AN EMPIRICAL ANALYSIS OF COMPETITIVE NONLINEAR PRICING*

GAURAB ARYAL† AND MARIA F. GABRIELLI‡

ABSTRACT. In this paper, we empirically analyze a market with competitive nonlinear pricing and multidimensional private information. We apply the method to a novel data set on Yellow Pages advertisements in Central Pennsylvania, U.S. First, we study the identification of the joint density of consumer preferences for the two directories, the marginal costs of printing, and common utility parameters. Since the data do not contain any other information about consumer characteristics, we rely on supply side optimality conditions to identify the model parameters. Second, we estimate the joint density of consumer preferences using Joe copula, and the cost and utility parameters using a nonlinear least squares method. Using the estimates in a counterfactual exercise we find evidence of asymmetric information with substantial heterogeneity among consumers, and estimate the welfare loss due to asymmetric information to be approximately 3.8% of the total sales.

Keywords: Nonlinear Pricing, Competition, Multidimensional Screening, Identification, Advertisement, Copula.

JEL classification: C14, D22, D82, L11, L13.

1. INTRODUCTION

In this paper we consider the problem of estimating multidimensional consumer preferences and the welfare cost of asymmetric information in a duopoly market, for differentiated products, with second-degree price discrimination. We build on

* Previously this paper was circulated under the title "Competition and Nonlinear Pricing in Yellow Pages." We are thankful to Yao Huang for her help with an earlier version of the paper. We want to thank Herman Bierens, Vijay Krishna, Mark Roberts, Neil Wallace and participants at Collegio Carlo Alberto, Columbia, Ohio State, Penn State, Queen Mary University of London, UNSW and UQ. Special thanks are due to Isabelle Perrigne and Quang Vuong for their intellectual impetus and suggestions. We are solely responsible for any error.

† University of Virginia; e-mail:aryalg@virginia.edu.

‡ CONICET and Universidad Nacional de Cuyo. e-mail: florgabrielli@gmail.com.

the Ivaldi and Martimort [1994] model of competitive nonlinear pricing and multi-dimensional preferences and determine optimality restrictions on the demand and the supply sides.

We apply this model to a novel data set (that was manually collected) on the advertisements placed by local businesses in Central Pennsylvania, U.S., with two Yellow Pages directories during the year 2006. Under the assumption that the data is an equilibrium outcome of the model, we first determine data features, and any additional conditions, that are sufficient to identify the joint density of consumer preferences, the marginal cost of printing and common utility parameters. The estimates support consumer heterogeneity, and asymmetric information, which results in 3.8% welfare loss.

There are several features in the data that guide my model choice. First, we observe the ads choices for *every* business units, it is limited to *only one* market and for *only one* year. Second, we only know the name of the business unit (e.g., name of a dental clinic) and nothing else that can be relevant to explain their ad choices. Third, the two directories are different, have different qualities (e.g., paper quality, colors) , number of circulation and both publishers compete with nonlinear pricing (e.g., give quantity-discount). Fourth, the differences in price per unit (we define the unit of measurement later) is smaller for smaller quantities and larger for larger quantities. Fifth, some consumers buy ads from both publishers even though, a priori the two should be close, if not perfect, substitutes. Sixth, even after we control for business types (e.g., cafe) there is a substantial variation in choices among the consumers. Seventh, price functions that are quadratic in ad (color adjusted) sizes provide a good approximation of the observed prices.

The first three features suggest that the observed product lines are endogenous, therefore the instruments in Berry [1994]; Berry, Levinsohn, and Pakes [1995] that are based on competition, or Fan [2013] instrument for endogenous product characteristics that is based on geographic variation in competition, are infeasible in this setting. Moreover, the fact that the sellers compete in nonlinear prices means that the average prices are different from the marginal prices, and since consumers respond to marginal prices and not average prices using average prices to estimate the demand will lead to incorrect estimates. On top of that, if consumer (taste) heterogeneous and unobserved by the sellers (and the researcher) then using average prices, and ignoring the fact that product varieties are endogenous will further compound the error.

In view of this, we posit that there is indeed an asymmetric information between the sellers and the consumers about the latter's willingness to pay for their products (ads). The sellers only knows the distribution of consumers' willingness to pay, and their costs. They compete by offering a menu of quantity-price pairs that satisfy incentive compatibility constraints. In doing so, we aim to put forth a message that whenever we have detail information about consumer choices from one market and one period, and nothing else, mechanism design models can provide sufficiently flexible tools to analyze markets with price discrimination with competition and multidimensional unobserved consumer preferences.

This paper seeks to contribute to the literature on "empirical mechanism design" starting with Ivaldi and Martimort [1994]. Also see Armstrong [2006] and Stole [2007] for excellent overviews of the theory of price discrimination with competition. For an original and important empirical analysis of Yellow Pages markets as two-sided markets see Rysman [2004], and for the relationship between competition and price discrimination in this market see Busse and Rysman [2005].

The idea that asymmetric information can be an important source of price discrimination is not new; see, for example, Mussa and Rosen [1978]; Maskin and Riley [1984] and Wilson [1993]. From the empirical perspective, asymmetric information is important because it can also lead to a significant welfare loss. However, recent empirical literature [Crawford and Shum, 2006; Einav, Finkelstein, and Cullen, 2010; Einav, Finkelstein, Ryan, Schrimpf, and Cullen, 2012; Einav, Jenkis, and Levin, 2012] has found that asymmetric information and its effect on welfare vary a lot across markets. This suggests that the nature of asymmetric information is inherently an empirical question to which this paper aims to contribute to.

As mentioned above, a nontrivial proportion of consumers buy ads from both publishers. One way to rationalize this is if we allow consumers to have different, possibly dependent, taste parameters for each ad, and not restrict them to choose from at most one seller. Allowing consumers to have two-dimensional taste/preference parameters leads to the problem of multidimensional screening *with* competition which is known to be a difficult problem.[1] Since there are too many papers in this area to list here the interested reader is referred to Oren, Smith, and Wilson [1983]; Epstein and Peters [1999]; Armstrong and Vickers [2001] and Martimort and Stole [2002] for more. Also see Armstrong [1996] and Rochet and Choné [1998] for multi-dimensional screening with a single seller, and Aryal [2016] for the identification of the latter. Empirical papers in this area are few and far between with the exception of Ivaldi and Martimort [1994].[2]

In this paper, we focus on the market for print ads published in two Yellow Pages where a potential consumer is anyone who has a phone number registered for business purpose in Central Pennsylvania. This consumer can choose from many options from either Verizon or Ogden, these are the two companies that publish and circulate two Yellow Pages directories in this area. Each of them offers a wide variety of ad options that differ in sizes and colors, and the data contains information about the ads bought by all consumers in the year 2006. Since we do not observe anything else about consumers we take their willingness to pay of ads as given, and focus on estimating these willingness to pay from choice and price data.

Even though an ad differs in size and color/quality, we use an aggregation method that translates size and color into a single dimensional attribute that we call the *quality-adjusted quantity*. Heuristically, this measure uses prices to transform sizes

[1] Rochet and Stole [2002] say, ".. because in a common setting of screening contracts – that of nonlinear pricing – an empirically reasonable model of consumer preferences arguably requires at least two dimensions of type uncertainty and a more general modeling of the participation decision."

[2] Alternatively, [Leslie, 2004; McManus, 2006; Cohen, 2008] focus only on data where consumer buys from at most one seller, and/or use only the demand side for estimation. Their identification rely on exclusion restriction and exogenous variation in covariates, they are infeasible here.

of an ad in terms of the most colorful ad option in each directory, e.g., a half page black and white ad could be equivalent to a quarter page ad colorful ad. On the one hand this allows me to keep the model tractable by avoiding the need to explicitly model each product as a collection of multidimensional attributes, on the other hand such aggregation leads to bunching –a phenomenon where two "types" of consumers choose the same option– that makes the identification difficult. However, the model predicts that in equilibrium every allocation rule is unique and monotonic up to bunching. In other words, higher type of bunched consumers will buy higher quality-adjusted ads. This monotonic relation between bunched types and choices means the allocation rules are invertible and identify these types.

The idea of inverting first-order condition(s) was pioneered in the context of auctions by Guerre, Perrigne, and Vuong [2000], and since then has been applied to price discrimination problems by Luo, Perrigne, and Vuong [2015]; Aryal [2016] and Aryal, Perrigne, and Vuong [2016], among others. In this regard this paper is also related to Ekeland, Heckman, and Nesheim [2002, 2004]; Heckman, Matzkin, and Nesheim [2010] and Chernozhukov, Galichon, Hallin, and Henry [2015] who rely on equilibrium conditions on both demand and supply side for identification.

If every consumer had bought ads from both directories then we could nonparametrically identify the joint density of consumer types. But most of the consumers choose ads from only one publisher. So one can only identify the truncated (marginal) distribution of each type, where the truncation is endogenously determined as the type who is indifferent between buying an ad and not buying an ad. We also show other parameters (marginal costs and common utility parameters) are identified under a location normalization. To combine the two marginals, we use a parametric copula and use the dependence structure from those who buy from both.

The estimates suggest: (a) Joe copula provides the best fit for the joint density of consumer tastes;[3] (b) there is substantial unobserved consumer heterogeneity; (c)

[3] We use Cramér-von Mises and Vuong [1989] tests on seven most widely used copulas in the literature, and Joe copula was selected by both tests.

competition is stronger at the lower end than at the upper end of the market; (d) consumers treat the two directories as (weak) substitutes; and (e) the welfare cost of asymmetric information is $\approx 3.8\%$ of sales revenue.

The paper is organized as follows: Section 2 describes the data, the model is presented in Section 3, the identification in Section 4, and the estimation and empirical findings in Section 5. Section 6 concludes.

2. DATA

The data contains information about print advertisement in two Yellow Pages directories placed by local business units in central Pennsylvania (State College and Bellefonte), U.S. for the year 2006. The two directories are published by Verizon (henceforth, VZ), which is a utility provider, and Ogden (henceforth, OG).[4] These two directories differ in terms of quality and size, but they cover the same geographic area. Each publisher offers consumers a wide variety of ad options to choose from. A consumer is a business unit (say a doctor) who has a phone registered for his/her clinic. We observe the ad placed by the clinic including the size and the color and the price(s) for the ad(s).

Data Sources. The information on the ads bought, their sizes and colors, were manually recorded from the two directories. Since it is a norm in the market to put names and addresses of each business unit corresponding to a phone number in the directory for free. This list provides the name and ad of every consumer in the market. The price data is gathered from the Yellow Page Association –an umbrella organization of Yellow Page publishers in the U.S. As members of this association, VZ and OG have to provide the association with their prices for all offered options. In view of the limited data and the focus of this paper we ignore the possibility of bargaining between a potential consumer and a publisher to lower the out of pocket price.[5]

[4] See Busse and Rysman [2005] for competition and nonlinear pricing in Yellow Pages markets across U.S.

[5] Bargaining is not likely to be of first order importance perhaps because both VG and OG are big players and because other features are clearly more important. Bargaining plus other features is too complex.

Differentiated Directories. The standard unit of measurement in this industry is called pica, which is approximately 1/6 of an inch. A standard listing equals 12 sq. pica in VZ and 9 sq. pica in OG, while a full page in VZ equals $3,020$ sq. pica and $1,845$ sq. pica in OG. This means that the VZ directory is slightly bigger than that of OG and has three columns while OG has only two. Not only that, VZ directory is thicker with higher paper quality (glossy) than OG. In terms of number of circulation, VZ distributes more than 215,400 copies while Ogden distributes 73,000, but in terms of geographic coverage both cover similar places. This would suggest that the consumers see these two directories as being differentiated products.

All the ads can be classified into three general categories: listing, space listing, and display. Both publishers offer different options within each category. For example, VZ offers three font sizes to just list the names, address and phone number(s). OG offers listing with only two font sizes. Listing with smallest font (called the standard listing) was offered for free by both to every one whose phone was registered as a business phone. The listing option accounts for 30% and 53% of the total ad sales in VZ and OG, respectively. Space listing refers to an option where a space is allocated within the column under an appropriate business heading (such as Doctors, Salons, etc.), Both VZ and OG offer five different options within this category and it accounts for 30% and 26% of the total ad sales in VZ and OG, respectively.

