) and market shares. In particular, we use the equilibrium pricing (first-order) conditions to eliminate prices from the relation between advertising and sales.[4] Then, we relate ad levels of the different ad types to other observable market variables, like market shares.[5]

---

[2] This is, for example, consistent with "hype" in the Johnson and Myatt (2004) taxonomy of demand shifts. We can though also reconcile our formulation with other advertising types. Most simply, the formulation is consistent with complementary advertising of the type propounded by Stigler-Becker (1977) and Becker and Murphy (1993). Indeed, one can readily append advertising in the standard discrete choice approach underpinning to the logit demand, as we present below. Alternatively, it is easy to formulate a representative consumer utility function to underlie the demand model, along the lines of Anderson, de Palma, and Thisse (1988), and introduce advertising into it.

[3] A somewhat similar approach is expounded in Harrington and Hess (1996). These authors treat positive and negative advertising by 2 politicians with given locations in a policy space. Negative advertising shifts a rival candidate away from the median voter, while positive advertising shifts a candidate closer. This framework would indeed provide an interesting base to develop a product market model.

[4] One advantage of this approach is that we bypass having to deal with price data, which involves multiple price points for multiple variants of the same brand, along with various other problems associated to price data.

[5] These variables are in turn determined simultaneously in a market equilibrium game between profit maximizing firms. Firms with a lot of advertising are also typically those with large market shares. They also tend to set high prices. This is of course not to say that high prices drive high market shares, nor, more subtly, that advertising creates high prices, nor indeed is it the high prices that create the desire to advertise. All of these variables are jointly determined, at a market equilibrium, and we show how they are determined within an industry from the firms' equilibrium choices. What drives the results is the intrinsic brand "qualities" and the marginal efficiency of advertising types across firms. See Anderson and de Palma (2001) for an analysis of how qualities correlate with market shares and prices, in a context without advertising. Here, with advertising in the choice set, and interacting with quality parameters, the results are more nuanced, though we still find some strong relations between market shares and advertising of various types.

To estimate the advertising first order conditions we first of all need to find out how much in practice is spent on comparative advertising. This is not a simple matter because advertising spending by firms, even when the data are available (which is already rare), is not broken down into comparative and non-comparative advertising. We must therefore look at each individual ad and determine whether or not it is comparative, and, if so, which is the target brand. This therefore requires a detailed coding of advertising content. Ideally, we should be able to analyze an industry for which comparative advertising is prevalent and represents a large fraction of industry sales, for which data on spending on ads is available for a full sample of firms and for a reasonably long period of time. Furthermore, video files (or audio files for radio ads or photographic files for newspaper/magazine ads) need to be available and their content readily coded for the desired information of comparison and targets. Fortunately, all these criteria are met with the US OTC industry.[6] We use data on national sales from AC Nielsen and advertising data on advertising expenditure (and movies) from TNS - Media Intelligence.

The crucial novelty of our approach is to code advertising content (focusing on comparative advertising) and organize it into a unique and original dataset.[7] We watched more than four thousands individual video files of all TV advertisements in the US OTC analgesics industry for the 2001-2005 time period and coded them according to their content. Specifically, we recorded whether the commercial had any comparative claims – whether the product was explicitly compared to any other products. If a commercial was comparative, we also recorded which brand (or class of drugs) it was compared to (e.g. to Advil or Aleve; or to Ibuprofen-based drugs).

There are two main methodological concerns that we need to address when estimating the advertising first order conditions: left-censoring of non-comparative and comparative advertising and endogeneity of market shares and advertising expenditures. Left-censoring occurs because in some periods some brands do not engage in non-comparative or comparative advertising (there are corner solutions). We control for the left-censoring by running Tobit regressions.

To control for the endogeneity of market shares and advertising expenditures, we use brand fixed effects and two sources of exogenous variation. First, we construct a dataset of news shock that hit the OTC analgesic markets in the time period of analysis.[8] These shocks might interact with the advertising decisions,

---

[6] Indeed, while explicit comparative advertising has flourished in the United States over the past 20 years (with the blessing of the FTC), its prevalence varies widely across industries. The US OTC analgesics industry (basically, medicine for minor pain relief, involving as major brands Advil, Aleve, Bayer Aspirin, and Tylenol) exhibits high advertising levels in general, and extraordinary levels of explicit comparative claims on relative performance of drugs.Most of the advertising expenditures are for television ads.

[7] See Liaukonyte (2009) for a related paper that uses this same dataset.

[8] As we discuss later on, we follow an approach similar to Chintagunta, Jiang and Jin (2007) when constructing our dataset of news shocks. In particular, between 2001 and 2005, the OTC analgesics market endured several major medical news related "shocks". The most notable, but by no means the only ones, of these were the following. The withdrawals of the Prescription NSAIDs Vioxx (October, 2004) and Bextra (April, 2005) affected the OTC NSAIDs market (which excludes Tylenol). Naproxen sodium, the active ingredient in Aleve was linked to increased cardiovascular risk, which led to a significant sales decrease for Aleve (December, 2004). The main idea here is that these shocks act as many natural experiments. The idea of using a natural

and thus we cannot use them straight up as instrumental variables. However, adding these news shocks improves our empirical analysis dramatically.

Second, we use data on the prices of the generic products to construct measures of the marginal costs that firms face to produce the corresponding branded product. Here, the generic price of a pill of Acetaminophen is used as an instrumental variable of the share of Tylenol, whose main active ingredient is Acetaminophen. Thus, the prices of the generic products are the variables that are excluded from the utility function and that we use as instrumental variables in the estimation.[9] We show that adding the news shocks remove most of the endogeneity bias we could uncover, and the exclusion restriction on the generic prices provides, in practice, only a marginal contribution to the empirical analysis.

The main results are the following. With regard to self-promotion advertising we find: i) higher market shares are associated with higher non-comparative advertising, with an elasticity of self-promoting advertising expenditures to shares estimated to be between 1 and 1.5; ii) outgoing attacks are half as powerful as direct non-comparative ads in raising own perceived quality; iii) every dollar spent by its competitors on incoming attacks has a statistically and economically strong effect on the perceived quality of the attacked brand..

With regard to comparative advertising we find that firms have a greater incentive to attack larger firms, and this incentive is increasing in the share of the attacker, with the elasticities of comparative advertising expenditures to own market shares and to market shares of the attacked firm equal to 1. This result has a nice and simple interpretation: the return to attacking a large firm is higher than the return to attacking a smaller firms, since by attacking a larger firm, the attacker can hope that a larger pool of consumers switch away from the attacked to the attacker. Similarly, a large firm has a stronger incentive than a smaller firm to attack because the probability that consumer switch to the larger firm is higher than the probability that consumers switch to the smaller firm. We also find that firms carry attacks on their competitors jointly.

The paper is organized as follows. In the next section we review the literature. Section 3 presents the theoretical model. Data and industry background are discussed in Section 4. We present the empirical specification and discuss identification of the model in Sections 5 and 6. Section 7 discusses results and Section 8 provides the robustness analysis. Section 9 concludes.

2   Literature Review on Advertising

A lot of the economics literature on the economics of advertising has been concerned with the functions of advertising, and whether market provision is optimal. We here take more of a marketer's stance that advertising clearly improves demand (otherwise firms would not do it), and we take a rather agnostic view of

experiment to study the effect of advertising (on prices) is the crucial insigth in Milyo and Waldfogel [1999].

[9]In addition we can interact these shocks with the price of the generic products and increase the number of instrumental variables that we use.

how it is the advertising actually works on individuals, and bundle it all into a single "persuasive" dimension. Since we do not cover here the normative economics of the advertising, this is excusable. The innovations we pursue are in advertising competition, and in the new strategic direction of comparative advertising.

2.1 Theoretical Literature

Much of the economic theory of advertising has been concerned with the mechanism by which advertising affects choice, and the welfare economics of the market outcome.[10] Moreover, much work has considered very particular market structures, most often monopoly.[11]

Persuasive Advertising. Much of the early work linked advertising to market power, and reached a fairly negative assessment that advertising is a wasteful form of competition. Kaldor (1950) and Galbraith (1958) saw the differentiation achieved by advertising as spurious and artificially created by persuasion. Such persuasive advertising was thought to decrease social welfare by deterring potential competition and creating barriers for new entrants. Dixit and Norman (1978), propose viewing persuasive advertising as shifting demand curves out, but they then take an agnostic view as to the welfare effects of the shift (i.e., whether the demand curve before or after the advertising is a better representation of the true consumer benefit from consuming the good).[12] Regardless, they suggest that there is a tendency for too much advertising.

Informative Advertising. The persuasive view and the idea that advertising fosters monopoly was first challenged by Telser (1964) who argued that advertising can actually increase competition through improving consumer information about products (see also Demsetz (1979)).[13] Butters (1977) later formalized a monopolistically competitive model of informative advertising about prices, in which the level of advertising reach is socially optimal. These results were tempered somewhat by Grossman and Shapiro (1984), who extended the advertising content to include (horizontal) product differentiation.[14]

Another informative role, albeit indirect information, is at the heart of "money-burning" models of signaling product quality. Nelson (1970, 1974) claims that advertising serves as a signal of quality, especially in experience good markets, and reasons that consumers will rationally conclude that a firm doing a lot of advertising must be selling a product of high quality. These insights were later formalized and further developed, most frequently by using repeat purchases as the mechanism by which a high-quality firm recoups

---

[10] See Bagwell (2009) for a comprehensive survey.

[11] Almost all the signaling literature considers monopoly, with the notable exception of Fluet and Garella (2002) who consider a duopoly. The classic Butters (1977) model of informative advertising considers monopolistic competition and a homogenous good with zero profits sent on each message. Grossman and Shapiro (1984) allow for oligopoly and product differentiation (around a circle), but they use symmetry assumptions liberally.

[12] This analysis is not uncontroversial: see the subsequent issues of the RAND journal for comments, replies, and rejoinders. Dixit and Norman (1978) posited that advertising increases demand, and then perform the welfare analysis using consumer surplus measures from that starting point, according to which demand curve embodies "true" tastes.

[13] Indeed, informative advertising can reduce consumers' search costs to learn about the existence of products, their prices, qualities, and specifications.

[14] Cristou and Vettas (2008) analyse a non-localized discrete choice version of the Grossman-Shapiro model.

its advertising investment.[15] Kihlstrom and Riordan (1984) show a role for dissipative advertising in a perfectly competitive model. Milgrom and Roberts (1986) break out different roles for signaling quality through (low) price and through advertising by a monopoly, again using a repeat purchase mechanism. Fluet and Garella (2002) show that under duopoly there must always be dissipative advertising by the high quality firm if qualities are similar enough.

Advertising as a Complementary Good. Another foundational role for advertising is proposed by Stigler and Becker (1977) and Becker and Murphy (1993), who argue that advertising can be viewed as part of consumers' preferences in the same way as goods directly enter utility functions, and that there are complementarities between advertising levels and goods' consumption. Hence, ceteris paribus, willingness to pay is higher the more a good is advertised. The complementary goods approach affords one clean way for advertising to affect directly consumer well-being, and so gives a way of thinking about persuasive advertising.

The specification we use in our model is most directly interpreted in this vein of complementary goods, insofar as we can interpret that advertising expenditures as boosting demand. However, since we will not be doing a welfare analysis with the model, we are not constrained to this interpretation, but instead our approach is broadly consistent with advertising as a demand shifter (as in Dixit and Norman (1978)).

2.2 Modeling Comparative Advertising

The theoretical economics literature on comparative advertising is quite scarce. Modeling comparative advertising presents several alternative potential approaches. In common with much of the economics of advertising, these are perhaps complementary rather than substitute approaches, and elements of each are likely present (in different strengths) in different applications. Each though has drawbacks, and sometimes the predictions (e.g., comparative static properties) differ in direction.

One early contribution is Shy (1995), who argues that comparative advertising of differentiated products informs consumers about the difference between the brand they have purchased in the past and their ideal brand. The model explains only brand switching behavior, because according to that setting comparative advertising is meaningless for the inexperienced consumer as she would not be able to comprehend an ad involving a comparison of the brands' attributes that she never consumed. Aluf and Shy (2001) model comparative advertising using a Hotelling-type model of product differentiation as shifting the transport cost to the rival's product.

Horizontal Match. Anderson and Renault (2009) model advertising as purely and directly informative

---

[15]Another mechanism is to suppose some consumers are informed already, so a low-quality firm has to distort its price so high to mimic the high-quality one that it does not wish to do so.

revelation of horizontal match characteristics of products.[16] Revelation of such information increases product differentiation, although this does not always increase firm profits. Comparative advertising in this context is modeled as revelation of characteristics (match information) of the rival product along with own characteristics. One key finding is that (under duopoly) comparative advertising is carried out by the smaller firm against its larger rival, and arises if firms are different enough.[17]

It is not immediately evident how these results extend to more firms, except insofar as an industry of roughly similar size firms would be expected to not deploy comparative advertising since individual incentives to broadcast own information should suffice. Otherwise, with firms of different sizes, there is a free-rider aspect to comparative advertising, that others (apart from the target) might benefit from it. A medium size firm might benefit from advertising relative to a large rival, but might lose relative to smaller ones. Small ones might have little to gain if indeed their small size stems from inherent disadvantages. However, it is not easy to introduce multiple firms in this context of asymmetric information divulging and hence asymmetric product differentiation.

The present model also relates the pattern of ads to market shares, but it treats the role of advertising differently. We do not model the informational content of the advertisement. Empirically we are unable to separate whether advertising was persuasive or informative, so we remain agnostic about the advertising effects and focus just on separation of comparative and non-comparative ads.

It is also important to note that the role of advertising in the Anderson-Renault (2009) model is only to divulge horizontal match information, which is two-edged sword – what characteristics one consumer likes, another dislikes. The analysis is phrased in terms of informing all consumers: it does not allow for advertising reach that tells only some. The same critique can be leveled at other models in the field, as well as (perhaps to a lesser degree) the model we actually propose here; and we return to this criticism in the conclusions.

Signaling. Another approach to modeling comparative advertising takes as staging point the signaling model of advertising, which goes back to insights in Nelson and was formalized in Milgrom and Roberts (1986). The original theory views advertising as "money-burning" expenditure which separates out low-quality from high quality producers. Equilibrium advertising spending, in this adverse-selection context, smokes out the low type because a low-type would never recuperate in repeat purchases the high level of spending indicated in

---

[16]That paper builds on Anderson and Renault (2006), who show that a monopoly firm might limit information about its product attributes even if advertising has no cost. This result identifies situations where a firm is hurt by information disclosure about its own product, so there might be incentives for competitors to provide that information through comparative ads.

[17]To understand the incentives to advertise requires understanding the benefits of more information on each firm's profits. With no information at all, firms are homogenous apart from the quality advantage, and the large firm can price out its advantage and still serve the whole market. It has no incentive to advertise because, while such advertising will raise the willingness to pay of consumers who discover they appreciate its product, it will also decrease the valuations of those who discover they like the product less than average, and so the firm will lose customers to its rival as well as having to price lower to staunch the loss of consumer base. This means that the large firm does not want to advertise, while the smaller rival does. These incentives extend to comparative advertising, which further enhances differentiation and further erodes the customer base (and price) of the larger firm to the advantage of the smaller one.

equilibrium. The comparative advertising version of this theory expounded in Barigozzi, Garella, and Peitz (2006) relies on the possibility of a law-suit to punish an untrue claim. Recently, Emons and Fluet (2008) also took a signaling approach to comparative advertising, although their analysis relies on advertising being more costly the more extreme are the claims it makes, instead of a law-suit.

Persuasion Games. In parallel work, we are developing another approach along the lines of the Persuasion Game of Milgrom (1981) and Grossman (1981). In this work the firms must (truthfully) announce levels of product characteristics their products embody. Comparative advertising, through this lens, involves announcing characteristics levels of rivals that those rivals would prefer to keep silent. However, the actual ads are quite vague for the most part in specifics of actual claims (e.g., a product may act "faster" than another, but it is not usually specified how much faster, or indeed what the response time in minutes is for the two products or the statistical significance of the difference across different individuals, etc.)