Display ad refers to a listing option with a space (that could cover up to two pages) where consumer can choose colorful pictures. VZ offers nine different variations and OG offers seven different variations within this category, which is also the most expensive of the three options. On top of that, VZ offers five color options – no color, one color, white background, white background plus one color and multiple colors including photos and OG offers the same options except for the 'white background' option.

These options are presented in Table A-1, from which we see that: (i) for any size, color accounts for most of the differences in prices, e.g., a full-page display ad with no color costs $\$18,510$ in VZ and $\$6,324$ in OG, and for multiple colors the prices

Size (No Color)	$ per Pica (Verizon)	$ per Pica (Ogden)
2.5% of page	10.84	10.65
10% of page	8.65	5.54
25% of page	7.98	3.93
Half Page	6.79	3.71
Full Page	6.12	3.42

TABLE 1. Quantity Discounts: Price per Pica for different sizes.

increase to $32,395 and $9,435, respectively; (ii) VZ's prices are significantly higher than Ogden's across all the comparable advertising options, e.g., a half-page display without color costs $10,093 in VZ and only $3,372 in OG; (iii) the price differences between VZ and OG are smaller for the lower-end options, such as listing, than for the upper-end options such as display, e.g., VZ's average price is 130% higher than OG's for the display option while the difference in prices is only 18% for space listing and 17% for standard listing with no color; and (iv) for a given color category, both VZ and OG offer a quantity discount: the price per sq. pica decreases with the ad size as is evident from Table 1. [6]

From Table A-2 we can see: (i) the display option, which is the most expensive option, accounts for more than 70% of the revenue for both VZ and OG; (ii) roughly 66% of consumers choose listing and 14% choose display in VZ, while 94% and 3.8% choose listing and display in OG; (iii) around 54% advertise exclusively with VZ, whereas only 2% advertise exclusively with OG, and 12% advertise with both, and the remaining choose the outside option (the standard listing) by default.

The average prices paid in each directory by the firms purchasing from both directories are higher than those who purchase from only one directory, which may indicate a higher valuation of advertising among this group. A similar pattern is observed with respect to ad sizes.

Two-Dimensional Preferences. Figure 1 shows a scatter plot of ads bought by consumers. Suppose each consumer could be indexed by a one dimensional parameter that govern the willingness to pay for the ad. Then, within the classic quasi-linear utility environment, this would mean that both publishers would value a consumer

[6] All Tables A- are in Appendix A.

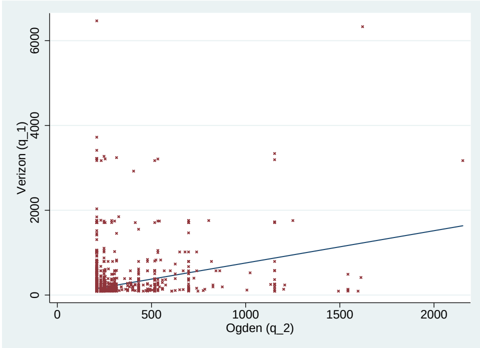

FIGURE 1. X-axis: Ogden ad; Y-axis: Verizon ad., in square pica

type the same way. If both publishers agreed on who are the lucrative consumers and who are not, then the observed ads bought would coalesce around an increasing straight line. This would mean that the ad sizes bought from VZ and OG would be highly correlated, but that is not supported by the scatter plot since the correlation between ad choices is 0.25, increasing up to only 0.32 for those who buy from both.

The correlation between the two ads is informative about the correlation between unobserved types. The Cramér-von Mises statistic for independence between the two sales was 1.66 ($p \approx 0$), rejecting the null of independence. The (normalized) rank plot (or the probability of a sale) of VZ and OG ad is presented in Figure 2.

So to explain this weak but significant correlation between choices we model consumers with two dimensional preferences. Although this leads to a multidimensional screening problem with competition, which is hard, the alternatives are even harder. For example, one way to rationalize the data that does not rely on multidimensional screening is to model the private "value" for an ad displayed to the "right" consumer, and the two publishers could have different stochastic matching technologies, so in equilibrium they could specialize in serving different types of consumers (e.g., low versus high income, etc.). An optimizing consumer would then place ads in both, by equating the marginal benefit per dollar spent. But this method is not applicable in the current situation because both VZ and OG distribute their directories to similar localities and have a significant overlap. (Without any

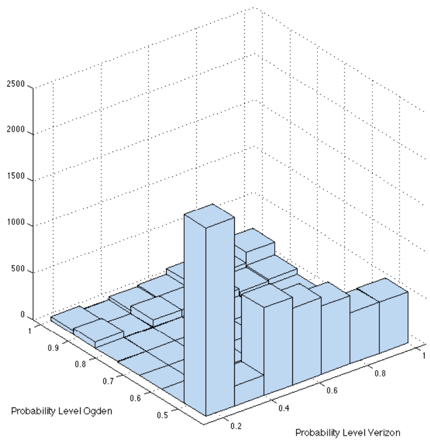

FIGURE 2. Rank plot of ads bought from Verizon and Ogden.

consumer covariates, we cannot estimate a model asymmetric stochastic matching technologies.)

Quality Adjusted Quantity. It is clear from the description above that ads vary in sizes and in colors. The question is if one should model the products to have multiple-attributes, which makes the model even more complicated. Instead of going this route we propose to use an aggregator that combines the two attributes into one aggregated quantity, which we refer to as quality-adjusted quantity, which allows me to keep the model tractable.

For this aggregation to work it must not be the case that the publishers use color as a tool for price-discrimination and the aggregation must not alter the ordering of ads choices from consumers' perspective. In fact we find that indeed that color is not used to discriminate across buyers. If both size and color were important then one should see the sellers using both dimensions to discriminate consumers, but once we control for the size, the relative price is constant across colors. In particular, discounts are offered for large ads while no such discounts are observed for ads with multiple colors and the ratio of the (marginal) prices for two different colors are constant across different sizes.

Moreover, from Maskin and Riley [1984] it is known that an optimal bundle of quantity and quality should lie on a unique curve in the quantity-quality space and the optimal quantity allocation should increase with quality along this curve, something that is not observed in the data. Both of these observations suggest that it might be possible to combine the discontinuous attributes (size and color) into one continuous aggregated attribute: quality-adjusted quantity.

To that end, we consider the price schedule for multicolored options and fit a continuous function that represents the price schedule. Then we can project each size into this multicolored price to get a new quantity (in picas) in terms of the multicolored size. Then we fit the following "quadratic" functions using OLS:

$$\begin{aligned} \widehat{T_1(q_{j1})} &= \gamma_1 + \alpha_1 q_{j1} - \frac{\beta_1}{2} q_{j1}^2, \\ \widehat{T_2(q_{j2})} &= \gamma_2 + \alpha_2 q_{j2} - \frac{\beta_2}{2} q_{j2}^2, \end{aligned} \tag{1}$$

where T_{ji} is the price in dollars for publisher i, and q_{ji} is the ad size for multicolored choices measured in square pica purchased by consumer j. The estimates are $\{\hat{\gamma}_1, \hat{\alpha}_1, \hat{\beta}_1/2\} = \{1512, 11.27, 0.00027\}$ and $\{\hat{\gamma}_2, \hat{\alpha}_2, \hat{\beta}_2/2\} = \{103, 6.25, 0.00066\}$, with $R^2 = 0.99$ and all estimates are significant at 1%.

Then the quality-adjusted quantities are constructed by plugging other (choices that are non-multicolored) onto these regression functions. For example, in VZ, a one-page ad with no color measures 3,020 sq. pica; the same size ad in multicolored measures 1,470 sq. pica. See Table (2) for the summary. This transformation automatically captures the data feature that Ogden offers only a relatively small menu of listing choices, especially within each color category, and hence it is competing only in a subset of Verizon's nonlinear tariff.

Note. One empirical regularity that has been observed, since Wilson [1993], and more recently by Chu, Leslie, and Sorensen [2011], is that the price schedules tend to be "simple," e.g., two-part tariff or quadratic, and yet approximately optimal.

	Min	1st Quartile	Median	Mean	Max
Verizon	92.27	92.27	114.50	171.82	6485.60
Ogden	209.5	209.5	209.5	230.3	2147.4

TABLE 2. Summary of Quality Adjusted Quantity

Given (1), in the model we can restrict the optimal prices to be a quadratic function, which will considerably simplify the model.

3. THE MODEL

In this section we present a model of competing nonlinear pricing that builds on Ivaldi and Martimort [1994]. Let $P1$ and $P2$ be two principals or sellers that stand for VZ and OG, respectively. In the market we study, it is not clear whether P1 and P2 move simultaneously, or sequentially (P2 being the follower). From Borenstein and Rose [1994]; Busse and Rysman [2005] we know that competition affects the equilibrium only through the price schedule and not through quantities (also known as allocation rules in mechanism design parlance). So when we estimate the model, we only use the equilibrium allocation rules for P1 and P2, but only use the price function for P2 (P2's reaction function), which is the same whether or not P2 moves at the same time as P1 or after P1.

Let $u(\mathbf{q}, \theta, A)$ be the gross utility that a consumer of type $\theta := (\theta_1, \theta_2)$ gets from choosing $\mathbf{q} := (q_1, q_2)$, and let A be the set of utility parameters that are common for all consumers. If a $(\theta_1, \theta_2)-$type consumer chooses (q_1, q_2), let the net utility be

$$\begin{aligned} U(q_1, q_2, \theta_1, \theta_2) &:= u(\mathbf{q}, \theta, A) - \sum_{i=1}^{2} T_i(q_i) \\ &:= \sum_{i=1}^{2}\left(\theta_i q_i - \frac{b_i q_i^2}{2}\right) + c q_1 q_2 - \sum_{i=1}^{2} T_i(q_i), \end{aligned} \tag{2}$$

where $T_i(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ is the pricing function chosen by Pi. So $A := \{b_1, b_2, c\}$ and let $b_i > 0, i = 1, 2$ and let $b_1 b_2 - c^2 > 0$ (for concavity). For each consumer, the (net) marginal utility from q_i is $MU_i = \theta_i - b q_i + c q_{-i} - T'(q_i)$, which increases with the type θ_i. In view of the data (Figures 1 and 2) We assume that the two ads are weak substitutes, i.e., $c \leq 0$, so MU_i decreases with q_{-i}.[7]

Equation (2) says that consumers have the same quasi-linear utility function but they differ in terms of their marginal utility for the two products. The quasi-linearity assumption is a simplifying assumption used in almost all mechanism design literature and discrete choice literature. Moreover, with multiple products, quadratic utility comes handy and has been used widely in the literature. Similar functional forms have also been used in hedonic models by Ekeland, Heckman, and Nesheim [2002]; and Ekeland, Heckman, and Nesheim [2004], among others.

Only the consumers know their own type (θ_1, θ_2). The publishers commonly know that they are distributed as $F(\cdot, \cdot)$. They also know each other's printing cost which is assumed to have a fixed cost (e.g., printing machine and distribution) and constant marginal cost of printing (e.g., ink, paper and labor). We begin with the following assumption

Assumption 1. (1) *The utility function is concave, i.e.* $b_1 b_2 - c^2 > 0$ *and* $c \leq 0$.
(2) $(\theta_1, \theta_2) \overset{i.i.d}{\sim} F(\cdot, \cdot)$, *with density* $f(\cdot, \cdot) > 0$ *on the support* $[\underline{\theta}_1, \bar{\theta}_1] \times [\underline{\theta}_2, \bar{\theta}_2]$.
(3) *Cost function:* $C_i(q_i) = K_i + m_i q_i$ *with* $K_i \geq 0$ *and* $m_i > 0$ *for* $i = 1, 2$.

Consumers' preference heterogeneity is captured by heterogeneity in (θ_1, θ_2), which affects the value for ads (in terms of dollars). In this paper we take these values as given and ignore any "deeper" game that could generate the value for ads. Implicitly this means that a consumer's willingness to pay for an ad is independent of the composition of the ads placed by others in the market. This modeling assumption is also the norm in the media economics literature; see, for example, Anderson and Waldfogel [2016] and Berry and Waldfogel [2016].

3.1. **Optimal Nonlinear Pricing.** $P1$ chooses $\{q_1(\cdot), T_1(\cdot)\}$ and $P2$ chooses $\{q_2(\cdot), T_2(\cdot)\}$. Then each consumer chooses (q_1, q_2) and pays accordingly. Given Equation (1) we

[7] If $c > 0$ then q_1 and q_2 should be positively-assortative given the concavity of the gross utility function and the two firms would not be competing.