2.3   Empirical Literature

In this Section we discuss the papers that are most closely related to ours and discuss the original contributions of our paper.[18] To do this, we identify four modeling choices that have to be made when empirically studying advertising: how to measure advertising; whether to use a static or a dynamic model of advertising; whether to have a partial or a full equilibrium model, where both consumer and firm sides of the market are explicitly modeled; and whether to model advertising as having only a persuasive or informative effect, or both. Next, we discuss how the literature has dealt with these choices.

Advertising Content. Ours is the first paper to code the content of advertising into non-comparative and comparative ads and use the information to address the incentives to use the different types of advertising. [19]   Previous papers have used total ad expenditures as the sole advertising explanatory variable (notable examples are Nevo [2000,2001] and Goeree [2008]). Here, because we have data on content, we break down the ad expenditures into comparative and non-comparative expenditures, and the comparative expenditures are further broken down into attacker-target pairs. We then look at the first order conditions of the advertising decisions, and so estimate the choice of advertising of the different types from the supply side. In related work with the same data, Liaukonyte (2009) estimates a model of demand where non-comparative and comparative advertising are found to have different quantitative effects on consumer choices.

Dynamic vs. Static and Partial vs. Full Equilibrium Models. We estimate a static model of firm

---

[18]For more detail on the broader findings of the literature, see Bagwell's [2007] superb review of the empirical literature on advertising.

[19]Contemporaneous and independent work by Crawford and Molnar (2009) looks at advertising content of TV ads for Hungarian mobile telephony. They estimate a demand model, in the same fashion as Liaukonyte [2009]. Anderson and Renault (2008) study newspaper ads for airlines, and they code their content. In the former case, only 5% of ads are comparative, and even fewer in the latter case. For Hungarian telephones, much of the advertising concerns prices; in analgesics, virtually none. For airlines, mainly the low-cost carriers emphasize prices.

behavior, where firms jointly choose product prices and advertising levels. We consider a full equilibrium static model of the advertising and product markets, where advertising is determined endogenously within the model. We use the first order conditions and demand equations for the product (analgesics) to solve the prices out of the first order conditions for advertising. This procedure yields simple relations between ad levels and market shares, which we term "quasi-reaction functions" (they are not the full reaction functions because they still include market shares, which in turn depend on all prices and all advertising). We estimate the structural parameters of the model from these advertising first order conditions.

Because advertising is likely to have long-run effects on demand, the decision to use a static model to study advertising needs to be carefully justified. This modeling decision is tightly linked to another one: whether or not to have a full equilibrium model of the advertising and product markets. In short, estimating a fully dynamic equilibrium model even of just the product market is beyond what is feasible at this stage of the literature.[20] Previous work in advertising has either estimated a dynamic model of demand (Hendel and Nevo [2006] and Gowrisankaran and Rysman [2009]) or has looked at a static model of demand and a dynamic model of supply (Roberts and Samuelson [1988], Dube, Hitsch, Manchanda [2005]).

Thus, a practical choice must be made. Either one models only one side of the market in a dynamic setting and must relinquish analyzing a full equilibrium model. Or else one can analyze a full equilibrium static model. In this paper we follow the second option. Clearly, these two approaches are complementary and provide different insights into the role of advertising. Most importantly, a static model simplifies the treatment of advertising as an endogenous variable. To our knowledge, all papers that study advertising in a dynamic context treat it as an exogenous variable (notable examples are Erdem and Keane [1996], Ackerberg [2001,2003] and Dube, Hitsch, Manchanda [2005]).[21]

Persuasive vs. Informative Advertising. The last modeling choice is about the way that advertising affects consumer choice. Ideally, one would like advertising to have both an informative and persuasive effect. The informative effect has been modeled using a Bayesian learning model (Erdem and Keane [1996], Ackerberg [2003]), a limited consumer information model based on information sets (Goeree [2008]), or horizontal match information models (Anderson and Renault [2008] and Anand and Shachar [2004]). The persuasive effect is easier to model, as advertising is simply introduced into the utility function (e.g. Nevo [2001], Shum [2004].[22] There are only two papers that allow for both effects to be present, both by Ackerberg [2001,2003].

---

[20]The problem is both computational complexity and multiplicity of solutions. One would have to solve for rational and consistent expectations that consumers and producers have on the future values of the state variables, which means solving for a fixed point. There might be multiple future values of the state values for which such consistency requirements hold (that is, there might be multiple equilibria).

[21]Although the latter paper presents a dynamic theoretical model of advertising, the econometric study estimates only the demand side parameters. These estimates are then used to calibrate the theoretical dynamic model.

[22]Gasmi, Laffont, and Vuong [1992], Kadiyali [1996], and Slade [1995] postulate a set of residual demand functions, which include advertising. Thus, the interpretation of the role of advertising as persuasive or informative is not transparent.

In order to identify the persuasive from the informative role, Ackerberg [2001,2003] analyzes consumer reactions to the advertising of a new product (the yogurt Yoplait 150). Essentially, advertising is only informative for first buyers, while it is both informative and persuasive for repeat buyers.[23] This is a clever identification device, but we cannot use it here because we have aggregate and not individual data (that is, we cannot identify first buyers).[24]

Our Push-Pull perspective on advertising is coherent with the persuasive view. In addition to positive persuasion on own quality, comparative advertising also gives negative persuasion on rivals.

Review of Similar Models of Advertising. We conclude this Section with a review of the three papers which deploy models of price and advertising competition that are close to ours.[25]

Gasmi, Laffont, and Vuong [1992] propose an empirical methodology for studying various types of collusive behavior in pricing and advertising. They derive two first order conditions (for prices and advertising) and one demand equation (for the product market, cola) for each firm and estimate them all jointly.[26]

Roberts and Samuelson [1988] estimate a model where demand is modeled statically, while supply is modeled dynamically. By assuming that firms have perfect foresight of future input prices, Roberts and Samuelson end up estimating a set of first order conditions for prices and advertising, as well as demand equations. Thus, even if they start from a dynamic supply model, in practice the system of equations they estimate is quite similar to the one considered by Gasmi, Laffont, and Vuong [1992].

Goeree [2008] considers a discrete choice consumer model under limited information, where advertising influences the set of products from which consumers choose to purchase, but does not enter into the utility function. She derives first order conditions for advertising and prices, as well demand functions for the products (computers).

In many ways our approach is similar to the ones used in these three papers. We also use a theoretical model to derive the first order conditions for prices and advertising. There are, however, important differences between our work and theirs. The main methodological differences are related to how we code advertising content, how we model demand, the nature of the exogenous variation that we use to identify the model, and how we estimate the parameters of the model.

---

[23] Ackerberg (2001, 2003) argues that the observed facts that "experienced" consumers (those who have previously bought Yoplait 150) are much less sensitive to advertising than inexperienced ones is strong evidence in favor of advertising fulfilling an informative role rather than a "prestige" one. However, he does not control for the content of the particular ads in his sample; nor does he allow for the possibility (in his interpretation) that advertising 'prestige' could exhibit strong threshold effects, which could also account for the observed behavior.

[24] This identification assumption excludes the possibility that a first buyer of a new product might have consumed other products of the same brand in the past, otherwise it is unclear that there is no persuasion effect for that type of buyer. Thus, while very clever, this assumption might not hold in practice.

[25] Other papers (e.g. Shum [2004] or Nevo [2000,2001]) that use static models assume that advertising is exogenous, though they justify that assumption in their contexts. Clearly, these papers do not include first order conditions for advertising.

[26] Kadiyali [1996] proposes an empirical methodology to investigate strategic entry and deterrence, where firms compete in prices and advertising. Since she closely follows Gasmi, Laffont, and Vuong [1992], the methodological differences between her paper and ours are the same as those between our paper and Gasmi, Laffont, and Vuong [1992].

First, all three look at the total advertising expenditure, while we distinguish between comparative and non-comparative advertising expenditures.

Second, our demand (as well as Goeree's [2008]) is derived from a discrete choice model, while Gasmi, Laffont, and Vuong [1992] and Roberts and Samuelson [1988] postulate a set of residual demand functions. We have in common with Roberts and Samuelson [1988] a market expansion effect and a share effect, although we do not have the possibility that rivals' demands can rise with own advertising.

Third, we use a combination of exogenous shocks and firm-specific generic prices to construct sources of exogenous variation in the data. Instead, Gasmi, Laffont, and Vuong [1992] use aggregate variables (e.g. the price of sugar). Roberts and Samuelson [1988]) use aggregate variables (e.g. cost of capital) and the number of own and rival brands.[27] Goeree [2008] uses type of instrumental variables introduced by Bresnahan [1987]: the characteristics of the products produced by the competitors. Because we look at brands and not products, such instrumental variables cannot be used in an obvious way (brands have many differentiated products).

Finally, our estimation methodology is different from those in the other papers. While they estimate a full set of simultaneous equations, we use the first order conditions for prices to solve the prices out of the advertising first order conditions. Thus, we fully exploit the theoretical model in the same way that they do, but we reduce the number of equations to be estimated. If the model is correctly specified (which is the maintained assumption in their studies, as in ours), then the estimation results should be the same under the two approaches.[28]

3   The Model

The theoretical model suggests certain regularities between market shares and both non-comparative and comparative advertising. Notice that the predictions for non-comparative advertising hold without the more specific functional form restrictions imposed later for the comparative advertising case. These size-advertising relations therefore hold in more general settings and also even when there is no comparative advertising, and so they constitute a contribution to the understanding of the size-advertising relation which is broader than the particular comparative advertising application developed in the sequel.

We first describe the demand side assumptions and then we derive the equilibrium predictions from the model. These take the form of advertising intensities as a function of market shares, and they form the basis of the estimation which follows. As we will see, the key predictions are all supported by the data.

We assume that each product is associated to a quality index and demand depends on the quality indices

---

[27] The numbers of own and rival brands are valid instruments as long as these numbers are determined prior to price and advertising choices. This type of instrument was first proposed by Bresnahan [1987] and has been widely used since. We cannot use such brand numbers, since these are constant over the time period.

[28] In our small sample, estimating the full model would likely to lead to more precise estimates. We leave the estimation of the full model to future work.

of all firms, in a manner familiar from, and standard in discrete choice analysis. These quality indices are influenced positively by own advertising (both non-comparative and comparative) and negatively by competitors' comparative advertising. They are also influenced by medical news shocks which unexpectedly indicate good news or bad news about the health effects of the product(s).

3.1 Demand

Suppose that Firm $j = 1, ... n$ charges price p_j and has perceived quality $Q_j(.)$, $j = 1, ... n$. We retain the subscript j on $Q_j(.)$ because when we get to the econometrics, exogenous variables such as medical news shocks and random variables summarizing the unobserved determinants of perceived quality will enter the errors in the equations to be estimated.

Firms can increase own perceived quality through both types of advertising, and degrade competitors' quality through comparative advertising. Comparative advertising, by its very nature of comparing, both raises own perceived quality and reduces the perceived quality of rival products. The corresponding arguments of $Q_j(.)$ are advertising expenditure by Firm j which directly promotes its own product, denoted by A_{jj}; "outgoing" advertising by Firm j targeted against Firm k, A_{jk}, $k \neq j$, which has a direct positive effect; and "incoming" comparative advertising by Firm k targeting Firm j, A_{kj}, $k \neq j$, which has a negative (detraction) effect on Firm j's perceived quality. Thus, we write j's perceived quality as $Q_j(A_{jj}, \{A_{jk}\}_{k \neq j}, \{A_{kj}\}_{k \neq j})$, $j = 1, ..., n$, which is increasing in the first argument, increasing in each component of the second (outgoing) group, and decreasing in each component of the third (incoming) group.[29]

The demand side is generated by a discrete choice model of individual behavior where each consumer buys one unit of her most preferred good. We will not estimate this demand model from (aggregate) choice data; we simply use it to frame the structure of the demand system. Preferences are described by a (conditional indirect) utility function:

$$U_j = \delta_j + \mu \varepsilon_j, \qquad j = 0, 1, ..., n, \tag{1}$$

in standard fashion, where

$$\delta_j = Q_j(.) - p_j \tag{2}$$

is the "objective" utility, and where we let the "outside option" (of not buying a painkiller) be associated to an objective utility $\delta_0 = V_0$. The parameter μ expresses the degree of horizontal consumer/product heterogeneity.[30]

---

[29] Throughout, we assume sufficient concavity that the relevant second order conditions hold.

[30] As in Anderson, de Palma, and Thisse (1992). This parameter is especially needed whenever we specialize the model to the multinomial logit. Note that econometric specifications often set a marginal utility of money parameter (often α) before the price term, and they normalize $\mu = 1$. This is therefore effectively setting $\alpha = 1/\mu$: we do not do this here because we shall shortly substitute out price term anyway, and the intuitions are cleaner without carrying around this α.

The structure of the random term determines the form of the corresponding demand function. At first, we do not impose further structure, but we later specialize (for the comparative advertising analysis) to the logit model to get a sharper set of benchmark properties. The corresponding market shares are denoted s_j, $j = 0, ..., n$, and each s_j is increasing in its own objective utility, and decreasing in rivals' objective utilities.[31] Assume that there are M consumers in the market, so that the total demand for product j will be Ms_j, $j = 0, ..., n$.

3.2 Profits

Assume that product j is produced by Firm j at constant marginal cost, c_j.

Firm $j's$ profit-maximizing problem is:

$$\underset{\{p_j, A_j\}}{Max}\, \pi_j = M(p_j - c_j)s_j - A_{jj} - \gamma \sum_{k \neq j} A_{jk} \qquad j = 1, ...n. \qquad (3)$$

Here $\gamma > 1$ reflects that comparative advertising may be intrinsically more costly because of the risk involved that a competitor might challenge the ad and it will have to be withdrawn and replaced with a less suitable one.[32]

The advertising quantities (the A's) are dollar expenditures.[33] The idea is that advertising expenditures will be optimally allocated across media (and times of day in the case of radio/TV). Then market prices for access to eyeballs (and eyeballs of different value to advertisers) should embody the condition that there should be no systematically better/cheaper way to reach viewers. The strong form of this (efficient markets) hypothesis implicitly assumes that there are enough advertiser types, and there is no great difference in the values of consumers to OTC analgesics advertising compared to other sectors.[34]

---

[31] For example, in the standard logit model, we have $s_j = \frac{\exp[\delta_j/\mu]}{\sum\limits_{k=0}^{n} \exp[\delta_k/\mu]}$, $j = 0, ..., n$.

[32] Hosp (2007) from Goodwin Procter LLP notes that "Comparative advertising is a useful tool to promote an advertiser's goods and to tout the superior quality of the advertiser's goods over those of its competitors. Comparative advertising, however, is also the form of advertising that is most likely to lead to disputes. In undertaking comparative advertising a company should be cognizant of the potential risks and pitfalls that can lead to costly disputes and litigation. The competitor will scrutinize the advertising, and is more likely to be willing to bear the expense of litigation or dispute resolution in an instance where the competitor itself has been targeted."

More formally, suppose that a comparative ad is successfully challenged with probability $\mathbb{P}$, and that when withdrawn it must be replaced with an ad of lower effectiveness, and the effectiveness is a fraction β of that of the preferred ad. Let p^A per denote the cost of airing a non-comparative (on a particular channel at a particular time). Then the cost of airing the comparative ad is $p^A\left((1 - s_j) + s_j/\beta\right)$. If we normalize the cost of the non-comparative advertising by setting $p^A = 1$, then we have the effective comparative ad cost as $\gamma = ((1 - s_j) + s_j/\beta) > 1$.