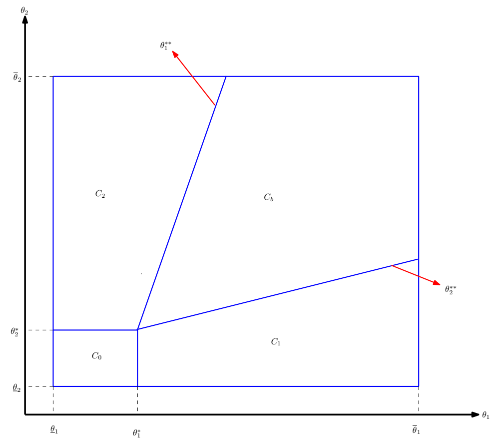

FIGURE 3. Consumer types: C_0 types choose (q_{10}, q_{20}); C_i types choose $(q_i > q_{i0})$ but q_{-io} and C_b types choose $(q_1 > q_{10}, q_2 > q_{20})$.

focus on quadratic pricing function, i.e.,

$$T_1(q_1) = \begin{cases} \gamma_1 + \alpha_1 q_1 + \frac{\beta_1}{2} q_1^2 & \text{if} \quad q_1 > q_{10} \\ 0 & \text{if} \quad q_1 \leq q_{10}, \end{cases} \tag{3}$$

such that $T_1(\cdot)$ is right differentiable at q_{10}. So P1 chooses $\gamma_1 > 0, \alpha_1 > 0$ and $\beta_1 < 0$. Although this assumption is consistent with the observed price schedule, it is nonetheless a strong assumption. Given the nature of the problem, focusing on quadratic pricing will considerably simplify the equilibrium conditions. Moreover, Wilson [1993] argues that simpler nonlinear pricing tends to be nearly optimal, so this assumption need not be as restrictive as it appears.

To determine the participation/individual rationality (IR) and incentive compatibility constraint (IC), consider the consumers' first order conditions

$$\begin{aligned} (\theta_1 - b_1 q_1 + c q_2 - T_1'(q_1))(q_1 - q_{10}) &= 0; \\ (\theta_2 - b_2 q_2 + c q_1 - T_2'(q_1))(q_2 - q_{20}) &= 0. \end{aligned} \tag{4}$$

These conditions determine four subsets of consumers: C_0 which denotes consumers who choose the outside option (q_{10}, q_{20}); C_1 and C_2 which denote consumers who only choose $P1$ and P_2, respectively; and C_b which denotes consumers who choose from both sellers. See Figure 3. These subsets depend on $T_1(\cdot)$ and $T_2(\cdot)$.

Consider the set C_0. For all $(\theta_1, \theta_2) \in C_0$ the marginal utility $MU_i(q_{10}, q_{20}; \theta_1, \theta_2) \leq 0$ for both $i = 1$ and $i = 2$. From (3), these two conditions can be simplified to

$$\begin{aligned} \theta_1 - b_1 q_{10} + c q_{20} &\leq \alpha_1 + \beta_1 q_{10} \\ \theta_2 - b_2 q_{20} + c q_{10} &\leq T_2'(q_{20}). \end{aligned}$$

Let (θ_1^*, θ_2^*) be the marginal type who choose (q_{10}, q_{20}), i.e.,

$$\begin{aligned} \theta_1^* &= \alpha_1 + (b_1 + \beta_1) q_{10} - c q_{20}; \\ \theta_2^* &= T_2'(q_{20}) + b_2 q_{20} - c q_{10}. \end{aligned}$$

So all consumers with type $(\theta_1, \theta_2) \ll (\theta_1^*, \theta_2^*)$ find it optimal to choose (q_{10}, q_{20}). Now, consider C_1 where consumers choose $q_1 > q_{10}$ but $q_2 = q_{20}$. The types must satisfy the following conditions:

$$\begin{aligned} \theta_1 - b_1 q_1 + c q_{20} &= \alpha_1 + \beta_1 q_1; \\ \theta_2 - b_2 q_{20} + c q_1 &\leq T_2'(q_{20}). \end{aligned}$$

From the first equality we get $q_1 = \frac{\theta_1 - \alpha_1 + c q_{20}}{b_1 + \beta_1}$, which together with the second inequality determines the threshold type θ_2^{**} such that all $\theta_2 \leq \theta_2^{**}$ consumers choose q_{20}. Since the marginal utility from q_2 depends on the choice of q_1, this threshold type is a function of θ_1 and is:

$$\theta_2^{**} = \left(b_2 - \frac{c^2}{b_1 + \beta_1} \right) q_{20} + T_2'(q_{20}) + \frac{c \alpha_1}{b_1 + \beta_1} - \frac{c}{b_1 + \beta_1} \theta_1. \tag{5}$$

Similarly, C_2 is the counterpart of C_1 and is determined in the same way. Let, θ_1^{**} be the threshold type such that any type with $\theta_1 \leq \theta_1^{**}$ buys q_{10} and is:

$$\theta_1^{**} = \left(b_1 + \beta_1 - \frac{c^2}{b_2} \right) q_{10} + \alpha_1 + \frac{c}{b_2} T_2'(q_2) - \frac{c}{b_2} \theta_2.$$

C_b corresponds to the "interior solution" where consumers choose positive amount of both. This set is determined by the two first-order conditions as in Equation (4) that can be simplified as

$$\theta_1 - b_1 q_1 + c q_2 = \alpha_1 + \beta_1 q_1; \tag{6}$$

$$\theta_2 - b_2 q_2 + c q_1 = T_2'(q_2). \tag{7}$$

Next, we will use these sets to determine the optimal nonlinear pricing for $P2$. Note that, even though consumers have two dimensional types (θ_1, θ_2), each seller has only one dimensional instrument q. This means that there will be bunching - where more than one type of consumers are allocated the same q. Then P2's objective is to determine the most profitable way to bunch, while preserving the incentive compatibility across types that are not bunched. This is known to be a hard problem to solve, and as far as we know, if we want to characterize equilibrium conditions we have to impose some functional form assumption on both the utility and the pricing function. Fortunately, it turns out that the assumptions on utility and $T_1(\cdot)$ (3) are sufficient for that purpose.

Next, using the consumer optimality conditions, we show that the two-dimensional screening problem can be transformed into a one-dimensional screening problem using a sufficient-statistic. Then P2's problem becomes a canonical one-dimensional screening problem, with respect to this sufficient statistic, which is considerably easier to solve.

For those consumers who buy $q_2 > q_{20}$, the corresponding q_1 can be determined from (6) as

$$q_1 = \begin{cases} \frac{\theta_1 - \alpha_1 + c q_2}{b_1 + \beta_1}, & \theta_1 > \theta^* \\ q_{10}, & \theta_1 \leq \theta_1^*. \end{cases} \tag{8}$$

Substituting (8) in (7) gives the necessary condition for q_2 to be optimal for type (θ_1, θ_2) consumer, i.e.,

$$\theta_2 + \frac{c\theta_1}{b_1 + \beta_1} = \frac{c\alpha_1}{b_1 + \beta_1} + \left(b_2 - \frac{c^2}{b_1 + \beta_1}\right) q_2 + T_2'(q_2). \tag{9}$$

Notice that consumer's type (θ_1, θ_2) appear only in the LHS of (9). So, given $T_1(\cdot)$, P2 can take the linear combination of (θ_1, θ_2) as exogenous. Letting $z_2 := \theta_2 + \frac{c\theta_1}{b_1+\beta_1}$ we can re-write (9) in terms of a one-dimensional statistic:

$$z_2 = \frac{c\alpha_1}{b_1+\beta_1} + \left(b_2 - \frac{c^2}{b_1+\beta_1}\right) q_2 + T_2'(q_2). \tag{10}$$

As can be seen from the definition of z_2, it aggregates the effect of θ_1, θ_2 and the price of P1, and hence it is a "sufficient statistic" for θ and prices. It increases with θ_2 as desired, and weakly decreases with θ_1 (those who value q_1 more, and q_2 less) and decreases with β_1 (q_2 is a weakly increasing price for VZ, ceteris paribus). Therefore, z_2 aggregates (θ_1, θ_2) is a sufficient statistic in the sense that a mechanism that depends on z_2 will do as good as a mechanism that depends on (θ_1, θ_2). An implication of this aggregation is that all consumers with same z will buy the same q_2 even though they might have different (θ_1, θ_2).

Let $G_2(\cdot)$ be the distribution of $z_2 \in [\underline{z}_2, \overline{z}_2]$ and $g_2(\cdot)$ its density, then

$$z_2 \sim g_2(z_2) := \int_{\bar{\theta}_2}^{\theta_2} f\left(\theta_1, z_2 - \frac{c\theta_1}{b_2+\beta_2}\right) \mathrm{d}\theta_1.$$

Now, $P2's$ optimization problem can be written in terms of z_2 as

$$\max_{T_2(\cdot), q_2(\cdot), z_2^0} \left\{ \mathbb{E}\Pi_2 = \int_{z_2^0}^{\overline{z}_2} \Big(T_2\left(q_2(z_2)\right) - m_2 q_2(z_2) \Big) g_2(z_2) \mathrm{d}z_2 - K_2 - m_2 q_{20} G_2(z_2^0) \right\}, \tag{11}$$

subject to the appropriate IC and IR constraints (see below). The threshold type z_2^0 that corresponds to the types who choose the outside option q_{20}, i.e., the area C_1 and C_0 types in Figure 3 are such that $z_2^0 = \theta_2^* + \frac{c\theta_1^*}{b_1+\beta_1}$ if $\theta_1 \leq \theta_1^*$ and $z_2^0 = \theta_2^{**} + \frac{c\theta_1}{b_1+\beta_1}$ if $\theta_1 \geq \theta_1^*$. Since the utility from q_2 depends on q_1, which depends on q_2, (see (6) and (7)), determining the incentive compatibility constraints for P2 needs some additional work. See Epstein and Peters [1999] for more on competing principal-agent models.

Let, $W_2(\theta_1, z_2)$ be z_2's indirect utility from (q_1, q_2), i.e.,

$$W_2(\theta_1, z_2) \quad := \quad \max_{q_1 \geq q_{10}, q_2 \geq q_{20}} \left[u\left(q_1, q_2; \theta_1, z_2 - \frac{c\theta_1}{b1+\beta_1}\right) - T_1(q_1) - T_2(q_2) \right],$$

and let $w_2(\theta_1, z_2)$ be the net utility that z_2 gets from (q_1, q_{20}), i.e.,

$$w_2(\theta_1, z_2) := \max_{q_1 \geq q_{10}} \left[u\left(q_1, q_{20}; \theta_1, z_2 - \frac{c\theta_1}{b_1 + \beta_1}\right) - T_1(q_1) \right].$$

And let $s_2(q_2, z_2)$ be such that

$$s_2(q_2, z_2) : \max_{q_2 \geq q_{20}} \left\{ \left(z_2 - \frac{c\alpha_1 - c^2 q_2}{b_1 + \beta_1}\right)(q_2 - q_{20}) - \frac{b_2}{2}(q_2^2 - q_{20}^2) - T_2(q_2) \right\}.$$

Then after some simplification one can see that

$$W_2(\theta_1, z_2) = w_2(\theta_1, z_2) + s_2(q_2, z_2),$$

which means one can decompose the indirect utility from (q_1, q_2) into a sum of the indirect utility from (q_1, q_{20}) and the additional utility from choosing $q_2 > q_{20}$. Since P2 can only affect $s_2(\cdot)$ it is the only relevant "utility function,"that P2 cares about. Thus the IC constraints can be expressed as follows:

$$s_2(q_2(z_2); z_2) \geq s_2(q_2(\tilde{z}_2); z_2), \quad \forall z_2, \tilde{z}_2 \in [\underline{z}_2, \bar{z}_2].$$

Moreover, $s_2(\cdot)$ is continuous, convex and satisfies the envelope conditions

$$s_2'(z_2) = q_2(z_2) - q_{20}, \qquad \forall z_2 \in (z_2^0, \bar{z}_2], \tag{12}$$

and

$$T(z_2) = \left(z_2 - \frac{c\alpha_1 - c^2 q_2}{b_1 + \beta_1}\right)(q_2 - q_{20}) - \frac{b_2}{2}(q_2^2 - q_{20}^2) - s_2(z_2). \tag{13}$$

From (12) and (13), we see that without loss of generality P2 can be viewed as choosing $s_2(z_2)$ as the rent function and charging $T_2(\cdot)$. Rochet [1987] showed that the global IC constraint is satisfied if and only if: (i) $s_2(z_2) = \int_{z_2^0}^{z_2} (q_2(t) - q_{20}) \mathrm{d}t + s^+, \forall z_2 \in [z_2^0, \bar{z}_2]$, where $s_2^+ \equiv \lim_{z_2 \downarrow z_2^0} s_2(z_2)$; and (ii) $s_2(\cdot)$ is increasing, or equivalently from (12), the allocation function is strictly increasing in z_2, i.e., $q_2'(z_2) > 0$.

Similarly, the participation or IR constraint becomes

$$W_2(\theta_1, z_2) = w_2(\theta_1, z_2) + s_2(z_2) \geq \max\{w_2(\theta_1, z_2), 0\},$$

or equivalently $s_2(z_2) \geq 0$.

Then, P2 solves:

$$\max_{q_2(\cdot), z_2^0, s_2^+} \Bigg\{ \mathbb{E}\Pi_2 = \int_{z_2^0}^{\bar{z}_2} \Bigg[\left(z_2 - \frac{c\alpha_1 - c^2 q_2(z_2)}{\beta_1 + b_1} \right) (q_2(z_2) - q_{20}) - \frac{b_2}{2}(q_2^2(z_2) - q_{20}^2)$$
$$-m_2 q_2(z_2) - s_2^+ - (q_2(z_2) - q_{20}) \frac{1 - G_2(z_2)}{g_2(z_2)} \Bigg] g_2(z_2) \mathrm{d}z_2 - K_2 - m_2 q_{20} G_2(z_2^0) \Bigg\},$$

subject to the following IC and IR constraints, respectively

$$q_2'(z_2) > 0; \qquad s_2(z_2) \geq 0, \qquad \forall z_2 \in [\underline{z}_2, \bar{z}_2].$$

To solve the problem, we follow the literature and ignore the second order IC constraint and verify ex-post that the solution satisfies the constraint. Since $s_2(\cdot)$ is increasing, $s_2(z_2^0) = 0$ implies $s_2(z_2) > 0$ (IR) for all $z_2 \in (z_2^0, \bar{z}_2]$. It is immediate to see that $s_2^+ = 0$ is optimal. Since competition is endogenously accounted for in z_2 this is a one-dimensional screening problem, for which a unique equilibrium exists.[8] The next result provides the characterization of the allocation and Ramsey pricing rule.

Proposition 1. *Let* $(1 - G_2(\cdot))/g_2(\cdot)$ *be decreasing, and* $b_2 > \frac{2c^2}{b_1 + \beta_1}$. *Then,*

(1) *The optimal allocation function is*

$$q_2(z_2) = \frac{z_2 - \frac{1 - G_2(z_2)}{g_2(z_2)} - m_2 - \frac{c^2 q_{20} + c\alpha_1}{b_1 + \beta_1}}{b_2 - \frac{2c^2}{b_1 + \beta_1}}, \forall z_2 \in (z_2^0, \bar{z}_2] \tag{14}$$

and $q_2(z_2) = q_{20}$, *if* $z_2 \in [\underline{z}_2, z_2^0]$ *where* z_2^0 *solves*

$$z_2^0 - \frac{1 - G_2(z_2^0)}{g_2(z_2^0)} = (b_2 - \frac{c^2}{b_1 + \beta_1}) q_{20} + m_2 + \frac{c\alpha_1}{b_1 + \beta_1}.$$

[8] This follows from the existence and uniqueness results in Rochet and Choné [1998]. The proof is available upon request.

(2) *$T_2(q)$ must satisfy (13), it must also satisfy the Ramsey rule:*

$$\frac{T_2'(q_2(z_2)) - m_2}{T_2'(q_2(z_2))} = \frac{1 - G_2(z_2)}{g_2(z_2)} \frac{1}{\frac{\partial s_2(q_2(z_2))}{\partial q_2}}. \tag{15}$$

Proof. Since the proof is standard in the literature [Stole, 2007] we will highlight only the main steps. The first step is to show that $\mathbb{E}\Pi$ is concave in q_2, and super modular in (q_2, z_2).

Let I be the integrand of the expected profit function. Then,

$$\begin{aligned}
\frac{\partial I}{\partial q_2} &= \left(\left(z_2 - \frac{c\alpha_1 - c^2 q_2}{\beta_1 + b_1}\right) + \frac{c^2}{b_1 + \beta_1}(q_2 - q_{20}) - b_2 q_2 - \frac{1 - G(z_2)}{g(z_2)} - m_2\right) g(z_2),\\
\frac{\partial^2 I}{\partial q_2^2} &= -(b_2 - \frac{2c^2}{b_1 + \beta_1}) g_2(z_2),\\
\frac{\partial^2 I}{\partial q_2 \partial z_2} &= \left(\left(1 - \frac{\partial}{\partial z_2}\frac{1 - G(z_2)}{g(z_2)}\right) - \left(b_2 - \frac{2c^2}{b_1 + \beta_1}\right) q_2'(\cdot)\right) g(z_2) = 0.
\end{aligned}$$

Since $g_2(\cdot) > 0$ and $b_2 > \frac{2c^2}{b_1 + \beta_1}$, concavity follows from the second equation. The last equation implies super modularity, i.e. $\frac{\partial^2 I}{\partial q_2 \partial z_2} \geq 0$. Optimal allocation q_2 can be determined by simple point-wise maximization of I:

$$\frac{c^2}{b_1 + \beta_1}(q_2 - q_{20}) + (z_2 - \frac{c\alpha_1 - c^2 q_2}{b_1 + \beta_1}) - b_2 q_2 - \frac{1 - G_2(z_2)}{g_2(z_2)} - m_2 = 0,$$

which determines the optimal allocation rule to be:

$$q_2(z_2) = \frac{z_2 - m_2 - \frac{c^2 q_{20} + c\alpha_1}{b_1 + \beta_1} - \frac{1 - G_2(z_2)}{g_2(z_2)}}{b_2 - \frac{2c^2}{b_1 + \beta_1}}.$$

The optimal z_2^0 is determined by the Euler method of differentiating the expected profit with respect to z_2^0:

$$-\left(z_2^0 - \frac{1 - G(z_2^0)}{g(z_2^0)} - m_2 - \frac{c\alpha_1 - c^2 q_2(z_2^0)}{b_1 + \beta_1}\right)(q(z_2^0) - q_{20}) + \frac{b_2}{2}(q_2^2(z_2^0) - q_{20}^2) = 0.$$

And since $q_2(z_2^0) = q_{20}$, z_2^0 solves $z_2^0 - \frac{1-G_2(z_2^0)}{g_2(z_2^0)} = (b_2 - \frac{c^2}{b_1+\beta_1})q_{20} + m_2 + \frac{c\alpha_1}{b_1+\beta_1}$. Differentiating (13) gives

$$T_2'(q_2) = z_2 - \frac{c\alpha_1}{b_1+\beta_1} - (b_2 - \frac{2c^2}{b_1+\beta_1})q_2 + \frac{c^2}{b+1+\beta_1}q_{20}, \tag{16}$$

from which the Ramsey equation (15) follows immediately. □

Equation (15) connects the quantity discount to the distribution of the demand (i.e. z_2). For instance, the markup is smaller if either the distribution of z_2 is skewed towards the lower end, i.e. $(1 - G(z_2))$ is smaller, or if $\frac{\partial s_2(q_2(z_2))}{\partial q_2}$ is higher.

Next, we solve for the optimal quantity function $q_1(\cdot)$, for which we follow the same logic as before with P2. That is, first using the consumer optimality conditions we determine a sufficient statistic $z_1 \sim g_1(\cdot)$. Second, we solve for the optimal allocation rule $q_1(\cdot)$ by solving the maximization problem for P1. To avoid repeating the same steps, we only present the result.

Proposition 2. *The optimal quantity allocation rule (contract) is given by*

$$q_1(z_1) = \frac{z_1 - \frac{1-G_1(z_1)}{g_1(z_1)} - m_1 - \frac{c\alpha_2+c^2q_{10}}{b_2+\beta_2}}{b_1 - \frac{2c^2}{b_2+\beta_2}}, \quad \forall z_1 \in (z_1^0, \overline{z}_1], \tag{17}$$

and $q_1(z_1) = q_{10}$, *if* $z_1 \in [\underline{z}_1, z_1^0]$.

4. Identification

In this section we study the identification of the model parameters. The parameters include the joint distribution of types $F(\cdot,\cdot)$, the set of utility parameters $[b_1, b_2, c]$ and the set of cost parameters $[m_1, m_2, K_1, K_2]$. In the data we observe the price functions which is completely characterized by two sets of triplets $\{\alpha_i, \beta_i, \gamma_i : i = 1, 2\}$, ads bought by all consumers $\{q_{1j}, q_{2j}\}_{j=1}^J$. Let ψ denote all the parameters excluding $F(\cdot)$. There are some additional assumptions imposed on these parameters. We assume that the joint distribution $F(\cdot,\cdot)$ is defined on $\Theta := [\underline{\theta}_1, \overline{\theta}_1] \times [\underline{\theta}_2, \overline{\theta}_2]$ such that $F(\cdot,\cdot)$ is absolutely continuous with continuously differentiable and nowhere vanishing density $f(\cdot,\cdot)$, and ψ is such that (i) $b_1b_2 - c^2 > 0$;

(ii) $b_i + \beta_i > 0$ for $i = 1, 2$; and (iii) $(b_1 + \beta_1)(b_2 + \beta_2) - 2c^2 > 0$, to ensure the utility function is concave and optimization problem is convex.

Two structures $\{F(\cdot,\cdot), \psi\} \neq \{\tilde{F}(\cdot,\cdot), \tilde{\psi}\}$ are said to be observationally equivalent if they imply the same probability distribution of the observed data, and the model is said to be identified if there are no two observationally equivalent structures.

For a fixed parameter set ψ, consider the following allocation rule

$$\mathbf{q}(\theta;\psi) = \begin{pmatrix} q_1(\theta_1,\theta_2;\psi) \\ q_2(\theta_1,\theta_2;\psi) \end{pmatrix} : \Theta \longrightarrow \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \in \mathbb{R}^2_+,$$

which can be written in terms of $z = (z_1, z_2)$ as

$$\mathbf{q}(z;\psi) = \begin{pmatrix} q_1(z_1;\psi) \\ q_2(z_2;\psi) \end{pmatrix} : \Theta \longrightarrow \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \in \mathbb{R}^2_+.$$

The uniqueness result implies that the mapping from type z to allocation/choice $\mathbf{q}$ is monotonic and unique. So, if we can show that this mapping is invertible, from choice $\mathbf{q}$ to type z then we can identify the joint density of z. Since $(z_1, z_2) \mapsto (\theta_1, \theta_2)$ is one-to-one, this will identify $f(\cdot,\cdot)$.

This, however, is infeasible because only few consumers choose ads from both publishers. As a result we have to study the nonparametric identification of the marginal densities of z_1 and z_2 separately. If the marginal densities are nonparametrically identified that is still not enough to recover the joint density. In view of that we propose using a copula to combine these two nonparametric marginals. In particular we propose to use the correlation between q_1 and q_2 among those who buy ads from both publishers to estimate the dependence for the copula. Under the assumption that the dependence between those who buy from both is the same as those who choose outside options, the joint density can be estimated.

One question that crops in the empirical analysis is which family of copula should we choose. Since the economic theory is silent about this, we rely on goodness of fit test and Vuong's model selection test to select the copula that provides the best fit. At the end we will choose the family that is selected by both of these tests.