[33] They therefore need to be deflated by an advertising price index: as long as the price per viewer reached has not changed in a manner systematically different from the general inflation rate, the CPI is a decent proxy, and will be used below.

[34] For example, suppose that each ad aired at a particular time on a particular channel cost $\hat{p}$ and delivered H "hits" (where the hit is measured in dollars). Then the equilibrium price of an ad delivering $H/2$ hits should be $\hat{p}/2$, etc.: the price per hit ought to be the same. Factoring in hits of different worth (the audience composition factor) follows similar lines. Notice though that such arbitrage arguments require sufficient homogeneity in valuations of at least some sub-set of advertising agents. The second caveat is that the arbitrage argument most directly applies to numbers of viewers hit, whereas here we deploy a demand form where ads enter a representative utility. It remains to be seen how consistent this is with an approach where heterogenous individuals (who see different numbers of ads) are aggregated up to give a market demand function (see for example Goeree (2008) for an empirical application, albeit in the context of informative ads / consideration sets).

We assume in what follows that pricing and advertising levels are determined simultaneously in a Nash equilibrium.

3.3 Firms' Optimal Choices

Pricing. Recalling that shares, s_j, depend on all the δ's, the price condition is determined in the standard manner by:

$$\frac{d\pi_j}{dp_j} = Ms_j - M(p_j - c_j)\frac{ds_j}{d\delta_j} = 0, \qquad j = 1,...n, \qquad (4)$$

which yields a solution $p_j > c_j$: firms always select strictly positive mark-ups.

Non-Comparative Advertising. The following analysis covers persuasive advertising generally, and is not confined to the specifics of the comparative advertising approach which follows.

Non-comparative advertising expenditures are determined by:[35]

$$\frac{d\pi_j}{dA_{jj}} = \frac{d\pi_j}{d\delta_j}\cdot\frac{\partial Q_j}{\partial A_{jj}} - 1 = M(p_j - c_j)\frac{ds_j}{d\delta_j}\frac{\partial Q_j}{\partial A_{jj}} - 1 \leq 0, \quad \text{with equality if } A_{jj} > 0 \quad j = 1,...,n, \qquad (5)$$

where the partial derivative function $\frac{\partial Q_j}{\partial A_{jj}}$ may depend on any or all of the arguments of $Q_j(.)$. The pricing first-order condition (4) can be substituted into the advertising one (5) to give the equilibrium conditions:[36]

$$Ms_j\frac{\partial Q_j}{\partial A_{jj}} \leq 1, \quad \text{with equality if } A_{jj} > 0, \qquad j = 1,...,n. \qquad (6)$$

Note this equation is general up to the assumption of quality not interacting with consumer type.[37] Thus, it is consistent with discrete choice models with an objective utility $u = Q - p + \mu\varepsilon$; also with a representative consumer model with $\delta = Q - p$ in the usual price places.

The interpretation is the following. Raising A_{jj} by \$1 and raising price by \$$\frac{\partial Q_j}{\partial A_{jj}}$ too leaves δ_j unchanged. This change therefore increases the revenue by \$$\frac{\partial Q_j}{\partial A_{jj}}$ on the existing consumer base (i.e., Ms_j consumers). This extra revenue is equated to the \$1 marginal cost of the change, the RHS of (6). We term the relation in (6) the non-comparative advertising quasi-reaction function. It is a function of whatever advertising variables are in Q_j (note that they all involve firm j as either emitter or target), along with j's share. This differs from a full reaction function because it still may include j's other advertising choices, and because it includes the market share, which in turn includes all prices and advertising.

The relationship in (6) already gives a strong prediction for markets where there is no comparative advertising (e.g., when comparative advertising is barred). Indeed, suppose that the perceived quality changes

---

[35] These conditions can be written in the form of elasticities. This yields Dorfman-Steiner conditions for differentiated products oligopoly; the comparative advertising conditions below can also be written in such a form.

[36] If $\frac{\partial Q}{\partial A_{jj}}$ were constant (which would arise if ads entered perceived quality linearly), then it is unlikely that the system of equations given by (6) has interior solutions. Below we (implicitly) invoke sufficient concavity of Q_j for interior solutions.

[37] This would happen in a vertical model, for example.

with advertising in the same (concave) manner for all firms. Then the firms with larger market shares will advertise more.[38] The intuition is that the advertising cost per customer is lower for larger firms. This is a useful characterization result for advertising in general: note (as per the discussion in the introduction) that it is not a causal relationship. The fundamental parameters of the model determine which firms will be large and advertise more. For example, if firms differ by intrinsic "quality" which is independent of the marginal benefit from advertising (this is the case for our parameter $\bar{W}_j$ in the econometric specification below in Section 5), then one might expect that firms with higher such quality will be those advertising more.[39] The same relation holds in the presence of comparative advertising, given some strong separability properties on $Q_j(.)$.

Proposition 1 *(Non-Comparative Advertising levels) Let $Q_j(.)$ be additively separable, and let the function $\frac{\partial Q_j}{\partial A_{jj}}$ be the same decreasing function of A_{jj} for all firms, $j = 1, ..., n$. Then, in equilibrium, firms with larger market shares will use more non-comparative advertising.*

Proof. From the relation (6), any firm which is active in non-comparative advertising will set its corresponding advertising level to satisfy $Ms_j \frac{\partial Q_j}{\partial A_{jj}} = 1$. Since $\frac{\partial Q_j}{\partial A_{jj}}$ is decreasing in A_{jj}, firms for which s_j is larger will advertise more (choose a higher value of A_{jj}) than those with smaller market shares. For firms with low enough market shares, from (4) the term $(p_j - c_j)\frac{ds_j}{d\delta_j}$ is small enough that the derivative $\frac{d\pi_j}{d\delta_j}$ in (5) is negative when $\frac{\partial Q_j}{\partial A_{jj}}$ is evaluated at $A_{jj} = 0$. ∎

Although we will not impose the strong separability in our estimation below (for reasons elucidated in Section), the Proposition is still a useful benchmark (and indeed covers the case of no comparative advertising), even though the conditions given are strong. For the model we estimate, the Proposition holds, without imposing additive separability, as long as other advertising levels are constant.

We now turn to comparative advertising levels, employing a further restriction on demands.

Comparative Advertising. The general problem is more opaque than for own ads, so we use a logit formulation. Then, assuming the idiosyncratic match terms are i.i.d. with the Type 1 Extreme Value Distribution, the market share for Firm j (fraction of consumers buying from Firm j) will be given by the logit formulation as:

$$s_j = \exp[\delta_j/\mu]/\sum_{k=0}^n \exp[\delta_k/\mu], \qquad j = 0, ..., n, \qquad (7)$$

This formulation has important properties (readily proved by simple differentiation) useful to the subsequent development. First, cross effects are given as:

---

[38] In this case, $Ms_j Q'(A_{jj}) = 1$, is the first order condition, with (temporarily) $Q(.)$ the production of quality from advertising. Clearly, the larger is the share, the smaller must be Q', and hence the higher must be ads. Note we did not use any symmetry property of the share formula: what did all the work was the same Q' function.

[39] This indeed can be shown to be the case in some specifications of the model.

$$\frac{ds_j}{d\delta_k} = -\frac{s_j s_k}{\mu}, \qquad j = 0, ..., n, \quad j \neq k, \tag{8}$$

which is also the expression for $\frac{ds_k}{d\delta_j}$ (such symmetry is a general property of linear random utility models: see Anderson, de Palma, and Thisse, 1992, Ch. 2, for example).

Second, the own effect is readily derived as:[40]

$$\frac{ds_j}{d\delta_j} = \frac{s_j(1 - s_j)}{\mu}, \qquad j = 0, ...n, \tag{9}$$

Using this expression, the price first-order condition (4) under the logit formulation is now

$$\frac{d\pi_j}{dp_j} = Ms_j - M(p_j - c_j)\frac{s_j(1 - s_j)}{\mu} = 0, \quad j = 1, ...n. \tag{10}$$

Recalling that the perceived quality is $Q_j(A_{jj}, \{A_{jk}\}_{k \neq j}, \{A_{kj}\}_{k \neq j})$, $j = 1, ..., n$, we can determine the advertising spending against rivals by differentiating (3) to get (for $k = 1, ..., n, \quad j = 1, ...n, \quad k \neq j$):

$$\begin{aligned}
\frac{d\pi_j}{dA_{jk}} &= \frac{d\pi_j}{d\delta_j}.\frac{\partial Q_j}{\partial A_{jk}} + \frac{d\pi_j}{d\delta_k}.\frac{\partial Q_k}{\partial A_{jk}} \\
&= \underbrace{M(p_j - c_j)\frac{s_j(1 - s_j)}{\mu}\frac{\partial Q_j}{\partial A_{jk}}}_{\text{own Q enhancement}} + \underbrace{M(p_j - c_j)(-\frac{s_j s_k}{\mu})\frac{\partial Q_k}{\partial A_{jk}}}_{\text{competitor's Q denigration}} - \gamma \leq 0,
\end{aligned}$$

with equality if $A_{jk} > 0$.

Inserting the price first-order conditions (10) gives (for $k = 1, ..., n, \quad j = 1, ...n, \quad k \neq j$):[41]

$$\frac{d\pi_j}{dA_{jk}} = Ms_j\frac{\partial Q_j}{\partial A_{jk}} - M\frac{s_j s_k}{(1 - s_j)}\frac{\partial Q_k}{\partial A_{jk}} \leq \gamma. \tag{11}$$

The relation between market share and comparative advertising takes a particularly clean form when the quality function embodies a perfect substitutability relation. This formulation includes the Net Persuasion form used below in the estimation. Suppose therefore that the quality function can be written as $Q_j(A_{jj}, \{A_{jk}\}_{k \neq j}, \{A_{kj}\}_{k \neq j}) = Q_j(A_{jj} + \lambda\Sigma_{k \neq j}A_{jk}, \{A_{kj}\}_{k \neq j})$, $j = 1, ..., n$, where $0 < \lambda < 1$ reflects

---

[40] These properties are related to the IIA property of the Logit model: as an option becomes more attractive, it draws customers from other products in proportion to the product of its own and their market shares.

[41] When the (pure) non-comparative advertising level is positive, its condition gives (as before):

$$Ms_j\frac{\partial Q_j}{\partial A_{jj}} = 1, \quad j = 1, ..., n.$$

Hence we can write the comparative advertising first- order condition (for positive A_{jk}) as:

$$\frac{\frac{\partial Q_j}{\partial A_{jk}}}{\frac{\partial Q_j}{\partial A_{jj}}} - \frac{s_k}{(1 - s_j)}\frac{\frac{\partial Q_k}{\partial A_{jk}}}{\frac{\partial Q_j}{\partial A_{jj}}} = \gamma, \ k = 1, ..., n, \quad j = 1, ...n, \quad k \neq j.$$

The first term on the LHS can naturally be interpreted as the marginal rate of substitution of the two ad types into perceived quality, the second term reflects the additional benefit from denigration, while the RHS is the relative price.

16

the idea that comparative advertising should not have a stronger *direct* effect than non-comparative advertising.[42] Suppose for the present argument that the solution for non-comparative ads is interior. Then, the non-comparative advertising condition $(Ms_j \frac{\partial Q_j}{\partial A_{jj}} = 1)$ implies that $Ms_j \frac{\partial Q_j}{\partial A_{jk}} = \lambda \phi$, and hence, using equation (11), we can write:

$$(0 <) - M \frac{s_j s_k}{1 - s_j} \frac{\partial Q_k}{\partial A_{jk}} \leq \gamma - \lambda. \tag{12}$$

The intuition is as follows. Raising A_{jk} by \$1 is equivalent to brand k raising its price by \$$\frac{-\partial Q_k}{\partial A_{jk}}$ (since the same δ_k is attained). Such a rival price change (which j thus effectuates through comparative advertising) causes j's market share to rise by $\frac{s_j s_k}{\mu}$. This increment is valued at $M(p_j - c_j)$. By the price first-order condition, $p_j - c_j = \frac{1}{\mu(1-s_j)}$, and (12) follows. This relation (12) generates two strong results that relate comparative advertising to market share. A sufficient condition for these results to hold is that the quality function takes one of the two following forms:

Q1. *Let the quality function be $Q_j(A_{jj} + \lambda \Sigma_{k \neq j} A_{jk}, \{A_{kj}\}_{k \neq j})$, with $Q_j(.)$ additively separable in incoming comparative ads, $\{A_{kj}\}_{k \neq j}$, with $\frac{\partial Q_j}{\partial A_{kj}}$ the same increasing function of A_{kj} for all firms, $j, k = 1, ..., n$.*

Q2. *Let the quality function be $Q_j(A_{jj} + \lambda \Sigma_{k \neq j} A_{jk} - \phi \Sigma_{k \neq j} A_{kj}, \{A_{kj}\}_{k \neq j})$, with $Q_j(.)$ additively separable in Net Persuasion, $A_{jj} + \lambda \Sigma_{k \neq j} A_{jk} - \phi \Sigma_{k \neq j} A_{kj}$, and incoming comparative ads, $\{A_{kj}\}_{k \neq j}$. Denote the marginal effect of A_{kj} on Q_j that does NOT come through Net Persuasion as $\frac{\partial Q_j^{Push}}{\partial A_{kj}} < 0$, and assume this is the same increasing function of A_{kj} for all firms, $j, k = 1, ..., n$.*

We are now ready to state the targeting share results.

Proposition 2 *(**Larger target more**) Let the quality function satisfy either Q1 or Q2. Then, in equilibrium, for all firms using a strictly positive level of non-comparative advertising, larger firms will use more comparative advertising against each target.*

Proof. Consider first firms using a strictly positive level of comparative advertising against target k. Then (12) holds with equality, i.e.,

$$-M \frac{s_j s_k}{1 - s_j} \frac{\partial Q_k}{\partial A_{jk}} = \gamma - \lambda.$$

We now consider the two different Q specifications.

Q1. For any given target k, note that the ratio $\frac{s_j}{(1-s_j)}$ on the LHS above is decreasing in market share, s_j. Hence $\frac{\partial Q_k}{\partial A_{jk}}(< 0)$ must be higher the larger is s_j, and the corresponding A_{jk} must be larger since $\frac{\partial Q_k}{\partial A_{jk}}$ is increasing and the same for all firms. For firms with low enough market shares, from (4) the term $(p_j - c_j)\frac{ds_j}{d\delta_j}$ is small enough that (12) holds with strict inequality when $\frac{\partial Q_k}{\partial A_{jk}}$ is evaluated at $A_{jk} = 0$.

Q2. We can break down the term $\frac{\partial Q_k}{\partial A_{jk}}$ into two parts, the one through Net Persuasion, and the other through the direct Push effect. The former is equal to $-\phi \frac{\partial Q_k}{\partial A_{kk}}$ while the latter is $\frac{\partial Q_k^{Push}}{\partial A_{jk}}$, which is assumed to be negative. Then we have, by substitution, $-M \frac{s_j s_k}{1-s_j} \left[-\phi \frac{\partial Q_k}{\partial A_{kk}} + \frac{\partial Q_k^{Push}}{\partial A_{jk}} \right] = \gamma - \lambda \phi$; recalling

---

[42] The Net Persuasion form used below has $Q_j(A_{jj} + \lambda \Sigma_{k \neq j} A_{jk}, \{A_{kj}\}_{k \neq j}) = Q_j(A_{jj} + \lambda \Sigma_{k \neq j} A_{jk} - \omega \Sigma_{k \neq j} A_{kj}, \{A_{kj}\}_{k \neq j})$.

that $Ms_k \frac{\partial Q_k}{\partial A_{kk}} = 1$ when k engages in non-comparative advertising, then this equation which determines comparative advertising becomes

$$\frac{s_j}{1 - s_j} \left[\phi - Ms_k \frac{\partial Q_k^{Push}}{\partial A_{jk}} \right] = \gamma - \lambda. \tag{13}$$

This yields the comparative advertising quasi-reaction function for the case at hand. For any given target k, the ratio $\frac{s_j}{(1-s_j)}$ on the LHS is decreasing in market share, s_j. Hence $\frac{\partial Q_k^{Push}}{\partial A_{jk}} (< 0)$ must be higher the larger is s_j, and the corresponding A_{jk} must be larger since $\frac{\partial Q_k^{Push}}{\partial A_{jk}}$ is increasing and the same for all firms. ∎

This follows from the logit property that the fall-out is greater from peeling off consumers from a larger rival. This suggests that the largest brands will also be those attacked most (Tylenol in our industry context.) The property also extends to the case when the quality function depends on net persuasion and incoming attacks.

Looking from the perspective of attack targets as a function of attacker size, we have:

Proposition 3 *(**Larger targeted more**) Let the quality function satisfy either Q1 or Q2. Then, considering attacks from firms with positive levels of non-comparative advertising, in equilibrium, larger firms suffer more attacks from each rival.*

Proof. For Q1, the proof is analogous to that of Proposition 2, noting that for any given rival j, the LHS of (12) is increasing in market share of the firm attacked, s_k. For Q2, the result follows from (12) by noting (on the LHS) that the larger is s_k, then the smaller must be $-\frac{\partial Q_k^{Push}}{\partial A_{jk}}$, which in turn means that A_{jk} must be larger. ∎

Before turning to the econometric specifications, we first discuss the data: note in particular that Table 2 below roughly supports the two preceding Propositions.

4 Description of Industry and Data

The OTC analgesics market is worth approximately \$2 billion in retail sales per year (including generics) and covers pain-relief medications with four major active chemical ingredients. These are Aspirin, Acetaminophen, Ibuprofen, and Naproxen Sodium. The nationally advertised brands are such familiar brand names as Tylenol (acetaminophen), Advil and Motrin (ibuprofen), Aleve (naproxen sodium), Bayer (aspirin or combination), and Excedrin (acetaminophen or combination). **Table 1** summarizes market shares, ownership, prices and advertising levels in this industry.[43]

---

[43] We exclude Midol and Pamprin from the sample because they are both aimed more narrowly at the menstrual pain-relief market and they both have small market shares.

We use three different data-sets: (1) sales (2) advertising, and (3) medical news data. Sales and advertising data were collected by AC Nielsen and TNS - Media Intelligence respectively, and we coded the advertising content. We constructed the medical news data-set from publicly available news archives.

4.1 Sales Data

The product level data consist of average prices, dollar sales, and dollar market shares (excluding Wal-Mart sales) of all OTC oral analgesics products sold in the U.S. national market during the 5 years from March of 2001 through December of 2005 (a total of 58 monthly observations).[44] Products vary in package size (the number of pills) and the strength of the active ingredient in milligrams.

Episode of Pain. We construct a measure of a *serving* of pain medication, or an *episode of pain*, so that we can aggregate across different package sizes and across different medication strengths.

First, we assign to each analgesic product in the sales dataset the strength of its active ingredient in milligrams. To do so, we combined the descriptive data in the Nielsen dataset with the data of milligrams of a specific active ingredient in a specific formula.[45] Since the strength information was given, we were able to match the milligrams of active ingredients of the products in our dataset with the products found on the brands' websites. From the amount of milligrams of the active ingredient we derived the maximum number of pills that a consumer can take of each particular product in 24 hours.[46]

We define the unit of consumption as an episode of pain. An episode of pain is given by the maximum number of pills (for OTC consumption) an individual can take over 24 hours, as defined and required by the FDA (e.g. 3 in the case of Aleve, and from 6 to 12 for Tylenol, depending on the acetaminophen formula) times the average number of pain days per month in the population. The average monthly number of pain days is three.[47]

Market Size, Brand Market Shares and Prices. The definition of market size follows immediately from this: we define the *market size* for OTC analgesic products as the US population 18 years or older. Then, we can compute each brand's market share as the fraction of total number of episodes of pain sold

---

[44]We have data on essential product attributes noted on the packages and the fraction of products sold of each such type: active ingredient, strength (regular, extra strength, etc. - as regulated by the FDA), pill type (caplet, tablet, gelcap, etc.), number of pills contained in the product, and purpose (menstrual, migraine, arthritis, general, children, etc.), although in the end we did not use these data. In this paper we look at the strategic interaction among brands, rather than products.

[45]In the case of Ibuprofen- and Naproxen Sodium- based pain relievers, the assignment was straightforward, since these OTC products can come only in 200mg (for Ibuprofen) and 220mg (for Naproxen Sodium). In the case with Aspirin and Acetaminophen, the situation is more delicate, since these products can come in varying strengths and as a combination with other analgesic agents.

[46]For a certain analgesic drug to be sold as an OTC drug, FDA requires that the daily (24 hours) dosage does not exceed a certain threshold (the thresholds are different for different active ingredients. For example, for acetaminophen the daily dosage is 4000 mg of this active ingredient). Recall, that the maximum number of pills that one is allowed to take in a day (according to FDA standards) is a crucial variable in defining the market share of a product.

[47]This information is from the Morbidity and Mortality Weekly Report, Centers for Disease Control and Prevention, Feb 27, 1998/47(07);134-140.

over the market size. The average price of an episode of pain is computed as the ratio of the total sales by a brand divided by the total number of episodes of pain sold in a month.

Generic Prices. Next, we construct the generic product price information which we use as the exogenous variation in our instrumental variable approach. For each month we calculate the average price of the unit of episode of pain relief for the generic brands. The resulting output is the time series of average prices of episodes of pain relief for each of the four active ingredients for the generic products. We interpret the generic prices as proxies of the marginal cost of providing care to an episode of pain.

4.2 Advertising Data

Our advertising dataset is from TNS-Media Intelligence and data is reported on monthly basis. The advertising data contain monthly advertising expenditures on each ad, and video files of all TV advertisements for the 2001-2005 time period for each brand advertised in the OTC analgesics category. The unit of observation in the raw dataset is a single ad. There are more than four thousands different ads. For each ad, we know the amount spent in each month and the number of times that creative was shown during the specific month. Each ad is also associated with a video file

Advertising Content. As discussed in the Introduction, we watched all the ads and coded according to their content. Specifically, we recorded whether the commercial had any comparative claims – whether the product was explicitly compared to any other products. If a commercial was comparative, we also recorded which brand (or class of drugs) it was compared to (e.g. to Advil or Aleve; or to Ibuprofen-based drugs).If an ad had no comparative claims, it was classified as a non-comparative ad. We are then able to gather information on the advertising relationships between all potential pairs of brands.[48] The unit of observation is a year-month-brand-attacked brand combination. For example, a line in this dataset tells how much Advil spent on comparative advertising against Tylenol in March 2004 Each month has thirty six pair combinations.

Indirect Attacks. One delicate issue is how to deal with indirect attacks. An indirect attack occurs when one brand, say Tylenol, makes a claim against "all other regular" brands.[49] Because it is not clear how to deal with this type of ads, we consider two solutions. First, we consider the case where indirect attacks are equivalent to direct attacks (e.g. Tylenol on Advil), but are divided among all the brands falling within the attacked category. So, for example, when Tylenol makes a claim against "all other regular" brands, each one of the other five brands is being attacked the amount of dollars spent on that advertisement divided by five.[50] Second, we consider the case where indirect attacks should simply be interpreted as self-promotion

---

[48]We also include combinations that never see any attack. For example, Advil never attacks Motrin.

[49]Or it could be an attack against NSAIDs (Non Steroidal Anti-Inflammatory drugs, which are all drugs in our sample except those with acetaminophen as an active ingredient).

[50]Because McNeil owns both Motrin and Tylenol and Bayer also owns Aleve, we consider both the case of "independent"

ads. We look at this second case in the Robustness section.

The Attack Matrix. **Table 2** presents the complete picture of cross targeting and the advertising expenditure on each of the rival brand targeting. This table shows *every* nationally advertised brand used comparative advertising during the sample period. However, the brands against which comparisons were made are only a subset of the nationally advertised brands. The targets are the "big Three:" Tylenol, Advil, Aleve, plus Excedrin.[51] Notice that these data provide some informal support for Propositions 2 and 3. The entries on the diagonal are zeroes through not attacking oneself.

4.3 News Shocks

Between 2001 and 2005, the OTC analgesics market endured several major medical news related shocks. We follow an approach similar to Chintagunta, Jiang and Jin (2007) to collect the data on these shocks. We used Lexis-Nexis to search over all articles published between 2001 and 2005 on topics related to the OTC analgesics industry.[52]

Definition of a News Shock. We recorded the article name, source and date. From a data-set of articles we then constructed a data-set of news shocks. First, multiple articles reporting the same news were assigned to a unique shock ID. Second, we checked whether a news shock was associated with any new medical findings that were published in major scientific journals. As a result of this data cleaning, our news shock data-set includes 16 news shocks between March of 2001 and December of 2005.

Major vs Minor Shocks. We classified the shocks by their impact. If a news shock was reported in a major national newspaper (USA Today, Washington Post, Wall Street Journal, New York Times), then we classified it as a major shock. Otherwise we classified it as a minor shock. This classification is useful to verify whether our identification strategy is robust to changes in the way we define news shocks. **Table 3** reports the news shocks, by their title, date, scientific publication, and impact (Major or Minor).

Measuring the Effect of the News Shocks. For each shock that happened during period t we construct a dummy variable which is equal to 1 in all the periods after and including t: t; t + 1; ...; T [53]. In the empirical analysis below, we interact each of the major shocks listed in **Table 3** with brand dummies. This

---

brands and "multi-brands" firms (stablemates). In the case of multi-brands firms, we assume that indirect attacks carried out by a firm do not negatively affect its other stablemate brands. Thus, for example, we maintain that an indirect attack by Tylenol on all the NSAIDs is not perceived by the consumer as an attack on Motrin. Section () deals with the multi-brand specification.

[51]Motrin does not attack Tylenol because the parent company is the same; likewise, Bayer does not attack Aleve for the same reason. However, we have effectively ignored these multi-product firm relations in the data.

[52]The keywords that we used consisted of brand names, such a "Aleve," "Tylenol," "Advil," "Vioxx," and the names of their active ingredients, such as "Naproxen," or "Acetaminophen." Then we made searches using generic terms such as "pain killers" or "analgesics."

[53]We experimented with allowing shocks to depreciate over time at varying rates, but found out that the version without depreciating had a better explanatory power. Also, allowing shocks to affect brands only in the short term (varying number of periods after the shock happened) did not prove to be an effective strategy as well.

approach enables us to let the data determine whether a medical news shock affected the demand (instead of us arbitrarily assigning which shock affected which brand in which way), and, if it did, whether a shock had a positive or negative effect on that brand. **Figure 1** presents the occurrence of the *eight* major shocks, highlighting the reaction of sales and advertising to those medical shocks.

5   Econometric Analysis

Here we first discuss the quality function upon which we base the empirical analysis. Then we illustrate the equations that we want to estimate. Finally, we deal with the sources of exogenous variation in the data that identify the parameters of the model.

5.1   A Quality Function

Quality Function.   We separate out the advertising contribution to perceived quality from the intrinsic, or "base quality." That is, we write

$$Q_j\left(.\right) = \bar{Q}_j\left(.\right) + \bar{W}_j,$$

where only $\bar{Q}_j\left(.\right)$ depends on advertising levels, and $\bar{W}_j$ is a variable specific to firm j which affects quality with no interaction with j's advertising.

Equations (6) and (11) implicitly define A_{jj} and A_{jk}, when holding with equality, as functions of market shares and other advertising levels. Using the implicit function theorem, the partial derivatives of these implicit functions may be derived from second derivatives and cross partials of Q with respect to the various ad levels. In what follows, we assume that Q is quadratic so these second derivatives are constants.[54] This in turn allows us to estimate (6) and (11) in a simple manner.

After *extensive* experimentation, we chose the following functional form for the base quality:

$$Q_j = \begin{aligned} & -\left(\bar{A}_j - \alpha_1\left(A_{jj} + \lambda \sum_{k\neq j} A_{jk}\right)\right)^2 - \phi\left(\bar{C}_j - \left(A_{jj} + \lambda \sum_{k\neq j} A_{jk}\right)\right)\sum_{k\neq j} A_{kj} \\ & + \sum_{k\neq j}\left(\bar{A}_{kj} - \alpha_2 A_{kj}\right)^2 - \beta \sum_{k\neq j}\sum_{k'\neq j, k'\neq k} A_{kj} A_{k'j} \end{aligned} \tag{14}$$

This quality function has five crucial properties.

First, this function is a combination of push and pull effects of advertising. The push effect is given by the weighted sum of the self-promotion advertising and the outgoing comparative advertising $(A_{jj} + \lambda \sum_{k\neq j} A_{jk})$. The pull effect is given by incoming advertising comparative ads (A_{kj}). Here, λ is a substitutability parameter of outgoing comparative ad with (outgoing, of course) non-comparative ads. In other words, λ measures how much must be spent on non-comparative advertising to replace \$1 spent on comparative advertising to generate the same "push" in (own) perceived. For example, $\lambda = 0.75$ means that the firm can raise its

---

[54]Thus, this quadratic form can be thought of as the quadratic approximation to the true quality function around the equilibrium point.

perceived quality by the same amount if it spends 1.33 dollars in comparative advertising or 1 dollar in non-comparative advertising. This parameter does not represent the full effect of comparative advertising relative to non-comparative advertising, as there is also the Pull effect which is directly denigrating the perceived quality of targeted competitors' brands. Were we to find $\lambda = 1$ then comparative and non-comparative advertising would have the same effect on the perceived quality of a brand. If $\lambda \neq 1$, then we should conclude that comparative and non-comparative advertising have different effects and should be coded separately. We expect $\lambda \in (0, 1)$ so that outgoing attacks increases the perception of own quality, although less effectively than non-comparative ads.

Second, this functional form ensures, under appropriate parameter conditions that the quality function is increasing and concave in the push effect and decreasing and convex in the pull effect. In particular, we need all parameters expressed in Greek letters to be positive. The concavity and convexity ensure that the maximization problem is well behaved. Regarding the interaction between ad levels, this specification has the property that outgoing comparative advertising expenditures by j are perfect substitutes with non-comparative advertising by j where λ is the marginal rate of substitution between the two. Thus, λ tells us how substitutable are own outgoing ads for self-promotion ads.

Third, treating incoming attacks as perfect substitutes as well would have very strong implications on the equilibrium levels of comparative advertising used against a given target. In any given period, each firm should be attacked by at most one of its competitors. The data suggests that this is not the case. Hence we introduce a term involving a non-constant marginal rate of substitution between attacks, the term involving the sum of the $\left(\bar{A}_{kj} - \alpha_2 A_{kj}\right)^2$.

Fourth, in order to allow for some positive reinforcement between attacks by different firms on the same target, the quality function has a positive cross derivative between the different attack levels A_{kj}. This cross effect is embodied in the last term with parameter β.

Finally, we capture the potential interaction between the push effect of own advertising and the pull effect of comparative advertising aimed at firm j through the second term with parameter ϕ. Here the interaction term is written to reflect a positive cross derivative, which is consistent with the estimation results.

Advertising Allure and Base Quality Variables. By contrast to the $\bar{W}_j$, the A variables with overbars interact with their corresponding advertising levels, and determine the marginal efficiency of non-comparative and comparative advertising. For example, the higher is $\bar{A}_{jj}$, the lower is the marginal efficiency of non-comparative advertising; while the higher is $\bar{A}_{kj}$, the lower the marginal efficiency of attacks by k against j, in the sense of less incremental pull-down. In the econometric specification, both types of variables will depend on some of the observed variables (for example news shocks) as well as some of the random shocks. Here, we refer to the $\bar{W}$ variables as base quality, while the $\bar{A}$ variables are called advertising base allure.