Focus only on $[z_i^0, \bar{z}_i]$ where $q_i \geq q_{i0}$, then the incentive compatibility constraint implies that the equilibrium allocation rule $q_i(\cdot) : [z_i^0, \bar{z}] \mapsto [q_{i0}, \bar{q}_i]$ is monotonic, hence it can be inverted to provide a (inverse) mapping $z_i(\cdot) \equiv q_i^{-1}(\cdot)$, from ad choices to types, where $\bar{q}_i := \max_j q_{ij}$ is the largest ad size in the data. Then $z_{ij} = z_i(q_{ij})$ is the consumer $j's$ type who chose $q_{ij} > q_{i0}$ from publisher i. Let $H(\cdot, \cdot)$ be the conditional joint distribution of (q_1, q_2) given $q_i > q_{i0}$, and let $H_i(q_i)$ be the corresponding marginal density of q_i given $q_i > q_{i0}$. This implies the following relationship between the distribution of ad with publisher i and the distribution of $z_i, i = 1, 2$:

$$\begin{aligned} H_i(q) &:= \int_{q_{j0}}^{\bar{q}_j} H(q, y)dy = \Pr[q_i \leq q | q_i \geq q_{i0}] \\ &:= \Pr[z_i \leq z_i(q) | z_i > z_i(q_{i0})] \\ &= \frac{G_i(z) - G_i(z_i^0)}{1 - G_i(z_i^0)}, \end{aligned} \tag{18}$$

where $z = z_i(q)$ and $z_i^0 = z_i(q_{i0})$.
Differentiating (18) gives the conditional density $h_i(q) = \frac{g_i(z)}{1-G_i(z_i^0)} z_i'(q)$, which then gives the following relationship that we use later for identification:

$$\frac{1 - G_i(z_i)}{g_i(z_i)} = \frac{1 - H_i(q)}{h_i(q)} z_i'(q), \tag{19}$$

where after differentiating (16) (and an analogous function for $i = 1$) we get

$$z_i'(q_i) = T_i''(q_i) + b_1 - \frac{2c^2}{(b_{-i} + \beta_{-i})}.$$

Identification of Cost Parameters. One of the results in the price-discrimination literature is that the seller never distorts the output meant for the highest type, even though the output is distorted for everyone else. In other words, the equilibrium allocation is such that the output meant for the highest type, i.e., the largest output, equates marginal utility to marginal cost. This feature of the nonlinear pricing is referred to as "no-distortion-on-top," and helps to identify the marginal costs.

Consider the seller $i = 2$. From the data one can identify $\bar{q}_2 = \max\{q_{2i}; j = 1, \ldots, J\}$, and the monotonicity of $q_2(\cdot)$ means $\bar{q}_2 = q_2(\bar{z}_2)$. Then the pricing function gives

$$T_2'(q_2(\bar{z}_2)) = T_2'(\bar{q}_2) = \alpha_2 + \beta_2 \bar{q}_2,$$

which when used in (15) identifies m_2 as

$$m_2 = \alpha_2 + \beta_2 \bar{q}_2. \tag{20}$$

Using the same argument one can identify the marginal cost as[9]

$$m_1 = \alpha_1 + \beta_1 \bar{q}_1. \tag{21}$$

Identification of Utility Parameters. In this section we will consider the identification of the utility parameters. Since the utility function is concave, and the parameters b_1 and b_2 and c govern the mix of q_1 versus q_2 we can identify these parameters from those who choose extreme ad choices. Suppose the utility was linear, i.e., $b_1 = b_2 = 0$, then consumers will care only about ad $(q_1 + q_2)$, but not the composition, so let $q_2 = 0$. But if $b_1 > 0$, the marginal utility from q_1 falls and q_2 starts to become important leading to $q_2 > 0$.

Heuristically speaking, as the utility function becomes more concave the choice will become less and less asymmetric. This constraint is therefore most applicable to the highest type $\bar{z}_1-$ who buys the most asymmetric options: the maximum q_1 and minimum q_2. Hence, the value of b_1 must be small enough to rationalize this choice, and since $\bar{z}_1$ has an interior solution, the choice $\bar{q}_1$ must equate marginal utility and marginal price, i.e.,

$$\bar{\theta}_1 - b_1 \bar{q}_1 + c q_{20} = \alpha_1 + \beta_1 \bar{q}_1.$$

[9] One can show that this identification is robust with respect to wether or not VZ and OG move simultaneously or sequentially.

This identifies c conditional on $\overline{\theta}_1$ and b_1. Similarly, the optimality for the $\overline{z}_2-$ type

$$\begin{aligned} & \overline{\theta}_2 - b_2\overline{q}_2 + cq_{10} = \alpha_2 + \beta_2\overline{q}_2 \\ \Rightarrow \quad & b_2 = \frac{cq_{10} + \overline{\theta}_2 - \alpha_2}{\overline{q}_2} - \beta_2 \end{aligned}$$

identifies b_2 as a function of $\overline{\theta}_2$. (See next step for identification of the supports of the densities.) Therefore, c and b_2 are identified from $\{\overline{\theta}_1, \overline{\theta}_2, b_1\}$. For any $q_i < \overline{q}_i$, we rewrite optimal allocation rule using (19) as

$$\alpha_i + \beta_i q_i = m_i + \frac{1 - H_i(q_i)}{h_i(q_i)} \left(\beta_i + b_i - \frac{2c^2}{b_j + \beta_j} \right), \quad i, j \in \{1, 2\}, i \neq j,$$

so that at $q_1 \neq \tilde{q}_1$ gives

$$\begin{aligned} b_1 + \beta_1 &= \frac{\alpha_1 + \beta_1 q_1 - m_1}{\frac{1 - H_1(q_1)}{h_1(q_1)}} + \frac{2c^2}{b_2 + \beta_2}, \\ b_1 + \beta_1 &= \frac{\alpha_1 + \beta_1 \tilde{q}_1 - m_1}{\frac{1 - H_1(\tilde{q}_1)}{h_1(\tilde{q}_1)}} + \frac{2c^2}{b_2 + \beta_2}. \end{aligned}$$

Equating these two equations identifies b_1 as

$$b_1 = \frac{1}{2} \left(\frac{\alpha_1 + \beta_1 q_1 - m_1}{\frac{1 - H_1(q_1)}{h_1(q_1)}} + \frac{\alpha_1 + \beta_1 \tilde{q}_1 - m_1}{\frac{1 - H_1(\tilde{q}_1)}{h_1(\tilde{q}_1)}} \right) - \beta_1. \tag{22}$$

Identification of Marginal Densities. First, we consider the identification of the support of the densities, which, as will be evident soon, requires location normalization. Evaluating $q_2(z_2)$ in (14) at $(\overline{z}_2)$ identifies

$$\overline{z}_2 = \overline{q}_2 \left(b_2 - \frac{2c^2}{b_1 + \beta_1} \right) + m_2 + \frac{c^2 q_{20} + c\alpha_1}{b_1 + \beta_1},$$

which together with the definition of $\overline{z}_2$ identifies $\overline{\theta}_2$ as

$$\overline{\theta}_2 = \overline{q}_2 b_2 + m_2 + \frac{c^2 q_{20} + c\alpha_1 - c\underline{\theta}_1 - 2c^2\overline{q}_2}{b_1 + \beta_1}. \tag{23}$$

Since some consumers choose (q_{10}, q_{20}), normalization is important to determine the support. We assume that the type space is $\Theta = [\underline{\theta}_1, \overline{\theta}_1] \times [0, \overline{\theta}_2]$ and normalize the utility that the lowest types, i.e., $(\underline{\theta}_1, \underline{\theta}_2)$, get from the lowest quantities, i.e., (q_{10}, q_{20}), to be zero.

Assumption 2. *Normalization: Let $\underline{\theta}_2 = 0$ and $u(q_{10}, q_{20}; \underline{\theta}_1, \underline{\theta}_2) = 0$.*

This then identifies $\underline{\theta}_1$ as

$$\underline{\theta}_1 = \frac{b_1}{2} q_{10} + \frac{b_2}{2} \frac{q_{20}^2}{q_{10}} - c q_{20}. \tag{24}$$

Next, we consider the identification of the marginal densities of types. For those in C_b, (17) and (14) can be inverted and use (19) to identify $G(\cdot, \cdot | z_1 \geq z_1^0, z_2 \geq z_2^0)$ from

$$\begin{pmatrix} z_{1j} \\ z_{2j} \end{pmatrix} = \begin{pmatrix} q_1^{-1}(q_{1j}) \\ q_2^{-1}(q_{2j}) \end{pmatrix} = \begin{pmatrix} q_{1j}\left(b_1 - \frac{2c^2}{b_2+\beta_2}\right) + m_1 + \frac{c\alpha_2 + c^2 q_{10}}{b_2+\beta_2} + \frac{(1-H_1(q_{1j}))}{h_1(q_{1j})} z_1'(q_{1j}) \\ q_{2j}\left(b_2 - \frac{2c^2}{b_1+\beta_1}\right) + m_2 + \frac{c\alpha_1 + c^2 q_{20}}{b_1+\beta_1} + \frac{(1-H_2(q_{2j}))}{h_2(q_{2j})} z_2'(q_{2j}) \end{pmatrix}, \tag{25}$$

where $z_i'(q_{ij}) = (\beta_i + b_i) - 2c^2/(b_{-i} + \beta_{-i})$, with $i = 1, 2; j = 1, \ldots, 6328$. The marginal distributions $G_i(\cdot | z_i \geq z_i^0)$ can be identified by inverting the appropriate allocation rule, i.e., $q_i(\cdot)$ for $i = 1, 2$.[10]

5. Estimation

We observe (q_1, q_2) for every consumer j where $j = 1, 2, ..., 6328$. Using the optimal allocation functions, define the econometric model as

$$q_{ij} = \begin{cases} [z_{ij} - \frac{1-G_i(z_{ij})}{g_i(z_{ij})} - m_i - \frac{c\alpha_{-i} + c^2 q_{i0}}{b_{-i}+\beta_{-i}}] / [b_i - \frac{2c^2}{b_{-i}+\beta_{-i}}], & z_{ij} \in (z_i^0, \overline{z}_i] \\ q_{i0}, & z_{ij} \in [\underline{z}_i, z_i^0] \end{cases} \tag{26}$$

with $i = 1, 2; j = 1, \ldots, 6328$. In this model, the latent consumer type z_{ij} is the source of randomness, in other words, z_{ij} plays the role of the "error" in the usual regression models. However, the model also depends on $\frac{1-G_i(\cdot)}{g_i(\cdot)}$ which is unknown, but

[10] Henceforth, we use (z_1, z_2), to mean one of these combinations: $(z_1, z_2), (z_1, z_2^0), (z_1^0, z_2)$ and (z_1^0, z_2^0), depending on whether (z_1, z_2) is in C_b, C_1, C_2 and C_0, respectively.

from the identification arguments we know it can be replaced with $\frac{1-H_i(q_{ij})}{h_i(q_{ij})}z_i'(q_{ij})$. Since $\frac{1-H_i(q_{ij})}{h_i(q_{ij})}$ can be estimated nonparametrically (described below), we can replace it with its estimate $\frac{1-\hat{H}_i(q_{ij})}{\hat{h}_i(q_{ij})}$. Besides these two equations, one for each seller, there are also the following vector of conditions that identify ψ:

$$s(\psi) = \begin{pmatrix} m_1 - \alpha_1 + \beta_1\overline{q}_1 \\ m_2 - \alpha_2 - \beta_2\overline{q}_2 \\ 2(b_1+\beta_1) - \frac{\alpha_1+\beta_1 q_1 - m_1}{\frac{1-H_1(q_1)}{h_1(q_1)}} - \frac{\alpha_1+\beta_1\tilde{q}_1 - m_1}{\frac{1-H_1(\tilde{q}_1)}{h_1(\tilde{q}_1)}} - \beta_1 \\ \overline{q}_2 c q_{10} - \alpha_2 - (b_2+\beta_2)\overline{q}_2 + \overline{\theta}_2 \\ \overline{\theta}_1 - \overline{q}_1(b_1-\beta_1) - \alpha_1 + cq_{20} \\ (b_2+\beta_2)\overline{q}_2 - cq_{10} - \overline{\theta}_1 + \alpha_2 \\ (\overline{\theta}_2 - \overline{q}_2 b_2 - m_2)(b_1+\beta_1) - c^2 q_{20} - c\alpha_1 + c\underline{\theta}_1 + 2c^2\overline{q}_2 \\ (\underline{\theta}_1 + cq_{20})2q_{10} - b_1 q_{10}^2 - b_2 q_{20}^2 \end{pmatrix} = 0 \tag{27}$$

Therefore the estimation procedure consists of the following two steps: (1) Estimate the inverse hazard function $(1-H_i(\cdot))/(h_i(\cdot))$ for $i = 1,2$ using a kernel density estimator; (2) plug-in these estimates in (26) and (27), and estimate the parameters using the nonlinear least squares method.