5.2 The Equations to Be Estimated

Self-Promotion. After taking the derivative with respect to A_{jj} of equation (14) we find the non-comparative ad equations:

$$A_{jj} = \max \left\{ -\bar{A}_{jj}^* - \frac{\alpha^*}{Ms_j} - \lambda \sum_{k \neq j} A_{jk} + \phi^* \sum_{k \neq j} A_{kj}, 0 \right\}, \quad j = 1, ..., n. \tag{15}$$

where $\bar{A}_{jj}^* = \frac{\bar{A}_{jj}}{\alpha_1}$, $\alpha^* = \frac{1}{2\alpha_1^2}$, $\phi^* = \frac{\phi}{2\alpha_1^2}$. We write this equation in terms of the quasi-structural parameters α^*, ϕ^*, and λ to facilitate the discussion of the statistical significance of the results. Then, we will use the results for α^* and ϕ^* to learn about the structural parameters α_1 and ϕ. One of the advantages of writing the regressions in terms of the quasi-structural parameters is that it illustrates how our regressions are about very intuitive relationship. Here, self-promoting advertising is a linear function of the shares of the brand; its outgoing comparative advertising and its incoming comparative advertising.

Comparative Advertising. Using the perfect substitutes property of the functional form, we have simply $Ms_j \frac{\partial Q_j}{\partial A_{jk}} = \lambda$. For the maintained quality function we have:

$$A_{jk} = \max \left\{ -\gamma^* \frac{1 - s_j}{Ms_j s_k} - \beta^* A_{kk} - \omega^* \sum_{l \neq k} A_{kl} + \varphi^* \sum_{l, j \neq k} A_{lk} + \bar{A}_k^*, 0 \right\}, \quad j = 1, ..., n. \tag{16}$$

where $\gamma^* = \frac{-(\gamma - \lambda)}{2\alpha_2^2}$, $\beta^* = \frac{\phi}{2\alpha_2^2}$, $\omega^* = \frac{\lambda \phi}{2\alpha_2^2}$, $\varphi^* = \frac{\beta}{\alpha_2^2}$, $\bar{A}_k^* = 2\alpha_2 \bar{A}_{jk} + \phi \bar{C}_k$. Again, notice that we write the comparative advertising equation as a function of the quasi-structural parameters γ^*, β^*, ω^*, and φ^* to facilitate the discussion of the statistical significance of the results. Then, we will see how these results can be translated into information about the deep structural parameters α^2, γ, and β. As discussed above, this regression is also about very intuitive relationships: outgoing comparative advertising is a function of the shares of the attacking and of the attacked firm; of the self-promotion advertising of the attacked brand; of the outgoing comparative advertising of the attacked brand; and of the other brands' incoming attacks on the attacked brand.

Notice that there are some cross equation restrictions. In particular, if we estimate the coefficient of $\sum_{k \neq j} A_{kj}$, $\frac{\phi}{2\alpha_1^2}$, to be positive in the self-promotion first order condition, then the coefficient of A in the comparative first order condition, $-\frac{\phi}{2\alpha_2^2}$, must be negative. These restrictions provide useful testable hypotheses.

6 Identification

We estimate the equations (15) and (16). As mentioned in the Introduction, there are two main concerns that we need to address: left-censoring of non-comparative and comparative advertising and endogeneity of

market shares and advertising expenditures. Left-censoring occurs because in some periods some brands do not engage in non-comparative or comparative advertising (there are corner solutions). Hence the variables A_{jjt}, A_{jkt}, $j, k = 1, ..., n$, are left-censored.[55] We control for the left-censoring by running Tobit regressions.

The Nature of Endogeneity. The endogenous variables are A_{jjt}, A_{jkt}, A_{kjt}, s_{jt}, $j, k = 1, ..., n$.[56] To clarify the nature of the endogeneity in our analysis, we start from equation (15). To further simplify the discussion we assume, just for the sake of exposition, that $\lambda = 0$ and $\phi^* = 0$. We will drop these two assumptions at the end of this section. Then (15) becomes, with the appropriate time subscripts:

$$A_{jjt} = \max \left\{ -\bar{A}_{jjt} - \frac{\alpha^*}{M s_{jt}}, 0 \right\}.$$

The term $\bar{A}_{jjt}$ captures the advertising base allure of a brand, which we write as follows:

$$\bar{A}_{jjt} = Z'_{jt} \Phi + \xi_{jt},$$

where Z_{jt} are observable determinants of the advertising base allure. In this paper, these are the news shocks. The ξ_{jt} are unobservable shocks to the base allure, so ξ_{jt} is a structural error. Notice that ξ_{jt} is here assumed to be observed by firms, but not by the econometrician.

Next, recall that the market share for brand j is written as:[57]

$$s_{jt} = \frac{\exp[\delta_{jt}/\mu]}{\sum\limits_{k=0}^{n} \exp[\delta_{kt}/\mu]}, \qquad j = 0, 1, ..., n$$

where

$$\delta_{jt} = \bar{Q}_{jt}(.) - p_{jt} + \bar{W}_{jt}. \tag{17}$$

Because firms observe ξ_{jt} when they choose advertising and because shares are a function of advertising (through Q, the perceived quality), then shares are a function of ξ_{jt}, and thus we will get inconsistent estimates of α^* and Φ if we run the following simple Tobit regression:

$$\begin{cases} A^*_{jjt} = -Z'_{jt} \Phi - \frac{\alpha^*}{M s_{jt}} - \xi_{jt}, & \xi_{jt} \sim N\left(0, \sigma^2\right) \\ A_{jjt} = \max\left(A^*_{jjt}, 0\right). \end{cases} \tag{18}$$

Top Brands vs. Other Brands. The first step to address the endogeneity of the market shares is to exploit the panel structure of our data to account for time-constant differences across brands. Essentially, we model the unobservable ξ_{jt} as follows:

$$\xi_{jt} = \bar{\xi}_j + \Delta \xi_{jt},$$

---

[55] As noted above, there are two brands, Pamprin and Midol, which are primarily menstrual formulations, and that we exclude them from the empirical analysis because of their negligible market shares. Interestingly, they never engage in non-comparative advertising, only in comparative advertising. Generic brands never engage in any type of advertising.

[56] Notice that prices, which are also endogenous, have been substituted out in the equations to be estimated.

[57] Notice that the generics are included here: abusing notation, generic drugs can be funneled into multiple options 0. However, as we will show later, we do not need to estimate the demand functions to estimate the relevant structural parameters.

where $\bar{\xi}_j$ is a brand fixed effect, while $\Delta\xi_{jt}$ are time specific idiosyncratic shocks. We have investigated various specifications for the fixed effects, and concluded that a specification where there are two fixed effects, one for the top brands (Advil, Aleve, Tylenol), and one for the other brands (Excedrin, Motrin, Bayer) fits our data best.[58] We provide in **Figure 2** a graphical description of the relationship between non-comparative advertising and market sales (Ms_j) for all brands and months.

Figure 2 shows that there are two types of brands in the market. Aleve, Advil, and Tylenol (the 'Top Brands') control large market shares compared to Excedrin, Bayer, and Motrin. This is consistent with the reported weighted market share descriptive statistics in **Table 1** . This observation parallels the economic intuition that 'Top Brands' have a larger advertising base allure which translates into larger inherent quality, $\bar{A}_{jj}$. Additionally, the linear fit between shares and non-comparative advertising has the same slope for the 'Top Brands' and the rest of the brands. We use the evidence from this figure to justify the construction and use of a dummy variable 'Top Brand'.

One route is then simply to specify conditions under which there is no remaining correlation, and proceed directly to the estimates. This is the essence of Assumption 1. If this is untenable, various exclusion restrictions can remove residual endogeneity. These are described in Assumptions 2. In our regressions, we will start with estimates under the simple Assumption 1, and then proceed to deploy the other Assumption. (Note that Assumption 1, if correct, obviates the other).

Using Timing to Identify the Parameters. The parameters of the regression (18) can be identified when $\Delta\xi_{jt}$ and $\frac{1}{s_{jt}}$ are uncorrelated by estimating a variant of (18) where the ξ_{jt} are allowed to have different means corresponding to the brand-group fixed effects. The (non-)correlation condition can be given a justification, paralleling a standard assumption in a large part of the literature estimating production functions with a particular assumption on the timing of the realizations of the errors.[59] More specifically, a sufficient condition is the following:

Assumption 1 *After controlling for the news shocks, which we assume to enter directly through Z_{jt}, and after including brand fixed effects, the time specific idiosyncratic error $\Delta\xi_{jt}$ is uncorrelated with s_j, that is* $E\left(\Delta\xi_{jt}|s_{jt}, Z_{jt}\right) = 0.$

Clearly, the news shocks are exogenous since they require new medical discoveries, which 'surprise' both the consumers and the firms. Here, variation in the knowledge of the health properties of the products is captured by the news shocks. One standard interpretation for this maintained assumption is that we are basically able to observe all the variables that the firms take into account when taking their decisions,

---

[58] One important reason, to which we will return later on, is that brand shares change little over time (except for Aleve, which suffered large losses after the negative news shock at the end of 2004). The identification of the share effect is mostly from cross section variation.

[59] See Griliches and Mairess [1999] for an illuminating review of the literature on the estimation of production functions.

including the news shocks (e.g. the information that consumers and firms have at any point in time). This means that neither the econometrician nor the firms observe $\Delta\xi_{jt}$ before taking their advertising and pricing decisions. When this assumption is untenable, identification can be achieved using exclusion restrictions. We now discuss the identification assumption of this paper.[60]

Exclusion Restrictions. We need variables that affect advertising only through shares, but not directly. We seek variables that affect shares through prices, p_{jt}, but do not affect perceived quality (such the cost of providing care to an episode of pain).[61]. To this end we make the following identification assumption:

Assumption 2 *The prices of the generic products are set equal to their marginal costs, which are assumed to be constant. The prices of the generics enter into each branded product's market share but are excluded from the equation (18). Formally, $E\left(\Delta\xi_{jt}|p_{jt}^G, Z_{jt}\right) = 0$, where p_{jt}^G is a vector of generic prices.*

First, the marginal cost of production of a generic product must be constant; otherwise, the price of the generic would depend on the quantity produced by the branded products, and so it would not be exogenous.[62] Second, Bertrand competition and free entry among generic producers of the drugs with the same active ingredient leads to pricing at marginal cost.[63] If, as to be expected, the cost of producing generic products is highly correlated with the cost of producing branded products, then generic prices have an additional indirect impact on branded products' market shares through branded prices.

In practice, there are two basic instrumental variables for (the inverse of) each share s_{jt}: the price of the generic product that uses the same active ingredient as the brand j; and the sum of the prices of the analogous instrumental variables for its five competitors.[64] In addition, we include the interaction of the first (the generic price) and the second one (the sum of the prices of the other generics); and the squared terms of the first and the second.

Then, we interact these two instrumental variables with the news shocks. While the news shocks enter directly in the equation (18), their interactions with prices are clearly excluded from that equation.

---

[60] Notice that one could assume that the news shocks affect the utility derived by consuming that product (and its demand) but do not affect the advertising base allure, which is then assumed to be independent of the clinical properties of the active ingredients of a product. More formally, the news shocks enter into $\bar{W}_{jt}$ but do not enter into Z_{jt}. Essentially, the advertising base allure is a function of the image or reputation of a brand, and the image and reputation is independent of the medical properties of a product. This would be the case if we believed that the consumer has a full knowledge of the medical properties of a product, and thus advertising cannot change the value of such properties to the consumer. Under this interpretation, the perceived quality of a product is not a function of its medical properties and the news shocks could be used as instrumental variables. However, we do find that news shocks play an important role as predictors of the advertising decisions in our first order conditions. Thus, the evidence is against using this identification assumption.

[61] Notice that the fact we have been able to substitute out prices from the advertising first-order conditions means that we need not worry about changes in prices affecting advertising. By substituting out prices, the impact of price on advertising goes through market share.

[62] The marginal cost for pharmaceuticals is reasonably constant, in the sense that there are not increasing returns to scale .

[63] Notice that we can allow generic brands to charge prices that are higher than marginal costs as long as this is explained by local conditions that national brands do not take into account when they set their prices.

[64] For example, the instrumental variables for the share of Tylenol at time t are the (average) price of its generic version that uses the same active ingredient, Acetaminophen; and the sum of the (average) prices of the the generic price of Advil (Ibuprofen), Bayer (Aspirin), Aleve (Naproxen), Motrin (Ibuprofen), and Excedrin (Excedrin).

To implement our estimation in our non-linear models, we use control functions (Heckman and Robb [1985,1986]).[65] Our methodology follows Blundell and Smith (1986) and Rivers and Vuong (1988).

Generalizing the Identification Strategy. In the above discussion we have focused on the first order condition (15) under the assumptions that $\lambda = 0$ and $\phi^* = 0$. It is quite clear that even if we let that λ and ϕ^* to be different from zero, we can use the same instrumental variables. This is exactly what we do. Essentially, we use variation in the generic prices (i.e. production costs) and their interactions with the news shocks to identify the effect of all of our endogenous variables.[66]

7 Results

7.1 Non-Comparative Advertising

Clearly, one of the great advantages of using the functional form (14) is the transparency and simplicity of the first order condition above. We have a simple and clean relationship between expenditures on non-comparative advertising (A_{jjt}) and shares ($\frac{1}{Ms_{jt}}$), outgoing comparative advertising ($\sum_{k\neq j} A_{jk}$), and incoming comparative ad attacks ($\sum_{k\neq j} A_{kj}$).

Baseline Regression. Column 2 of **Table 4** provides the estimates of α^*, ϕ^*, and λ when we run the following simple Tobit regression:

$$\begin{cases} A_{jjt}^* = -\frac{\alpha^*}{Ms_{jt}} - \lambda \sum_{k\neq j} A_{jkt} + \phi^* \sum_{k\neq j} A_{kjt} - \xi_{jt}, \ \ \xi_{jt} \sim N\left(0, \sigma^2\right), \\ A_{jjt} = \max\left(A_{jjt}^*, 0\right). \end{cases} \tag{19}$$

The coefficient α^* is small and is not estimated very precisely. To provide an economic interpretation of the coefficient α^* we compute the elasticity of non-comparative advertising to shares:

$$e_{A_{jj},s_j} = \frac{dA_{jj}}{ds_j} \frac{s_j}{A_{jj}}.$$

---

[65] In practice, the estimation is made in two steps. First, we run the LHS endogenous variables (here market shares) on all exogenous variables, including those excluded from the second stage relationship. Then, we run the second stage regression (advertising levels here) now including the residuals from the first regression as an additional explanatory variable (the "Control Function") to all the second stage explanatory variables. For example, if we want to estimate the parameters of the non-comparative advertising first order condition (ads on sales), we first run shares on generic prices and news shocks, and compute the residuals. Then we run a Tobit where ads are explained by market share, news shocks (if not excluded) and the residuals.

[66] Thus, there are no exogenous variables that identify shares but not the other advertising variables. We know that advertisers must meet the Federal Trade Commission (FTC) standard of truthful and not misleading advertising claims. All material claims must be substantiated by a reasonable basis of support and firms need to evaluate whether their promotional message is likely to be challenged by a competitor or ad monitoring institution. Failure to have robust substantiation for a commercial may result in serious and costly consequences among which are failure to gain network approval and high litigation costs. The most common serious consequence is the publicized disruption of the ad campaign, sunk costs invested in the ad campaign and negative press related to the brand name. Over the five year period, we observe 15 OTC analgesics advertising claims challenged by the FTC, National Advertising Division (NAD), a competitor or a consumer. The problem with using these data is that the challenges are a function of the amount of advertising expenditures. So they cannot be considered exogenous in our regressions. This problem is not different from the one that it is encountered when we estimate market power and we do not have information on the marginal cost. Adding more equations (the first order condition for price and the demand equation) would let us identify α_I, γ, and β.

We find the median elasticity to be equal to 0.312, which means that a 10 percent increase in market share, s_j, implies a 3% increase in non-comparative advertising. This is clearly a fairly weak relationship.

The substitutability parameter, λ, is estimated to be 0.700. This means that each dollar spent on comparative ad increases the perceived quality of the attacking brand by the same amount as 70 cents spent on non-comparative ad. Notice that comparative advertising also pulls down the rivals, which is what we discuss next.

The parameter, ϕ^* is estimated to be 0.590. This suggests that incoming attacks do have a sizeable negative effect on the perceived quality of the attacked firm. Keeping shares constant (shares are endogenous, so they will change with changes in A_{jkt}), every dollar spent on incoming attacks requires 59 cents to mitigate.

Top Brand Dummy. As discussed in Section (**??**), one simple way to control for the endogeneity of shares and advertising expenditures is by adding the Top Brand dummy. Formally, we then have $\bar{\xi}_j = \bar{\xi}_{TB}$ for $j \in \{Advil, Aleve, Tylenol\}$ and $\bar{\xi}_j = \bar{\xi}_{OB}$ for $\{Motrin, Excedrin, Bayer\}$ (for obvious collinearity reasons, only the fixed effect for Top Brand will be reported). Given our relatively small sample, it helps to reduce the number of brand fixed effects. Another useful advantage of having such group-type fixed effects is that we avoid the incidental parameter problem that would have been there with the nonlinear Tobit regression and individual brand-specific fixed effects.[67]

This dummy controls for the Top Brands' advertising base allure advantage, so that it picks up any persistent component of such advantage. The remaining source of endogeneity in our regressions then comes from any potential correlation between $\Delta\xi_{jt}$ and s_{jt}.

In **Column 3** of **Table 4** the Top Brand fixed effect, $\bar{\xi}_{TB}$, has a negative sign, which means that the larger firms, Aleve, Tylenol and Advil have inherently higher advertising base allure than the other brands. This result is not very robust across specifications.

Confirming that the endogeneity concern is relevant in our context, **Column 3** shows that the coefficient α^*, equal to -0.212, is much larger than in **Column 2** and is estimated very precisely. The corresponding median elasticity of self-proming ads to shares is now equal to 2.208, which means that a 10 percent increase in market share, s_j, implies a 22.08% increase in non-comparative advertising. This is a very strong relationship, which confirms the suggestion of Proposition 1, that larger firms engage in more non-comparative advertising to push up perceived quality and demand.

The results for the parameters λ and ϕ^* are unchanged.

Major News Shocks. Column 4 adds on the *major* news shocks vector Z. Thus, we estimate their effects on the amount spent on non-comparative advertising by getting estimates for Φ. Under Assumption 1, we

---

[67]Notice, however, that even with individual brand specific fixed effects that incidental parameter problem would be marginal for two reasons. First, the time dimension grows over time, while the number of brands remains equal to six. Second, the incidental parameter problem is less important with a Tobit than with a Probit .

get consistent estimates of the parameters of the model. The idea is that any component of the unobservable that remains, after controlling for persistent advertising base allure advantages (picked up by the Top Brand dummy) and news shocks, is not observed by the firms before making their advertising and pricing decisions. Formally, we estomate the regression (19), where A_{jjt}^* is now written as follows:

$$A_{jjt}^* = -\frac{\alpha^*}{Ms_{jt}} - \lambda \sum_{k \neq j} A_{jkt} + \phi^* \sum_{k \neq j} A_{kjt} - Z_{jt}'\Phi - \bar{\xi}_{TB} - \Delta\xi_{jt}, \ \ \Delta\xi_{jt} \sim N\left(0, \sigma^2\right)$$

The way we deal with news shocks is the following. We interact each news shock with brand dummies for all brands. This leads to six (brands) times eight (shocks) variables to include in the regression. This way to deal with the shocks lets the data pick up which shocks had an impact on the firms' decisions and, also, it allows the shocks to have different effects on different brands. Because of the large number of variables, we do not report the results for the shocks.[68]

We now estimate α^* to be equal to -0.146. The corresponding share elasticity of non-comparative advertising is now down to 1.520 from 2.208. The dummy variable *TopBrand* is now equal to -0.097, and is not statistically significant, indicating now a lack of evidence that the top brands have better advertising base allure. The estimates of λ, equal to 0.575, and ϕ^*, equal to 0.395, are now smaller in magnitude than before. These results provide support to the idea that news shocks are important determinants of advertising decisions, and not including them in the regressions would bias the estimation results.

All News Shocks. **Column 5** includes all the news shocks, both those which were reported in media with national coverage and those which had only a limited impact on the media. The results are sufficiently different in **Columns 4** and **5** to infer that news shocks have an impact on this market even if they are not covered by the major newspapers or TV channels. We interpret this as evidence that, possibly through the advice of the more informed physicians, people in pain reduce (increase) the consumption of analgesics that are shown to have negative (positive) side effects.

λ is now equal to 0.452, down from 0.700 in **Column 1** and 0.575 in **Column 4**. As we control for more components of the unobservable, we reduce the extent to which different types of advertising expenditures are estimated to be correlated. One way to interpret this finding is that we estimate correlations that are biased upward when do not control for variables (here, news shocks) that are observed by firms and consumers. Similarly, ϕ^* is now equal to 0.367, down from 0.590 in **Column 1** and 0.395 in **Column 4**.

Generic Prices as Instrumental Variables. **Columns 6-8** include shocks as controls and use generic prices and their interactions with the news shocks as instrumental variables for the three endogenous variables: shares, outgoing comparative ads, and incoming comparative ads. Essentially we use Assumption 2

---

[68]Notice that what we are doing is different from using time dummies. Here we are using each news shock as a natural experiment which is allowed to have a different effect on the utilities of each of the six brands. The results are available from the authors.

to identify the parameters of the model. **Columns 6** and **Column 7** differ in that the first considers shares as endogenous variables to be instrumented, while the second assumes that the Top Brand fixed effects and the news shocks control for all the source of endogeneity of shares. **Column 8** shows that the results are robust to a small change in the way we enter shares in the regression equation.

In addition to the estimates estimates of α^*, ϕ^*, λ, the Top Brand fixed effects, and the constant term, we also report the coefficient estimates of the control functions associated with the endogenous variables. As discussed in Blundell and Smith (1986), statistically significant coefficients of the control functions would suggest that, even after including the Top Brand fixed effects and accounting for all the news shocks, the relationship between shares and advertising expenditures would still be estimated with a bias if we did not use generic prices as instrumental variables.

First, in **Column 6** we see that the coefficient estimate (5.261) of the control function of $\frac{1}{Ms_{jt}}$ is not statistically significant. There are two possible explanations. First, shares are indeed not endogenous, after including the Top Brand fixed effect and news shocks. Second, generic prices do not explain much of the variation in shares, and so they are not appropriate instruments for shares in the second stage. To choose between two explanations, we regress shares on just the Top Brand fixed effect and the news shocks, and we find that R^2 of this regression is equal to 0.9857. That is, there is only less than 2 percent of the variation in shares that still needs to be explained, and the generic prices explain a good fraction of it, as the R^2 of the this residual variation on the generic prices is 0.3181. Moreover, we can reject the null hypothesis that all the coefficients of the generic prices in the first stage regression are equal to zero (the F test is reported in the table as well). Thus, we conclude that the parameters associated with shares are identified off cross-sectional variation, rather than by within-brand variation, and we also conclude that the endogeneity of the shares is pretty much controlled for by the inclusion of the Top Brand fixed effects and the news shocks. Our interpretation is then that, after including the Top Brand fixed effect and news shocks, the remaining endogeneity of the shares (that is, their correlation with the remaining unobservables) is not empirically significant.

The conclusions are not much different as far as the estimates of ϕ^*, λ. We find that the generic prices do a fair job at explaining the first stage variation in outgoing comparative advertising and in incoming attacks. In particular, the F tests lead to the rejection of the null hypotheses that generic prices do not explain any of the first stage variation, and the generic prices explain a fair amount of the variation in the dependent variables that is not explained by the second stage exogenous variables (the R^2 of the residual variations are equal to 0.4483 and 0.3239).

A Brief Summary. We summarize our empirical analysis of the first order condition 15) as follows. First, Proposition 1's suggestion that higher shares, ceteris paribus, are associated with higher non-comparative advertising is confirmed. In particular, the results in **Table 4** imply the elasticity of self-promoting advertising

expenditures to shares is between 1 and 1.5.

Second, as we expected, we find evidence of a clear endogeneity of market shares (and other advertising variables) in the advertising first order conditions, which creates a substantial downward bias on the coefficient of market shares and upward bias on the coefficients of outgoing and incoming comparative advertising. We find that the inclusion of a Top Brand fixed effect and of brand-specific news shocks controls for the endogeneity in the variables. In the specifications where it is estimated precisely, the Top Brand fixed effect, $\bar{\xi}_{TB}$, has a negative sign, which means that the larger firms, Aleve, Tylenol and Advil have inherently higher advertising base allure than the other brands.

Finally, the estimates of the components of the Net Persuasion function lie within the expected ranges. Outgoing attacks are half as powerful as direct non-comparative ads in raising perceived quality. For given shares, incoming attacks draw down a brand by around 40 cents, in terms of the non-comparative ads that restore Net Persuasion.

7.2 Comparative Advertising

The second relation that we test is the comparative ad relation (16). The unit of observation now is a *pair of brands*, as we study attacks of one brand, j, on another brand, k. **Table 5** follows the same structure as **Table 4**: we start with a simple tobit specification; then, we add (pair-specific) brand dummies; then, we add shocks. We conclude the analysis by looking how the results change if we use generic prices as instrumental variables.

Baseline Regression. Column 2 of **Table 5** provides the estimates of γ^*, β^*, δ^*, and φ^* when we run the following simple Tobit regression:

$$\left\{ \begin{array}{c} A_{jkt}^* = -\gamma^* \frac{1-s_{jt}}{M s_{jt} s_{kt}} - \beta^* A_{kkt} - \omega^* \sum_{l \neq k} A_{klt} + \varphi^* \sum_{l,j \neq k} A_{lkt} + \xi_{jkt}, \ \xi_{jkt} \sim N\left(0, \sigma^2\right) \\ A_{jkt} = \max\left(A_{jkt}^*, 0\right) \end{array} \right\} \quad (20)$$

We estimate γ^* precisely and equal to 2.457. This result provides evidence in support of the theoretical model developed in Section (3). Propositions 2 and 3 predict that firms have a greater incentive to attack larger firms, and this incentive is increasing in the share of the attacker, thus γ^* should be positive. To provide a sense of the economic interpretation of this result, we can again compute elasticities. The median elasticity with respect to s_{jt}, $e_{A_{jk},s_j} = \frac{dA_{jk}}{ds_j} \frac{s_j}{A_{jk}}$, is equal to 1.580 and the one with respect to s_{kt} is equal to 1.504. That is, a 10 percent higher market share (of the attacker or of the attacked) implies that the comparative ads against that brand are higher by approximately 15 percent.

Next, we find that the coefficient estimate of β^* is equal to -0.0432 and that of ω^* is equal to -0.033. These are economically small numbers. For example, the first says that for each dollar spent by the attacked k in self-promoting advertising, and for given shares, the attacker firm j lowers its attacks by 4 cents If we

look across the columns of the table, we notice that in all the specifications these coefficients are estimated to have small magnitudes and, with the exception of **Column 2**, they are never statistically significant. Overall, the evidence in **Table 5** suggests that neither the attacked's self-promoting advertising, A_{kkt} nor its outgoing comparative ads, $\sum_{l\neq k} A_{klt}$, have a direct effect on the comparative advertising decision of the attacking firm j.

Finally, we estimate φ^* equal to 0.307. This means that firm j spends 30.7 cents attacking firm k for each dollar that the j's competitors $(l \neq k, j)$ spend attacking firm k. This is a strong effect and provides evidence that firms carry attacks in a jointly fashion.

Pair-Specific Brand Dummies. In **Column 3** we add pair specific brand dummies. In particular, we run the tobit regression (20), where ξ_{jkt} is replaced by $\bar{\xi}_{TB,TB} + \bar{\xi}_{TB,OB} + \bar{\xi}_{OB,OB} + \Delta\xi_{jkt}$, with $\Delta\xi_{jkt} \sim N\left(0, \sigma^2\right)$. Here, $\bar{\xi}_{jk} = \bar{\xi}_{TB,TB}$ if j and k are both Top Brands, $\bar{\xi}_{jk} = \bar{\xi}_{TB,OB}$ if j is a Top Brand (i.e., Advil, Aleve, Tylenol) and k is an Other Brand, and likewise for $\bar{\xi}_{OB,TB}$ and $\bar{\xi}_{OB,OB}$ (one is omitted because we include a constant term in the regression). For example, $\bar{\xi}_{TB,TB}$ is the pairwise group-fixed effect (to be estimated) if both the 'attacker', j, and the 'attacked', k, are top brands.

Most notably, γ^* is estimated quite smaller, as it is now equal to 0.867 and, correspondingly, the median elasticity with respect to s_{jt} is now equal to 0.558 and the one with respect to s_{kt} is equal to 0.531. Notice, however, that the brand dummies are all very small and two out of three are not statistically significant. This suggests that simply adding brand dummies is not an appropriate modeling choice, since it seems to be introducing just noise in the estimation.

News Shocks. Column 4 and 5 add, respectively, major and minor shocks in the tobit regression. The results are pretty analogous in these two columns. Adding the news shocks increases the magnitude of γ^*, bringing it to 1.571. Correspondingly, the median elasticity with respect to s_{jt} is now equal to 1.010 and the one with respect to s_{kt} is equal to 0.962. These numbers are remarkably close to the values of elasticities that we found in **Table 4**, where the elasticity of self-promoting advertising expenditures to shares is between 1 and 1.5.

Generic Prices as Instrumental Variables. Columns 6, 7, and **8** show the results when we instrument the endogenous variables with generic prices and their interactions with the news shocks. There are two main result out of these three columns. First, a direct comparison of **Column 6** (and **7** and **8**) with **Column 5** shows that φ^* might actually be downward biased when we do not instrument $\sum_{l,j\neq k} A_{lkt}$. It is now estimated equal to 0.443, up from 0.354. Second, γ^* is still estimated positive, confirming that firms have a greater incentive to attack larger firms, and this incentive is increasing in the share of the attacker. Notice that in **Column 8**, where the interaction term of s_{jt} and s_{kt} enters linearly, the median elasticities of outgoing comparative ads with respect to s_{jt} and with respect to s_{kt} are much.larger.

A Brief Summary. We summarize our empirical analysis of the first order condition (16) as follows. First, we find evidence in support of propositions 2 and 3, which predict that firms have a greater incentive to attack larger firms, and this incentive is increasing in the share of the attacker.

Second, neither the attacked's self-promoting advertising, A_{kkt} nor its outgoing comparative ads, $\sum_{l \neq k} A_{klt}$, have a direct effect on the comparative advertising decision of the attacking firm j.

Finally, firm j spends more than 40 cents attacking firm k for each dollar that the j's competitors $(l \neq k, j)$ spend attacking firm k, for given shares.

7.3 The Structural Parameters of the Model [In Progress]

Notice that all the structural parameters of the model are identified.

If we use the specifications with all the shocks, and without instrumenting using the generic prices, we find the following:

$$
\begin{aligned}
\alpha_1 &= 2.1320 \\
\phi &= 3.3273 \\
\lambda &= 0.451 \\
\gamma &= 1.4714 \\
\alpha_2 &= 14.2857 \\
\beta &= 5.1329
\end{aligned}
$$

Remarks:

- The marginal cost of comparative advertising is found to be equal to $\gamma = 1.4714$, while the cost of self-promoting advertising is equal to 1. This confirms our intuition that comparative ads are more expensive as they carry more risk. They carry more risk because firms might be challenged and they might have to drop an ad.