5.1. **Estimating Marginal Densities.** Let N_1^* and N_2^* denote the number of firms purchasing advertising space strictly larger than q_{10} and q_{20}, respectively and q_{ij} denotes the quantity purchased by each of those firms from $i = 1,2$. To estimate $H_i(\cdot)$ and $h_i(\cdot)$ we use the empirical distribution and kernel density estimators, respectively:

$$\hat{H}_i(q) = \frac{1}{N_i^*}\sum_{j=1}^{N_i^*} \mathbb{1}(q_{ij} \le q), \text{for } q \in [q_{i0}, \overline{q}_i]; \tag{28}$$

$$\hat{h}_i(q;\xi) = \frac{1}{N_i^*}\sum_{j=1}^{N_i^*} \frac{1}{\xi} K\left(\frac{q-q_{ij}}{\xi}\right), \tag{29}$$

where ξ is a bandwidth, and $K(\cdot)$ is a kernel.

It is known that: (a) the kernel density estimation suffers from lack of local adaptability, i.e. it is sensitive to outliers and spurious bumps [Marron and Wand, 1992; Terrell and Scott, 1992]; (b) it surfers from boundary bias; and (c) the most widely used data-driven bandwidth selection method, the plug-in method, is adversely affected by the normal-reference rule [Jones, Marron, and Sheather, 1996; Devrôye, 1997]. So, we use the adaptive kernel density estimator proposed by Botev, Grotowski, and Kroese [2010], which addresses these problems while ensuring that the estimated inverse hazard function is well defined, i.e. $\lim_{q_i \to \overline{q}_i} \frac{1-\hat{H}_i(q_i)}{\hat{h}_i(q)} = 0$ for $i = 1, 2$. For the bandwidth selection procedure see Botev, Grotowski, and Kroese [2010].

The next lemma characterizes the consistency properties. It is a direct application of the results in [Guerre, Perrigne, and Vuong, 2000], and therefore we omit the proof.

Lemma 1. *Suppose all the assumptions mentioned so far are valid. Then:*

(1) $\sup |\hat{\overline{q}}_i - \overline{q}_i| \overset{a.s}{\to} 0$ *and* $\sup |\hat{q}_{i0} - q_{i0}| \overset{a.s}{\to} 0$.
(2) $\hat{\overline{q}}_i = \overline{q}_i + O_{a.s}[(\log \log N_i^*)/N_i^*]$.
(3) $\sup_{q \in (q_{i0}, \overline{q}_i]} || \log[(1 - \hat{H}_i^*(q))/(1 - H_i^*(q))]|| \overset{a.s}{\to} 0$.
(4) *For any* $q_i \in (q_{i0}, \overline{q}_i)$, $\sup_{q_i \in (q_{i0}, \overline{q}_i]} |\hat{z}_i(\cdot) - z_i(\cdot)| \overset{p}{\to} 0$ *as* $N_i^* \to \infty$.
(5) $\sup_{z_i \in (z_i^0, \overline{z}_i]} |\hat{g}_i^*(z_i) - g_i^*(z_i)| \overset{a.s}{\to} 0$ *as* $N_i^* \to \infty$, *where* $g_i^*(\cdot)$ *is the conditional density given* $z_i > z_i^0$.

Recall that $\hat{\overline{q}}_i$ is the sample estimate of the highest quality offered by publisher i, likewise $\hat{z}_i$ is the pseudo aggregated type and $\hat{g}_i^*(\cdot)$ is the estimate of the conditional density given $z_i > z_i^0$. Next, we address the estimation of the joint density of types.

5.2. **Estimating Joint Density.** We are interested in estimating the joint distribution $F(\cdot, \cdot)$ of (θ_1, θ_2). Since the truncated distributions of z_1 and z_2 are nonparametrically identified, without further assumption the best we can do is (Fréchet) bound the joint CDF as

$$\max\{G_1(z_1) + G_2(z_2) - 1, 0\} \leq G(z_1, z_2) \leq \min\{G_1(z_1), G_2(z_2)\}.$$

Instead of estimating a bound, we use copula to estimate the joint density. The basic idea is simple. Once we pick a one-parameter family of copula, we can use the sample of consumers who buy from both sellers to estimate the dependence (correlation) between z_1 and z_2. Then under the assumption that this dependence is the same for those who choose the outside option (standard listing) we can extend the density everywhere.

A function $C : [0,1]^2 \to [0,1]$ is a two-dimensional copula if $C(\cdot,\cdot)$ is the joint distribution of a random variable in $[0,1]^2$ with uniform marginal distributions. Since $G_i(\cdot)$ is a uniform random variable, the copula representation of $G(z_1, z_2)$ is $C(z_1, z_2) := C(G_1(z_1), G_2(z_2))$. Sklar's Theorem [Nelson, 1999] guarantees that $C(\cdot,\cdot)$, is unique but is not identified.

Let $C(\cdot,\cdot)$ be known up to one parameter κ, i.e., $C(\cdot,\cdot) \in \mathcal{C}_0 = \{C_\kappa : \kappa \in \Gamma \subset \mathbb{R}\}$, where Γ is the parameter set that contains κ. If we knew the copula family , i.e., if the null $H_0 : C \in \mathcal{C}_0$, was known to be true, then the parameter κ could be estimated either by maximizing the joint likelihood function or by matching some measure of dependence such as Kendall's τ, or Spearman's ρ. One of the key limitation in implementing the copula method is that often in empirical research the family of copula is unknown. There are many one-parametric families of copulas, such as Gaussian, $t-$ copula, Clayton, Archemedean etc.[11]

Given this indeterminacy problem, we proceed in two steps: First, using a model selection criteria we select the copula family that provides the best fit; Second, we estimate the parameter of that family using pseudo maximum likelihood. Since the marginal distribution of z_i is unspecified, we can replace it by its empirical counterpart $\hat{G}_i(\cdot) = \frac{1}{J}\sum_{j=1}^{J} \mathbb{1}(z_{ij} \leq \cdot)$. It is easier to work with (imputed) rank $\hat{r}_{ij}$ instead

[11] For example, Clayton copula is $C(G_1, G_2) = \left(\max\left\{G_1^{-\kappa} + G_2^{-\kappa} - 1; 0\right\}\right)^{-1/\kappa}$.

of the variable z_{ij} and view the copula to be based on a collection of pseudo values $(u_1,\dots,u_J) \in \mathbb{R}^{2J}$ where $\hat{u}_{ij} := r_{ij}(J+1) = \hat{G}_i(z_{ij}) \times J/(J+1)$.[12]

The first thing we check is if z_1 and z_2 are independent, even if q_1 and q_2 are not. If they are independent then the joint distribution is simply the product of their marginal distributions. We test the following hypothesis:

$$H_0 : \forall (u_1,u_2) \in [0,1]^2, C(u_1,u_2) = u_1u_2.$$
$$H_A : \exists (u_1,u_2) \in [0,1]^2, C(u_1,u_2) \neq u_1u_2.$$

Let $C_J(u_1,u_2) = J^{-1}\sum_{j=1}^J \mathbb{1}(\hat{u}_{1j} \leq u_1, \hat{u}_{2j} \leq u_2)$ be the empirical copula. Then we compute the classic Cramér- von Mises statistic

$$T_N = \int_{[0,1]^2} J\left\{C_J(u_1,u_2) - u_1u_2\right\}^2 du_1du_2.$$

The test statistic is estimated to be $\hat{T}_J = 1.66467$ with the $p-$ value equal to 0.0005. Therefore we conclude that z_1 and z_2 are not independent.[13]

Since $C_J(\cdot,\cdot)$ is a consistent estimator of $C(\cdot,\cdot)$ [Fermanian, Radulović, and Wegkamp, 2004] the distance between the estimated copula $C_{\kappa J}$ and $C_J(\cdot,\cdot)$ can be used as a goodness-of-fit criteria –under the null that the said family is true. Let $\hat{\kappa}$ be the estimate of the parameter that maximizes the pseudo-log-likelihood, [Genest, Ghoudi, and Rivest, 1995; Genest, Quessy, and Rémillard, 2006], i.e.,

$$\hat{\kappa} = \arg\max_{\kappa\in\Gamma}\left\{ l(\kappa) := \sum_{j=1}^N \log\left[C_\kappa(\hat{u}_{1j},\hat{u}_{2j})\right]\right\},$$

Genest, Rémillard, and Beaudoin [2009] show that the Cramér-von Mises statistic

[12] This transformation is without loss of generality because copulas are invariant to continuous, strictly increasing transformations. The scaling factor $J/(J+1)$ ensures that the copula is well behaved at the boundary of $[0,1]^2$.

[13] Since (a) the asymptotic distribution of T_J under the null is not distribution free, [Genest and Rémillard, 2004]; (b) the distribution is also affected by the first-step errors from estimating the pseudo z_1 and z_2; we compute the critical values using the bootstrap procedure proposed by Genest and Rémillard [2004] and Kojadinovic and Holmes [2009].

Family	$\hat{\kappa}$	CvM $p-$ value	Vuong Test
Gumbel-Hougaard	1.12	0	2
Clayton	0.09	0	-6
Frank	0.24	0	-2
Gaussian	0.63	0	-4
Plackett	1.618	0	2
t (df=4)	0.16	0	2
Joe	1.2	0.12	6

TABLE 3. Goodness-of-Fit and Vuong test results: Estimated parameters of copula based on pseudo-MLE and $p-$ values of the Cramér-von Mises statistic are computed using 10,000 bootstrap replications and the rank in Vuong test.

$$\mathbb{T}_J := \sum_{i=1}^{J} \left\{ C_J(\hat{u}_{1j}, \hat{u}_{2j}) - C_{\kappa_J}(\hat{u}_{1j}, \hat{u}_{2j}) \right\}^2$$

can be used as a consistent goodness-of-fit criteria, where its asymptotic distribution follows from results in weak convergence properties of empirical copula; see Fermanian, Radulović, and Wegkamp [2004].

Using this result, the $p-$ value can be approximated from the limiting distribution of $\mathbb{T}_N$. Approximating the $p-$ value, however, is computationally costly because the limiting distribution depends on the asymptotic behavior of $\mathbb{C}_J$ and on the estimator $\hat{\kappa}$.[14] Therefore, the approximate $p-$ values can only be obtained from the bootstrap procedure outlined by Genest and Rémillard [2008]. In practice this is a slow method so we use the multiplier central limit theorem to determine the large sample distribution of the test statistic; see Kojadinovic and Yan [2011]. We implement Cramér-von Mises test for seven widely used families of copula and for each family estimate the test statistic, the corresponding parameter $\hat{\kappa}$ and the $p-$ value based on 10,000 replications. The $p-$ values are reported in the third column of Table 3. It is evident then that only the Joe copula provides the best fit, with the estimated parameter $\hat{\kappa} = 1.206497$ $(s.e. = 0.0137)$.

[14] Using pseudo-log-likelihood is just one of many ways to estimate the parameters in the literature. For robustness, we also estimated the parameters that maximize Kendall's tau and Spearman's Tho, and reach the same conclusion.

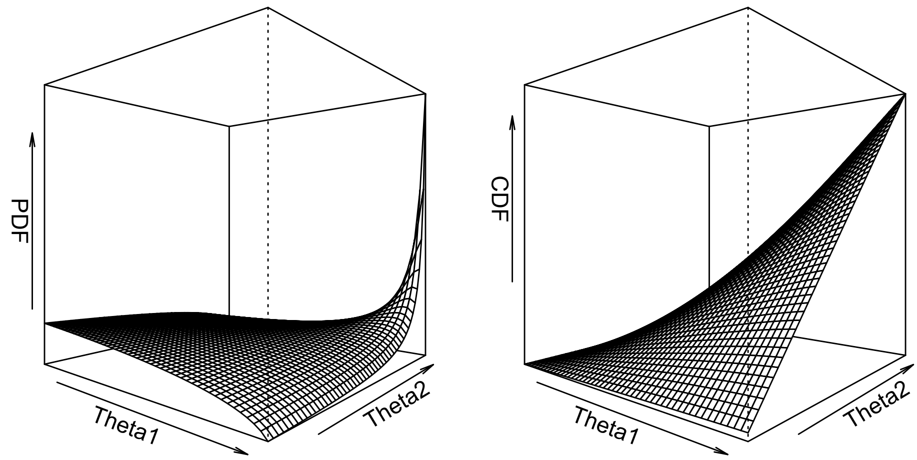

FIGURE 4. Estimated Joint pdf and cdf of recovered $(\hat{\theta}_1, \hat{\theta}_2)$

For robustness, we also consider Vuong [1989] non-nested model selection tests, where $p-$ values are calculated using the bootstrap procedure in Clarke [2007]. To implement the test, we perform a pairwise comparison for all copula families from the seven families, and give $+1$ to the one that is selected and -1 to the other family. The results are in the last column in Table 3. As can be seen, Joe copula has the highest score of 6, meaning it was selected in all pairwise comparison using Vuong's test.