- From these numbers we can compute the marginal rate of substitutions of one type of ad for the other. λ is the $MRS_{A_{jj}, A_{jk}}$, or the marginal rate of substitution of self-promoting ads for outgoing comp ads. Finding $\lambda = 0.451$ tells us that for each dollar spent in comp ads, one can save 45.1 cents in self promotion and keep the perceived quality unchanged.

8 Robustness [IN PROGRESS]

In our view, the main issue that we have to deal with is whether by omitting dynamic effects, we introduce a bias in the estimation of the relationships between the main variables of the model. There are two related

dynamic features that our static model might be missing. First, $\bar{A}_{jj}$ and $\bar{A}_{jk}$ might be related to the goodwill of a firm, and that goodwill might depend on past advertising decision of the firm. We can check the importance of this aspect by adding lags in our regressions. Second, as Dube, Hitsch, Manchanda [2005] show in their descriptive analysis, pulsing might play an important role in advertising decisions depending on the industry that we look at. In this section, we look at these two features and check, indirectly, whether omitting them from the analysis might bias our results. Because we are just checking for the robustness of the results, we only look at the non-comparative advertising first order condition.

[TO BE COMPLETED]

Goodwill. Advertising goodwill is the idea that past advertising is like an investment over time which creates a stock at any moment. This stock, in turn, is subject to depreciation as the consmuer "forgets" past ads. If there are strong stock effects (depreciation is not quick), then firms are engaged in a dynamic game. Solving such a game and writing the appropriate structural model would be substantially more involved than the simple static model characterized above.

Here we essentially estimate the regression (19) after including the one month lagged value of A_{jjt}.

[TO BE DONE]

Pulsing. Pulsing is the phenomenon of uneven advertising levels over time. A campaign will have a specific start date, and a series of ads will be run at quite high intensity. In many industries, there is a considerable lag (or at least a lull) until the next campaign starts up (a new "media blitz). This pattern is thought more effective than running ads at a steady level, in part because of attention thresholds for individuals' perception, etc.

One very simple way to test whether pulsing occur in this industry is the following: We compare how the results change if we use quarterly instead of monthly data. Dube, Hitsch, Manchanda [2005] show very irregular episodes of advertising to test their theory of pulsing. Clearly, the more one aggregates the data over time, the less irregular the episodes of advertising become. So our idea is that if there is pulsing in our monthly data, and if accounting for pulsing would affect our results radically, then we should see sizeable differences in the estimates that we get by using quarterly instead of monthly data.

Hence we estimate the regressions (19) and (20), but with quarterly instead of monthly data.

[TO BE DONE]

9   Conclusions [Preliminary]

The paper proposes a novel oligopoly model of advertising, based on persuasive advertising which shifts ("pulls up") perceived product qualities. The model also introduces comparative advertising as having both a pull up effect on own perceived quality, and a "pull-down" effect on a targeted rival's quality. The empirical

results for the non-comparative advertising are very clean. First, half of a comparative ad constitues pure push, insofar it has the same effect on own perceived quality as half the dollar amount spent on a non-comparative ad. What happens to the other half is damage inflicted on the target of the comparative ad. First, it takes approximately 50 cents of non-comparative ads to offset every incoming dollar of attack, and that is just in terms of the net persuasion part of advertising. The other part of the harm to a rival is in the Pull-Down effect, which involves a further loss.

The (linearized) comparative ads estimates indicate that there is a strong positive effect of larger size in comparative advertising, and a much stronger (in terms of elasticity) positive effect of larger size of the target. This concurs with the theoretical predictions of the Push-Pull model, and it is apparent in the raw data that the largest target is the largest firm (Tylenol).

The effects of advertising in this Push-Pull set-up are channeled through quality differences. This gives quite a negative view of comparative ads, in the sense that there is much wasteful battling between brands to and fro just to stay afloat.[69] This feature is reminiscent of the Zero-Sum Game critique of advertising; that it serves solely to reshuffle demand and firms are better off if they could agree not to do it (they would save the expense). The critique is a fortiori true of comparative advertising, at least as modeled though the Push-Pull model. Firms would be better off if they could agree not to do it. This reason might partially explain why it is not prevalent in many industries. As a form of quasi-collusion, firms do not begin the process because they realize it might trigger responses. It is noteworthy in this regard that comparative advertising is being used more and more, coinciding with a recession, when collusion typically has more trouble surviving.

When consumers have different tastes over different characteristics, comparative advertising (done by different parties in different characteristics directions) may serve to enhance the perceived horizontal differentiation between products. This effect is closed down in the current model, but introducing it would likely give both better estimates as well as an improved perspective on the social benefits of the practice, at least insofar as the advertising informs heterogenous consumers about true product performance differences.

References

[1] Ackerberg, Daniel A. (2001), Empirically Distinguishing Informative and Prestige Effects of Advertising. *RAND Journal of Economics,* 32(2), 316-333.

[2] Ackerberg, Daniel A. (2003), Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review,* 44(3), 1007-1040.

[3] Aluf, Yana and Oz Shy (2001), Comparison advertising and competition. Mimeo, University of Haifa.

---

[69]One notable context of comparative ads is the case of negative political ads, which sometimes emphasize the negative features of rivals more than the positive features of the candidate on whose behalf the ad is aired.

[4] Anand, Bharat and Ron Shachar (2004), Advertising, the Matchmaker. *Harvard Business School Strategy Unit Working Paper No. 264206.*

[5] Anderson, Simon P. and André de Palma (2001), Product Diversity in Asymmetric Oligopoly: Is the Quality of Consumer Goods Too Low? *Journal of Industrial Economics*, 49, 113-135.

[6] Anderson, Simon P., André de Palma, and Jacques-Francois Thisse (1988), A Representative Consumer Theory of the Logit Model. *International Economic Review*, 29, 461-466.

[7] Anderson, Simon P., André de Palma and Jacques-Francois Thisse (1992), *Discrete choice of product differentiation*. Cambridge, MA: The MIT Press.

[8] Anderson, Simon P. and Régis Renault (2006), Advertising content. *American Economic Review*, 96, 93-113.

[9] Anderson, Simon P. and Régis Renault (2008), Advertising: the Persuasion Game. Mimeo, University of Virginia.

[10] Anderson, Simon P. and Régis Renault (2009), Comparative Advertising: disclosing horizontal match information. Forthcoming, *RAND Journal of Industrial Economics*.

[11] Bagwell, Kyle (2007), The Economic Analysis of Advertising. In Mark Armstrong and Rob Porter (eds.) *Handbook of Industrial Organization*, 3, 1701-1844. Elsevier, Amsterdam, North Holland.

[12] Barigozzi, Francesca, Paolo Garella, and Martin Peitz (2006): With a little help from my enemy: comparative vs. generic advertising. Mimeo, University of Bologna.

[13] Becker, Gary S. and Kevin M. Murphy (1993), A Simple Theory of Advertising as a Good or Bad. *Quarterly Journal of Economics*, 108, 942–964.

[14] Blundell, Richard W. and Richard J. Smith (1986), "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply," *Econometrica*, 54(3), 679-685.

[15] Bresnahan, Timothy F. (1987), Competition and Collusion in the American Automobile Market: The 1955 Price War. *Journal of Industrial Economics*, 35(4), 457-482.

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

[17] Chintagunta, Pradeep K., Renna Jiang, and Ginger Zhe Jin (2009), Information, Learning, and Drug Diffusion: the Case of Cox-2 Inhibitors. Mimeo, University of Maryland.

[18] Christou, Charalambos and Nikolaos Vettas (2008), On informative advertising and product differentiation. *International Journal of Industrial Organization*, 26(1), 92-112.

[19] Crawford, Gregory S. and Jozsef Molnar (2008), The Empirical Consequences of Advertising Content in the Hungarian Mobile Phone Market. Mimeo, University of Arizona.

[20] Demsetz, Harold (1979), Accounting for Advertising as a Barrier to Entry. *Jounal of Business*, 52(3), 345-360.

[21] Dixit, Avinash K. and Norman, Victor D. (1978), Advertising and Welfare. *Bell Journal of Economics*, 9(1), 1-17.

[22] Dube, Jean-Pierre, Günter Hitsch, and Puneet Manchanda (2005): An Empirical Model of Advertising Dynamics. *Quantitative Marketing and Economics*, 3(2), 107-144.

[23] Emons, Winand and Claude Fluet (2008), Non-comparative versus comparative advertising as a quality signal. Mimeo, University of Bern.

[24] Erdem, Tülin and M. Keane (1996), Decision-Making under Uncertainty: Capturing Dynamic Brand Choices in Turbulent Consumer Goods Markets. *Marketing Science,* 15(1), 1–20.

[25] Fluet, Claude and Paolo G. Garella (2002), Advertising and prices as signals of quality in a regime of price rivalry. *International Journal of Industrial Organization*, 20(7), 907-930.

[26] Galbraith, John K. (1958), *The Affluent Society.* Houghton Mifflin, New York.

[27] Gasmi, F., J. J. Laffont, and Q. Vuong (1992), Econometric Analysis of Collusive Behavior in a Soft-Drink Market. *Journal of Economics & Management Strategy,* 1(2), 277–312.

[28] Goeree, Michelle S. (2008), Limited Information and Advertising in the US Personal Computer Industry. Forthcoming in *Econometrica.*

[29] Gowrisankaran, Gautam and Marc Rysman (2009), Dynamics of Consumer Demand for New Durable Goods. NBER Working Paper 14737.

[30] Griliches, Zvi and Jacques Mairesse (1999), Production Functions: the Search for Identification. In *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium,* ed. Steiner Strom. Cambridge: Cambridge University Press.

[31] Grossman, Gene M. and Carl Shapiro (1984), Informative Advertising and Differentiated Products. *Review of Economic Studies*, 51, 63-81.

[32] Grossman, Sanford J. (1981), The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics*, 24(3), 461-83.

[33] Harrington, Joseph Jr. and Gregory D. Hess (1996), A Spatial Theory of Positive and Negative Campaigning. *Games and Economic Behavior*, 17(2), 209-229.

[34] Heckman, J.J. and Robb, R. (1985), Alternative methods for evaluating the impact of interventions: an overview. *Journal of Econometrics* 30, 239–67.

[35] Heckman, J.J. and Robb, R. (1986), Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In *Drawing Inferences from Self-Selected Samples*, ed. H. Wainer. New York: Springer. Repr. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.

[36] Hendel, Igal and Aviv Nevo (2006), Measuring the Implications of Sales and Consumer Inventory Behavior. *Econometrica*, 74(6), 1637-1673.

[37] Hosp, D. R. (2007), Weighing The Risks: A Practical Guide To Comparative Advertising. excerpted chapter in *Advertising and Marketing Litigation Best Practices*. Aspatore Books.

[38] Johnson, Justin P. and David P. Myatt (2006), On the Simple Economics of Advertising, Marketing, and Product Design. *American Economic Review*, 96(3) 756-784.

[39] Kadiyali, Vrinda (1996), Entry, Its Deterrence and Its Accommodation: A Study of the U.S. Photographic Film Industry. *RAND Journal of Economics*, 27(3), 452-478.

[40] Kaldor, Nicholas (1950), The Economic Aspects of Advertising. *Review of Economic Studies*, 18(1), 1-27.

[41] Kihlstrom, Richard E. and Michael H. Riordan (1984), Advertising as a signal. *Journal of Political Economy*, 92(3), 427-450.

[42] Liaukonyte, Jura (2009), Is comparative advertising an active ingredient in the market for pain relief? Mimeo, University of Virginia.

[43] Milgrom, Paul R. (1981): Good News, bad news: representation theorems and applications. *Bell Journal of Economics*, 12(2), 380-391.

[44] Milgrom, Paul R., and John Roberts (1986), Prices and Advertising Signals of Product Quality, *Journal of Political Economy*. 94(4), 796-821.

[45] Milyo, Jeffrey and Joel Waldfogel (1999), The Effect of Price Advertising on Prices: Evidence in the Wake of 44 Liquormart. *American Economic Review*, 89(5),1081-1096.

[46] Nelson, Phillip J. (1970), Information and Consumer Behavior. *Journal of Political Economy*, 78(2), 311-329.

[47] Nelson, Phillip J. (1974), Advertising as information. *Journal of Political Economy*, 82(4), 729-754.

[48] Nevo, Aviv (2000), Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry. *RAND Journal of Economics*, 31(3), 395-421.

[49] Nevo, Aviv (2001), Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica*, 69(2): 307-342.

[50] Rivers, Douglas and Quang H. Vuong (1988), Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models. *Journal of Econometrics*, 39(3), 347–366.

[51] Roberts, Mark J. and Larry Samuelson (1988), An Empirical Analysis of Dynamic, Nonprice Competition in an Oligopolistic Industry. *RAND Journal of Economics*, 19(2), 200-220.

[52] Shum, Matthew (2004), Does Advertising Overcome Brand Loyalty? Evidence from the Breakfast-Cereals Market. *Journal of Economics and Management Strategy*, 13(2), 241-272.

[53] Shy, Oz (1995), *Industrial Organization: Theory and Applications*. MIT Press, Cambridge, MA.

[54] Slade, Margaret (1995), Product Rivalry with Multiple Strategic Weapons: An Empirical Analysis of Price and Advertising Competition. *Journal of Economics and Management Strategy*, 4(3), 445-476.

[55] Stigler, George J. and Gary S. Becker (1977), De Gustibus Non Est Disputandum. *American Economic Review*, 67(2), 76-90.

[56] Telser, Lester G. (1964), Advertising and Competition. *Journal of Political Economy*, 72(6) , 537–562.
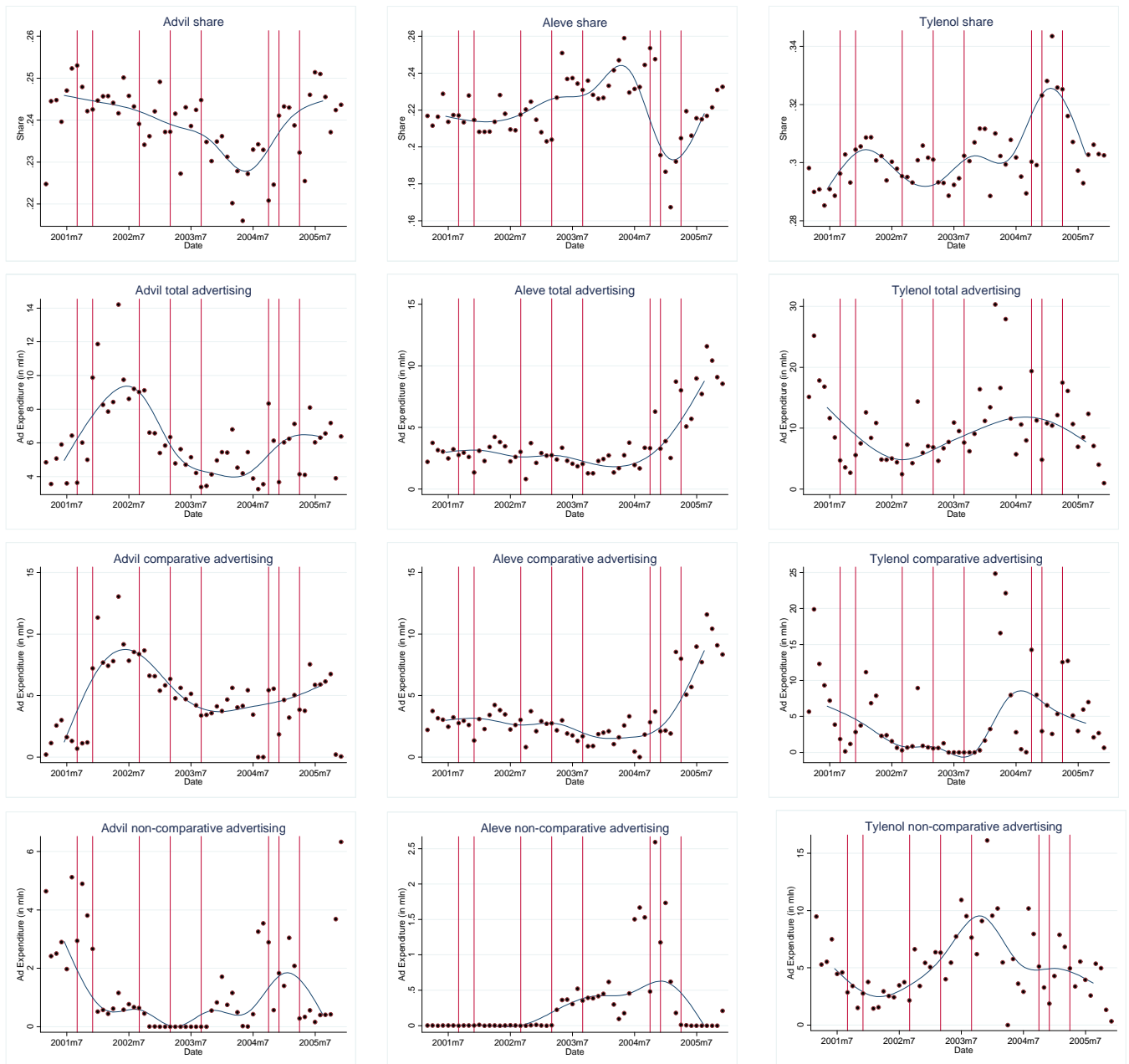
Year	Month	Description	Year	Month	Description
2001	9	Early Vioxx/Celebrex safety concerns	2003	9	NSAIDs inhibit cardioprotective benefits of Aspirin
2001	12	Ibuprofen counteracts Aspirin	2004	10	Vioxx withdrawal
2002	9	FDA panel calls for stronger warnings on NSAIDs	2004	12	Aleve is associated with increased cardio risk
2003	3	Aspirin prevents colorectal adenomas	2005	4	Bextra withdrawal

Figure 1. Timelines of Advertising Expenditures, Market Shares and Medical New Shocks

41

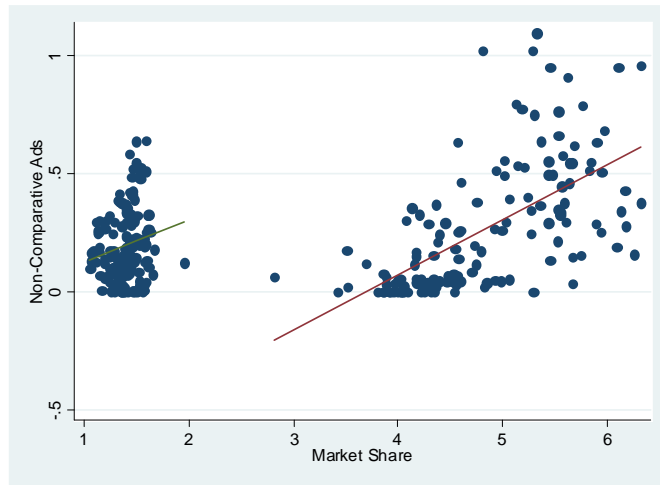Figure 2. Relationship between Non-Comparative Ads and Market Shares

Table 1: Brands, market share and advertising levels of OTC analgesics market

Brand	Active Ing.	Price per serving	Sales Share	Brand Vol. Share	Weighted Share	Max Pills	TA/ Sales	CA/ Sales	CA/ TA	Owner- ship
Tylenol	ACT	$2.15	29.16%	38.90%	30.51%	7.22	17.34%	4.98%	28.71%	McNeil
Advil	IB	$1.61	17.15%	22.87%	24.21%	5.90	20.00%	14.60%	72.99%	Wyeth
Aleve	NS	$.84	8.25%	11.00%	22.40%	3	26.56%	23.82%	89.71%	Bayer
Excedrin	ACT	$2.41	8.80%	11.74%	8.28%	9.22	26.42%	4.02%	15.22%	Novartis
Bayer	ASP	$1.85	5.73%	7.65%	6.98%	10.07	28.82%	8.80%	30.53%	Bayer
Motrin	IB	$1.73	5.83%	7.78%	7.68%	5.86	20.39%	8.07%	39.58%	McNeil
Generic	ACT	$1.17	8.00%							
Generic	IB	$.66	9.25%							
Generic	ASP	$.82	6.08%							
Generic	NS	$.57	1.66%							

Table 2: Comparative advertising and target pairs

Adv- ertiser ⇓	TARGET: Advil	Aleve	Bayer	Excedrin	Motrin	Tylenol	Total
Advil	-	17.80 [26]	-	4.26 [20]	-	160.20 [56]	182.26 [102]
Aleve	2.64 [9]	-	2.64 [9]	3.12 [16]	2.64 [9]	134.31 [58]	145.36 [101]
Bayer	13.17 [25]	2.05 [8]	-	-	2.05 [8]	15.69 [37]	32.95 [78]
Excedrin	-	1.96 [6]	2.15 [7]	-	-	19.96 [14]	24.08 [28]
Motrin	18.84 [25]	18.79 [25]	-	-	-	-	37.63 [50]
Tylenol	23.07 [43]	45.11 [51]	28.10 [40]	4.27 [21]	15.64 [39]	-	116.18 [194]
Total	57.72 [102]	85.71 [116]	32.89 [56]	11.66 [57]	20.33 [56]	330.15 [165]	538.47 [552]

1

Table 3: Medical News Shocks

No.	News Shock Description	Date	Source
	Major		
1	Risk of Cardiovascular Events Associated With Selective COX-2 Inhibitors	8/21/2001	Journal of the American Medical Association (JAMA)
2	Ibuprofen Interferes with Aspirin	12/20/2001	New England Journal of Medicine
3	FDA Panel Calls for Stronger Warnings on Aspirin and Related Painkillers	9/21/2002	FDA Public Health Advisory
4	Aspirin Could Reduce Breast Cancer Risk/ NSAIDs Protect Against Alzheimer's	4/8/2003/ 4/2/2003	JAMA American Academy Of Neurology
5	Anti-Inflammatory Pain Relievers Inhibit Cardioprotective Benefits of Aspirin	9/9/2003	Circulation
6	Vioxx Withdrawn From the Market	9/30/2004	
7	Long Term Use of Naproxen Associated with Increased Cardiovascular Risk	12/23/2004	FDA Public Health Advisory
8	Bextra Withdrawn	4/7/2005	
	Minor		
9	Ibuprofen May Prevent Alzheimer's	11/8/2001	Nature
10	Aspirin May Prevent Prostate Cancer	3/12/2002	Mayo Clinic Proceedings
11	Aspirin May Prevent Pancreatic Cancer	8/6/2002	J. of the National Cancer Institute
12	Aspirin Prevents Colorectal Adenomas	3/6/2003	New England Journal of Medicine
13	Misusing acetaminophen, can be deadly	1/23/2004	FDA Public Health Advisory
14	Myocardial infarction associated with Vioxx	4/19/2004	Circulation
15	Celebrex and Vioxx increases risk of acute myocardial infarction or cardiac death	8/25/2004	Annual meeting of the International Society for Pharmacoepidemiology
16	Acetaminophen, NSAIDs Increase Women's Hypertension Risk	8/15/2005	Hypertension

	Baseline	Brand Dummy	Major News Shocks	All News Shocks	Full IV	Partial IV	Partial IV Linear
$\frac{1}{Ms_{jt}}$	-0.0300	-0.212***	-0.146*	-0.120	-0.0810	-0.113	
	(0.0198)	(0.0596)	(0.0826)	(0.0731)	(0.0772)	(0.0730)	
Ms_{jt}							9.971***
							(2.655)
$\sum_{k\neq j} A_{jk}$	-0.700***	-0.657***	-0.575***	-0.452***	-0.466***	-0.468***	-0.455***
Outgoing Comp Ads	(0.0760)	(0.0758)	(0.0635)	(0.0607)	(0.0615)	(0.0616)	(0.0601)
$\sum_{k\neq j} A_{kj}$	0.590***	0.596***	0.395***	0.367***	0.401***	0.399***	0.332***
Incoming Comp Ads	(0.0620)	(0.0610)	(0.0650)	(0.0655)	(0.0671)	(0.0672)	(0.0686)
$\bar{\xi}_T$		-0.305***	-0.0973	-0.0556	-0.00492	-0.0479	-0.266**
Top Brand FE		(0.0943)	(0.121)	(0.110)	(0.115)	(0.109)	(0.109)
Constant	0.234***	0.637***	0.517***	0.455***	0.376**	0.440***	0.0399
	(0.0417)	(0.131)	(0.171)	(0.152)	(0.160)	(0.152)	(0.0573)
Ctr Fct $[\frac{1}{Ms_{jt}}]$					5.261		
					(4.270)		
Ctr Fct $[\sum_{k\neq j} A_{jk}]$					0.0192	0.0185	0.0166
					(0.0218)	(0.0218)	(0.0213)
Ctr Fct $[\sum_{k\neq j} A_{kj}]$					-0.0965**	-0.0952**	-0.0801*
					(0.0479)	(0.0480)	(0.0472)
Elasticity (Ms_{jt})	0.312	2.208	1.520	1.244	.8421	1.1712	1.5412
Observations	348	348	348	348	348	348	348
Log Likelihood	8.699	13.82	131.6	152.6	155.4	154.6	160.3
Major News Shocks	No	No	Yes	Yes	Yes	Yes	Yes
Minor News Shocks	No	No	No	Yes	Yes	Yes	Yes
F Test (1st Stage, $\frac{1}{Ms_{jt}}$)					←--	F(3,344)=52.12	--→
						Prob>F=0.000	
First Stage Full R^2, $\frac{1}{Ms_{jt}}$						$R^2=0.9857$	
First Stage Residual R^2, $\frac{1}{Ms_{jt}}$						$R^2=0.3181$	
F Test (1st Stage), $\sum_{k\neq j} A_{jk}$					←--	F(3,344)=52.12	--→
						Prob>F=0.000	
First Stage Full R^2, $\sum_{k\neq j} A_{jk}$						$R^2=0.5436$	
First Stage Residual R^2, $\sum_{k\neq j} A_{jk}$						$R^2=0.4483$	
F Test (1st Stage), $\sum_{k\neq j} A_{kj}$					←--	F(3,344)=52.12	--→
						Prob>F=0.000	
First Stage Full R^2, $\sum_{k\neq j} A_{kj}$						$R^2=0.7622$	
First Stage Residual R^2, $\sum_{k\neq j} A_{kj}$						$R^2=0.3239$	

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
1) The three first stage regressions are the same for the last three columns.
2) F Test (1st Stage) is a test of whether the coefficients of the ivs are all equal to zero in the first stage.
3) First Stage Full R^2 is the R^2 of the first stage regression, without including the ivs.
4) First Stage Residual R^2 is the R^2 of the regressions of the residuals of the first stage regression
     without ivs on the ivs. It says how much of the residual variation in the first stage is explained by the ivs.

Table 5: Comparative Advertising

	Baseline	Pair Brand Dummies	Major News Shocks	All News Shocks	Full IV	Partial IV	Partial IV Linear
$\frac{1-s_j}{Ms_j s_k}$	-2.457***	-0.867**	-1.571**	-1.678**	-1.617**	-1.564**	
	(0.206)	(0.407)	(0.665)	(0.694)	(0.752)	(0.749)	
$Ms_j s_k$							7.112***
							(0.770)
A_{kk}	-0.0432**	-0.0306	0.00210	-0.00833	-0.0703		
Targeted Self-Promotion	(0.0215)	(0.0194)	(0.0270)	(0.0291)	(0.0717)		
$\sum_{k\neq l} A_{kl}$	-0.0330	0.00926	0.00811	0.00708	0.0239		
Targeted Outgoing Comp Ads	(0.0262)	(0.0260)	(0.0319)	(0.0349)	(0.0427)		
$\sum_{l\neq k,j} A_{lk}$	0.307***	0.342***	0.343***	0.354***	0.443***	0.410***	0.355***
Targeted Incoming Comp Ads	(0.0220)	(0.0214)	(0.0319)	(0.0349)	(0.0698)	(0.0475)	(0.0460)
$\bar{\xi}_{TB,TB}$		0.109**	-0.0241	-0.0363	-0.0490	-0.0358	-0.436***
Top Brand-Top Brand FE		(0.0505)	(0.0763)	(0.0788)	(0.0844)	(0.0840)	(0.0723)
$\bar{\xi}_{TB,OB}$		0.0455	-0.0275	-0.0362	-0.0284	-0.0205	-0.0259
Top Brand-Other Brand FE		(0.0402)	(0.0600)	(0.0622)	(0.0668)	(0.0666)	(0.0306)
$\bar{\xi}_{OB,OB}$		-0.0417	-0.166***	-0.177***	-0.195***	-0.186***	-0.173***
Other Brand-Other Brand FE		(0.0400)	(0.0624)	(0.0649)	(0.0702)	(0.0701)	(0.0341)
Ctr Fcn $[s_{jt}]$					1.163	1.258	-6.685***
					(2.014)	(2.003)	(2.072)
Ctr Fcn $[s_{kt}]$					-0.621	-0.731	-8.414***
					(1.997)	(1.976)	(2.054)
Ctr Fcn $[A_{kk}]$					0.0608		
					(0.0782)		
Ctr Fcn $[\sum_{k\neq l} A_{kl}]$					-0.0344		
					(0.0345)		
Ctr Fcn $[\sum_{l\neq k,j} A_{lk}]$					-0.137*	-0.110*	-0.0555
					(0.0804)	(0.0648)	(0.0621)
Constant	-0.00608	-0.104*	0.0148	0.0286	0.0243	0.00595	-0.236***
	(0.0116)	(0.0550)	(0.0797)	(0.0828)	(0.0902)	(0.0885)	(0.0288)
Elasticity (Ms_{jt})	1.580	0.558	1.010	1.079	1.0398	1.0060	6.0073
Elasticity (Ms_{kt})	1.504	0.531	0.962	1.027	0.9898	0.9577	6.0073
Observations	1740	1740	1740	1740	1740	1740	1740
Log Likelihood	-114.0	-34.42	134.1	138.9	141.6	140.4	181.4
Major News Shocks	No	No	Yes	Yes	Yes	Yes	Yes
Minor News Shocks	No	No	No	Yes	Yes	Yes	Yes
F Test (1st Stage), A_{kk}					←--	F(27,238)=4.25	--→
						Prob>F=0.000	
First Stage Full R^2, A_{kk}						$R^2=0.652$	
First Stage Residual R^2, A_{kk}						$R^2=0.101$	
F Test (1st Stage), $\sum_{k\neq l} A_{kl}$					←--	F(30,247)=2.80	--→
						Prob>F=0.000	
First Stage Full R^2, $\sum_{k\neq l} A_{kl}$						$R^2=0.540$	
First Stage Residual R^2, $\sum_{k\neq l} A_{kl}$						$R^2=0.444$	
F Test (1st Stage), $\sum_{l\neq k,j} A_{lk}$					←--	F(29,230)=10.08	--→
						Prob>F=0.000	
First Stage Full R^2, $\sum_{l\neq k,j} A_{lk}$						$R^2=0.767$	
First Stage Residual R^2, $\sum_{l\neq k,j} A_{lk}$						$R^2=0.307$	

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
1) The three first stage regressions are the same for the last three columns.
2) F Test (1st Stage) is a test of whether the coefficients of the ivs are all equal to zero in the first stage.
3) First Stage Full R^2 is the R^2 of the first stage regression, without including the ivs.
4) First Stage Residual R^2 is the R^2 of the regressions of the residuals of the first stage regression
   without ivs on the ivs. It says how much of the residual variation in the first stage is explained by the ivs.
5) The F test and R^2 for the first stage for Ms_{jt} are given in Table 4