Then the estimated joint density of (z_1, z_2) is

$$\hat{g}(z_1, z_2) = (\hat{\kappa} - 1)\left(1 - \prod_{i=1}^{2}\left\{(1 - (1 - \hat{G}_i(z_i))^{\hat{\kappa}})(1 - \hat{G}_i(z_i))^{\hat{\kappa}-1}\hat{g}_i(z_i)\right\}\right).$$

and the corresponding joint density of (θ_1, θ_2), see Figure 4, evaluated at $z_j(\theta) \equiv z_j(\theta_1, \theta_2)$ is

$$\hat{f}(\theta_1, \theta_2) = \left(1 - \prod_{j=1}^{2} \left\{(1 - (1 - \hat{G}_i(z_i(\theta)))^{\hat{\kappa}})(1 - \hat{G}_i(z_i(\theta)))^{\hat{\kappa}-1}\hat{g}_i(z_i(\theta))\right\}\right) \times (1 - \hat{\kappa}) \left(1 - \frac{\hat{c}^2}{(\hat{b}_1 + \hat{\beta}_1)(\hat{b}_2 + \hat{\beta}_2)}\right).$$

5.3. **Estimation Results.** The estimated gross utility function becomes

$$\hat{u}(q_1, q_2, \theta_1, \theta_2) = \theta_1 q_1 - \frac{1.45}{2} q_1^2 + \theta_2 q_2 - \frac{0.414}{2} q_2^2 - 0.02 \times q_1 \times q_2.$$

As can be seen $\hat{c} < 0$, which shows that the two ads can be treated as substitutes, although the rate of substitution is weak. The marginal cost of printing for VZ at $\hat{m}_1 = 7.768$ is twice as that of OG at $\hat{m}_2 = 3.145$, which captures the differences in paper size and quality. The support is estimated to be $[109.39, 896.15] \times [0, 896.15]$. Recall that z_i^0 is the threshold type below which consumers buy q_{i0}.

Armstrong [1996] showed that in a multidimensional screening, it is always optimal for the seller to price the goods in such a way that some positive fraction of consumers are not served. The threshold type z_i^0 then depends on the density of consumer type, e.g., if $G_i(\cdot)$ has thicker lower tail than upper tail then z_i^0 should be closer to $\underline{z}_i$ as fewer types should be excluded and vice versa. The estimates of the threshold types are $z_1^0 = 978.51$ and $z_2^0 = 298.83$, for VZ and OG, respectively, which suggests that $\hat{g}_2(\cdot)$ has relatively more mass at the lower end than $\hat{g}_1(\cdot)$.

This means that competition between VZ and OG at the lower end is more severe than at the upper end. This is reflected in the differences in prices: the difference in average price per pica widens as we move from lower category to higher, see Table 1. And the fact that VZ's prices are consistently higher across comparable categories than of OG's suggests that VZ enjoys a higher brand effect.

	$\hat{\theta}_1$	$\hat{\theta}_2$
no. of Firms	5.25 (3.717)	0.16 (0.1751)
Sq. no. of Firms	-0.56 (0.25) (**)	-0.0004 (0.001)(*)
Avg. Size	10.53 (0.25) (**)	1.46 (0.45) (**)
Std. Size	0.69 (.39) (*)	-0.01 (0.11)
National	-626.34 (379.77)(*)	106.57 (24.22) (**)
Guide	669.36 (249.17) (**)	328.06 (16.13)(**)

TABLE 4. OLS: Result of regression of pseudo types $\hat{\theta}$ on number of business units and its square under the same heading, the average size of ad bought under the heading, the standard deviation of the size, whether or not the firm is national and if it opts for guide option. Standard errors are reported in the parenthesis and $(^*)$ denotes significance at 10% and $(^{**})$ at 5%.

Although we treated the willingness to pay for ads as given, using the estimated (θ_1, θ_2) we verify, albeit ex-post, if the number of consumers in the same business-category affects the willingness to pay. In other words, one can check if a monopoly-dentist has lower taste for ads than, say, a duopoly-dentist. The null hypothesis of the structural model is that the willingness to pay should not depend on the number of other consumers in the same category.

To test this hypothesis we control for other covariates, such as whether or not the consumer j is a national brand ($Nat_{ij} \in \{0,1\}$) and whether or not j opts for guide option ($Guide_{ij} \in \{0,1\}$) with publisher i. We control for guide option because it provides additional advertising space by listing specialities and covers attorneys, dentists, physicians, insurance companies, etc. So we expect it to have a positive effect on consumer type. We estimate the following model for each publisher:

$$\hat{\theta}_{ij} = a_{i0} + a_{i1}\#C_{ij} + a_{i2}(\#C_{ij})^2 + a_{i3}\mu(q_{ij}) + a_{i4}\sigma(q_{ij}) + a_{i5}Nat_{ij} + a_{i6}Guide_{ij} + \epsilon_{ij},$$

where $\hat{\theta}_{ij}$ is consumer $j's$ estimated preference for publisher i's ads, $\#C_{ij}$ is the number of consumers who share the same sub-heading (e.g., hair salon) with consumer j and $(\#C_{ij})^2$ is its square. $\mu(\cdot)$ and $\sigma(\cdot)$ are the average and standard deviation of ads bought by consumers in the same industry as j. (Average ad bought within an industry should have a positive effect on θ.) The result from the regression for VZ and OG are presented in Table 4.

As can be seen, the result is consistent with the hypotheses. The number of consumers under the same business sub-heading does not affect the consumer's valuation for ads. On the other hand, the square of number of consumers in the same business sub-heading has negative effect on valuation for ads. Moreover, the estimates suggest that consumers who have national presence have lower value for ads in VZ while higher value for ads in OG. The estimates for VZ are consistent with the hypothesis that for a national brand the need to advertise locally might be less important than for a local brand. The result for OG, however, is surprising. This could be because the dummy variable captures the effect of differences in VZ and OG and the effect of consumer being a national brand. To tease out these effects separately we need model the demand for ads, and it is left for future research.

5.4. **Cost of Asymmetric Information.** What is the welfare implication of asymmetric information between the sellers and the buyers in this market? In this section we estimate the welfare cost using a simple counterfactual exercise where both sellers know consumers' type (θ_1, θ_2). Under asymmetric information, if sellers offer distorted allocation (that does not equate marginal utility with marginal cost) then without asymmetric information we would expect efficient allocation and hence higher welfare. The difference between the new total welfare and what is observed in the data is the cost of asymmetric information.

To quantify the loss, we consider the case where VZ is the leader in the market and moves before OG and they know each consumer's type.[15] In the second stage (θ_1, θ_2)- consumer who buys $\tilde{q}_1$ from VZ and pays t_1 and buys q_2 from OG gets gross utility $D(q_2; \tilde{q}_1; \theta) = u(q_2, \tilde{q}_1; \theta) - t_1$. For such q_2 the maximum price she is willing to pay is

$$t_2(q_2) = \theta_2(q_2 - q_{20}) - \frac{b_2}{2}(q_2^2 - q_{20}^2) + c\tilde{q}_1(q_2 - q_{20}). \tag{30}$$

[15] We maintain the assumption that both sellers offer q_{10} and q_{20} for free. Alternatively, we could also assume that the two move simultaneously. The welfare cost under leader-follower model would generate the lower bound on welfare cost.

OG will make a take-it-or-leave-it offer of q_2 at t_2 that maximizes the profit $t_2(q_2) - m_2 q_2$. From Equation (30), OG's best response is $q_2(\tilde{q}_1) = \frac{\theta_2 + c\tilde{q}_1 - m_2}{b_2}$. Now, in the first period the maximum price VZ can charge for any q_1 is

$$\begin{aligned} t_1(q_1) &= \theta_1(q_1 - q_{10}) + \theta_2(q_2(q1) - q_2(q_{10})) - \frac{b_1}{2}(q_1^2 - q_{10}^2) \\ &- \frac{b_2}{2}(q_2(q_1)^2 - q_2(q_{10})^2) + c(q_1 q_2(q_1) - q_{10} q_2(q_{10})) - (t_2(q_1) - t_2(q_{10})), \end{aligned}$$

where $t_2(q_{10})$ can be determined by evaluating (30) at q_{10}. Then, $q_1 = (\theta_1 - m_1)/b_1$ maximizes the profit $t_1(q_1) - m_1 q_1$ and the corresponding q_2 (as a function of q_1) is $q_2 = [b_1(\theta_2 - m_2) + c(\theta_1 - m_1)]/[b_1 b_2]$.

Let $D_2(q_1^*, q_2; \theta)$ be the residual demand for OG when VZ sells q_1^*, then the profit function for OG is $\int_{q_{20}}^{q_2} D_2(q_1^*, y)\mathrm{d}y - K_2 - m_2 q_2$. Thus the best response is to choose q_2^* such that $D(q_1^*, q_2^*) = m_2$, which equates the marginal benefit of q_2^* to the marginal social cost of producing q_2^*. The optimal allocation for VZ can be determined along OG's best response function.

We find that VZ gains \$2,651,052,914 while OG gains \$48,330,062 and the consumers lose \$2,699,115,638. The resulting net social welfare gain is in the order of \$267,337. One would expect that under full information, the seller will extract all consumer surplus, but because (q_{10}, q_{20}) is free, the consumer's indirect utility under complete information will not be zero but will be equal to its valuation for (q_{10}, q_{20}), which increases with (θ_1, θ_2).

Table 5 presents the quantity pair under incomplete information, under full information, and the corresponding difference in utility. As predicted by the theory, since the quantity allocation is not distorted for the highest type, the difference in the quantity under the two informational regime decreases with the allocation under incomplete information. The total welfare loss amounts to approximately 3.8% of the sales revenue.

Qt: Incomplete Info.	Complete Info.	# Obs	Δ Utility
(101, 210)	(104, 210)	230	$87.706
(106, 231)	(108, 243)	53	$99.017
(137, 231)	(139, 243)	27	$138.871
(137, 248)	(139, 260)	31	$144.479
(137, 288)	(139, 300)	28	$159.312
(237, 432)	(240, 442)	9	$425.865
(572, 517)	(575, 527)	4	$1,900
(572, 843)	(575, 851)	1	$2,200
(1709, 697)	(1711, 706)	6	$15,000
(1709, 1154)	(1711, 1160)	3	$16,000
(3171, 2153)	(3173, 2153)	1	$55,000
(6330, 1621)	(6330, 1624)	1	$210,000

TABLE 5. Welfare cost of asymmetric information: Comparison of welfare under incomplete information and a counterfactual of complete information. The change in utility is in 000s of 2006 dollars.

6. CONCLUSION

In this paper we estimate a model of competitive nonlinear pricing using novel data on advertisements placed with two Yellow Pages directories. We use a model of competition among sellers who compete with nonlinear prices to attract consumers who are heterogenous with respect to their value for the two ads. We show that the model parameters can be identified if we exploit first-order conditions that characterize equilibrium allocation. We estimate the joint density of consumers' preferences, the marginal costs of the publishers and the common utility parameters.

This exercise highlights how we can use supply side theory to compensate for limited data. In my case we have very limited information about consumers, and we have data only from one market, even if it for only one period. In the process we are not only able to treat observed product varieties as endogenous variables, we can also explain differential competition observed in the data. That is to say we can rationalize the observation that the two sellers compete strongly for the lower end of the market than the upper end. This explains why the per unit prices diverge as we move up the size of an ad, which would not be possible under a linear price model.

The estimates support the hypothesis of heterogenous preferences and asymmetric information in the market for Yellow Pages ads. Using a counterfactual exercise, we find the welfare cost of asymmetric information to be approximately 3.8% of the observed sales.

A next step in this line of research would be to simulate the effect of a merger on product lines and subsequently welfare. To do that, we could use the estimate of $F(\cdot,\cdot)$ and solve the optimal nonlinear pricing for the multi-product monopoly using Rochet and Choné [1998]. Since this is a hard problem to solve analytically, we might be able to use the numerical approximation method suggested by Ekeland and Moreno-Bromberg [2010]. This is a promising line of research that we leave for future work.

Appendix A. Tables

				Color Category 1		Color Category 2		Color Category 3		Color Category 4	
VZ Picas	VZ Percentage	OG Picas	OG Percentage	VZ Price	OG Price	VZ Price	OG Price	VZ Price	OG Price	VZ Price	OG Price
Listing											
12	0.4%	9	0.5%	$0	$0						
18	0.6%	12	0.66%	$151	$134		$147				
27	0.89%	15	0.83%	$290	$240	$492	$278				
36	1.19%			$492		$845					
Space Listing											
54	1.79%	46	2.49%	$504	$490		$528				
72	2.38%	92	4.98%	$781	$587		$650				
108	3.58%	138	7.46%	$1,134	$1,008	$1,789	$1,096	$2,873			
144	4.77%	184	9.95%	$1,436	$1,154	$2,242	$1,231	$3,592			
216	7.15%	230	12.44%	$2,080	$1,276	$3,289	$1,363				
Display											
174	5.76%	211	11.43%	$1,638	$1,118	$2,458		$2,609	$1,398	$2,873	$1,624
208	6.90%			$1,915		$2,861		$3,049		$3,326	
355	11.77%	438	23.74%	$3,074	$1,722	$4,612		$4,927	$2,254	$5,381	$2,655
537	17.77%			$4,473		$6,703		$7,145		$7,812	
735	24.34%			$5,872		$8,808		$9,388		$10,256	
1,110	36.76%			$8,341		$12,512		$13,344		$14,579	
		592	32.11%		$2,163				$2,814		$3,328
1,485	49.18%	908	49.19%	$10,093	$3,372	$15,133		$16,128	$4,420	$17,640	$5,084
		1,220	66.15%		$4,491				$5,875		$6,936
3,020	100.00%	1,845	100.00%	$18,510	$6,324	$27,770		$29,610	$8,290	$32,395	$9,435
6,039	200.00%			$34,272		$51,434		$54,835		$60,002	

TABLE A-1. Menus (size-color and prices) offered by Verizon (VZ) and Ogden (OG).

Verizon	# Purchases	% Sales	Revenue	% Revenue
Standard Listing	2,302	33.74%	$0	0%
Listing	2,222	32.56%	$614,143	10.42%
Space listing	1,374	20.14%	$1,002,857	17.02%
Display	925	13.56%	$4,275,642	72.56%
Total	6,823	100.00%	$5,892,642	100.00%
Ogden				
Standard Listing	5,913	86.66%	$0	0%
Listing	484	7.09%	$105,805	12.75%
Space listing	167	2.45%	$98,341	11.85%
Display	259	3.80%	$625,441	75.40%
Total	6,823	100.00%	$829,587	100.00%

TABLE A-2. Distribution of Sales and Revenues by Sizes.

REFERENCES

ANDERSON, S. P., AND J. WALDFOGEL (2016): "Preference Externalities in Media Markets," in *Handbook of Media Economics*, ed. by S. P. Anderson, J. Waldfogel, and D. Stromberg, vol. 1A, chap. 1. Elsevier. 13

ARMSTRONG, M. (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, 64, 51–75. 4, 33

——— (2006): "Recent Developments in the Economics of Price Discrimination," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by R. Blundell, W. K. Newey, and T. Persson, vol. 2, pp. 97–141. New York: Cambridge University Press. 3

ARMSTRONG, M., AND J. VICKERS (2001): "Competitive Price Discrimination," *The RAND Journal of Ecoomics*, 32, 579–605. 4

ARYAL, G. (2016): "Identifying a Model of Screening with Multidimensional Consumer Heterogeneity," *Working Paper*. 4, 5

ARYAL, G., I. PERRIGNE, AND Q. H. VUONG (2016): "Identification of Insurance Models with Multidimensional Screening," *Working Paper*. 5

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890. 2

BERRY, S. T. (1994): "Estimation of Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25(2), 242–262. 2

BERRY, S. T., AND J. WALDFOGEL (2016): "Empirical Modeling for Economics of the Media: Consumer and Advertiser Demand, Firm Supply and Firm Entry Models for Media Markets," in *Handbook of Media Economics*, ed. by S. P. Anderson, J. Waldfogel, and D. Stromberg, vol. 1A, chap. 3. Elsevier. 13

BORENSTEIN, S., AND N. ROSE (1994): "Competition and Price Dispersion in the U.S. Airline Industry," *Journal of Political Economy*, 102, 653–683. 12

BOTEV, Z. I., J. F. GROTOWSKI, AND D. P. KROESE (2010): "Kernel Density Estimation Via Diffusion," *The Annals of Statistics*, 38(5), 2916–2957. 28

BUSSE, M., AND M. RYSMAN (2005): “Competition and Price Discrimination in Yellow Pages Advertising,” *The RAND Journal of Ecoomics*, 36, 378–390. 3, 6, 12

CHERNOZHUKOV, V., A. GALICHON, M. HALLIN, AND M. HENRY (2015): “Monge-Kantororvich Depth, Quantiles, Ranks, and Signs,” *Working paper*. 5

CHU, S., P. LESLIE, AND A. SORENSEN (2011): “Bundle-Size Pricing as an Approximation to Mixed Bundling,” *American Economic Review*, 101(1), 263–303. 11

CLARKE, K. A. (2007): “A Simple Distribution-Free Test for Nonnested Model Selection,” *Politica; Analysis*, 15(3), 347–363. 32

COHEN, A. (2008): “Package Size And Price Discrimination In The Paper Towel Market,” *International Journal of Industrial Organization*, 26(2), 502–516. 4

CORTS, K. S. (1998): “Third-Degree Price Discrimination in Oligopoly: All-Out Competition and Strategic Commitment,” *The RAND Journal of Economics*, 29(2), 306–323.

CRAWFORD, G., AND M. SHUM (2006): “Monopoly Quality Degradation in Cable Television,” *Journal of Law and Economics*, 49, 181–209. 3

DEVRÔYE, L. (1997): “Universal Smoothing Factor Selection in Density Estimation: Theory and Practice,” *Test*, 6, 223–320. 28

EINAV, L., A. FINKELSTEIN, AND M. R. CULLEN (2010): “Estimating Welfare in Insurance Markets using Variation in Prices,” *Quarterly Journal of Economics*, 125(3), 877–921. 3

EINAV, L., A. FINKELSTEIN, S. RYAN, P. SCHRIMPF, AND M. CULLEN (2012): “Selection on Moral Hazard in Health Insurance,” *American Economic Reveiw*, Forthcoming. 3

EINAV, L., M. JENKIS, AND J. LEVIN (2012): “Contract Pricing in Consumer Credit Markets,” *Econometrica*, 80(4), 1387–1432. 3

EKELAND, I., J. HECKMAN, AND L. NESHEIM (2002): “Identifying Hedonic Models,” *American Economic Reveiw*, 92(2), 304–309. 5, 13

——— (2004): “Identification and Estimation of Hedonic Models,” *Journal of Political Economy*, 112, 60–109. 5, 13

EKELAND, I., AND S. MORENO-BROMBERG (2010): "An Algorithm for Computing Solutions of Variationsal Problems wuth Global Convexity Constraints," *Numerische Mathematik*, 115, 45–69. 38

EPSTEIN, L. G., AND M. PETERS (1999): "A Revelation Principle for Competing Mechanisms," *Journal of Economic Theory*, 88, 119–160. 4, 17

FAN, Y. (2013): "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market," *American Economic Reveiw*, 103(5), 1598–1628. 2

FERMANIAN, J. D., D. RADULOVIĆ, AND M. H. WEGKAMP (2004): "Weak Convergence of Empirical Copula Processes," *Bernoulli*, 95, 119–152. 30, 31

GENEST, C., K. GHOUDI, AND L.-P. RIVEST (1995): "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrica*, 82(3), 543–552. 30

GENEST, C., J.-F. QUESSY, AND B. RÉMILLARD (2006): "Goodness-of-fit Procedures for Copula Models Based on the Probability Integral Transformation," *The Scandinavian Journal of Statistics*, 33, 337–366. 30

GENEST, C., AND B. RÉMILLARD (2004): "Tests for Independence and Randomness Based on Empirical Copula Process," *Test*, 13(2), 335–369. 30

GENEST, C., AND B. RÉMILLARD (2008): "Validity of the Parametric Bootstrap for Goodness-of-fit testing in semiparametric Models," *Annales de l'Institut Henri Poincaré: Probabilités et Statistiques*, 44, 1096–1127. 31

GENEST, C., B. RÉMILLARD, AND D. BEAUDOIN (2009): "Goodness-of-fit Tests for Copulas: A Review and Power Study," *Insurance: Mathematics and Economics*, 44, 199–213. 30

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–574. 5, 28

HECKMAN, J., R. L. MATZKIN, AND L. NESHEIM (2010): "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78(5), 1569–1591. 5

HOLMES, T. (1989): "The Effects of Third-Degree Price Discrimination in Oligopoly," *The American Economic Review*, 79(1), 244–250.

IVALDI, M., AND D. MARTIMORT (1994): "Competition under Nonlinear Pricing," *Annales d'Économie et de Statistique*, 34, 71–114. 2, 3, 4, 12

JONES, M. C., J. S. MARRON, AND S. J. SHEATHER (1996): "A brief survey of bandwitdh selection for density estimation," *Journal of Americal Statistical Association*, 91, 401–407. 28

KOJADINOVIC, I., AND M. HOLMES (2009): "Tests of Independence Among Continuous Random Vectors Based on Cramér-von Mises Functionals of the Empirical Copula Process," *Journal of Multivariate Analysis*, 100, 1137–1154. 30

KOJADINOVIC, I., AND J. YAN (2011): "A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems," *Statistics and Computing*, 21(1), 17–30. 31

LESLIE, P. J. (2004): "Price Discrimination in Broadway Theatre," *The RAND Journal of Economics*, 35, 520–541. 4

LUO, Y., I. PERRIGNE, AND Q. VUONG (2015): "Structural Analysis of Nonlinear Pricing," *Working Paper*. 5

MARRON, J. S., AND M. P. WAND (1992): "Exact Mean Integrated Error," *The Annals of Statistics*, 10, 712–736. 28

MARTIMORT, D., AND L. A. STOLE (2002): "The Revelation and Delegation Principles in Common Agency Games," *Econometrica*, 70(4), 1659–1673. 4

MASKIN, E., AND J. RILEY (1984): "Monopoly with Incomplete Information," *The RAND Journal of Ecoomics*, 15, 171–196. 3, 11

MCMANUS, B. (2006): "Nonlinear Pricing in an Oligopoly Market: The Case of Speciality Coffee," *The RAND Journal of Ecoomics*. 4

MUSSA, M., AND S. ROSEN (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317. 3

NELSON, R. B. (1999): *An Introduction to Copulas*. Springer-Verlag New York, Inc. 29

OREN, S. S., S. A. SMITH, AND R. WILSON (1983): "Competitive Nonlinear Tariffs," *Journal of Economic Theory*, 29(1), 49–71. 4

ROCHET, J.-C. (1987): "A necessary and sufficient condition for rationalizability in a quasi-linear context," *Journal of Mathematical Economics*, 16(2), 191–200. 18

ROCHET, J.-C., AND P. CHONÉ (1998): "Ironing, Sweeping, and Multidimensional Screeing," *Econometrica*, 66(4), 783–826. 4, 19, 38

ROCHET, J.-C., AND L. A. STOLE (2002): "Nonlinear Pricing with Random Participation Constraints," *Review of Economic Studies*, 69, 277–311. 4

RYSMAN, M. (2004): "Competition Between Networks: A Study of the Market for Yellow Pages," *Review of Economic Studies*, 71, 483–512. 3

STOLE, L. A. (2007): "Price Discrimination in Competitive Environment," in *The Handbook of Industrial Organization*, ed. by M. Armstrong, and P. R. H., vol. III. Elsivier. 3, 20

TERRELL, G. R., AND D. W. SCOTT (1992): "Variable Kernel Density Estimation," *The Annals of Statistics*, 20, 1236–1265. 28

VUONG, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57(2), 307–333. 5, 32

WILSON, R. (1993): *Nonlinear Pricing*. Oxford University Press. 3, 11, 14