

# AN EMPIRICAL ANALYSIS OF COMPETITIVE NONLINEAR PRICING\*

GAURAB ARYAL<sup>†</sup>

**ABSTRACT.** In this paper I estimate unobserved consumer heterogeneity when the data contains information *only* about the demand/sales and prices from a single market, so that the canonical method to estimate demand is infeasible. In the market that I consider two sellers sell differentiated products by nonlinear prices, and I use competitive principal-agent model to formalize the supply side optimality conditions, which are then be used for identification. To rationalize the sales pattern, I model consumers to have two-dimensional types that are private information, which requires me to solve a competitive multidimensional screening problem. I use an endogenous aggregation method that reduces the two-dimensional types into one dimensional sufficient statistic. I apply this model to the data from a Yellow Pages advertisement market in central Pennsylvania where local business-units place ad with two Yellow Pages directories. I can identify only the truncated marginal densities of the sufficient statistics, so to estimate the joint density I use the copula family selected by both Cramér-von Mises and Vuong’s model selection tests. I find that the Joe copula provides the best fit and rationalizes some key data features. I also find there is a substantial heterogeneity among consumers and a substantial (3.8% of the sales) loss of welfare due to asymmetric information.

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

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

## 1. INTRODUCTION

The objective of this paper is to estimate consumer’s heterogeneity in a (single) competitive (duopoly) market where: i) two sellers offer a menu of options (quantity-price pairs) and consumers buy different amounts from either or both the sellers; iii) the data is from only a *single* market, moreover there is no other information about consumers’ characteristics. With such data the estimation procedures proposed by [Berry, 1994; Berry, Levinsohn, and Pakes, 1995] is infeasible, because their identification relies on detailed consumer characteristics and demand across multiple markets. So, I propose to rely on the optimality of the supply side to identify the (density of) consumer heterogeneity. In my model consumers have private information about their marginal willingness to pay for the two

---

*Date:* June 11, 2014.

\* Previously this paper was circulated under the title “Competition and Nonlinear Pricing in Yellow Pages.” I am thankful to Yao Huang for her help with an earlier version of the paper. I 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. I am solely responsible for any error.

<sup>†</sup> University of Chicago; e-mail: [aryalg@uchicago.edu](mailto:aryalg@uchicago.edu).

products (ad in my case) but the sellers only know the joint density of these types. Thus, I posit that the data is generated by a model of competitive nonlinear pricing (competitive principal-agent model). The fact that some buyers buy from both sellers, while others buy from only one or at all, suggests that buyers have at least two dimensional taste parameters that are unobserved by the sellers. This leads to a multidimensional screening model with competition. Using this model I study a market in central Pennsylvania where two Yellow Pages publishers offer many options, in the form of ad sizes and colors, to all local businesses. The data contains information on the universe of ad bought by each local business units and the sticker prices for the year 2006, but does not contain any information on other characteristics of the buyers.<sup>1</sup> Under the assumption that the observed data is the equilibrium outcome of the competitive nonlinear pricing, I show how the model parameters can be identified. In the process, the paper also highlights how the theory of screening can be used to estimate demand and supply when the data contains information about disaggregated sales in *only one* market, with very little or no information about consumer heterogeneity. The estimates suggest that there is a presence of asymmetric information in the market.

The presence of private or hidden information is now a widely accepted characteristic for most markets and is studied under the rubric of principal-agent problem. In contrast, however, the literature on demand estimation assumes that the sellers know everything about the consumer preferences, and any unobserved consumer heterogeneity is only from the point of view of the econometricians. However, without asymmetric information between the sellers and the consumers, it is hard to rationalize some of the key features of my data (see Section 2 for more) such as the fact that both the sellers offer a wide variety of quantity-price options, not just a uniform price, and some of the consumers choose to buy from both. This also motivates why I model the environment as competitive nonlinear pricing; see [Wilson, 1993] for a comprehensive treatment on nonlinear pricing. The theory of nonlinear pricing for a monopoly seller with only one dimensional private information is well understood, [Mussa and Rosen, 1978; Maskin and Riley, 1984; Wilson, 1993], and has subsequently been used in empirical analysis by [Crawford and Shum, 2006; Einav, Finkelstein, and Cullen, 2010; Einav, Jenkis, and Levin, 2012; Einav, Finkelstein, Ryan, Schrimpf, and Cullen, 2012] among others.

Most markets, however, are served by more than one seller and often consumers buy from more than one such seller, like in my data. Any empirical analysis of such markets would require us to allow for imperfect competition with multidimensional private information. But competitive nonlinear pricing is not without its own difficulties. For instance, in a duopoly model the revelation principal fails [Epstein and Peters, 1999; Martimort and Stole, 2002] and solutions can be determined

---

<sup>1</sup> Throughout the paper I use the terms business units, firms or consumers. The ad bought by each business units were manually recorded by reading the two directories. See the second section for more. I also know the type of the business, e.g. restaurant, salon, etc, which I never use.

only under assumptions that can be restrictive from empirical perspective. [Yang and Ye, 2008], for instance, assume that consumers can buy from only one seller and their vertical and horizontal preferences are independent and *uniformly* distributed, all of which are untenable with the data used in this paper (Figures 1 - 2).<sup>2</sup> Alternatively, if one only cares only about the demand side, then the random utility framework can be used as [Leslie, 2004; McManus, 2006; Cohen, 2008]. My aim will be to impose as few assumptions as possible about the joint density. The only way to achieve that goal is to also use the optimality of the supply side, which entails using an optimal allocation rule that maps unobserved types to observed sales. Moreover, ignoring the supply side limits the scope of such models when it comes to quantifying the effect of adverse selection on consumer welfare and to understand the effect of merger on welfare directly via price changes and indirectly via changes in product varieties offered. Also see [Borenstein, 1991; Lott and Roberts, 1991; Shepard, 1991; Borenstein and Rose, 1994; Verboven, 2002; Busse and Rysman, 2005]. Solving a principal-agent model with multidimensional taste parameters is a difficult problem, see [Armstrong, 1996; Rochet and Choné, 1998], and it is even harder to extend these models to incorporate more than one seller. For some analysis of this problem see [Oren, Smith, and Wilson, 1983; Armstrong and Vickers, 2001; Stole, 2007]. In the market I consider (explained shortly) ad vary in size and color, but color is not used to discriminate, which suggests that we can combine these two features into single dimensional attribute, *quality-adjusted quantity*. So each seller can be thought of as choosing quality-adjusted quantity while consumer vary in two dimensions, leading to bunching, i.e. the equilibrium allocation functions are many-to-one.<sup>3</sup> So the next question is how best to bunch consumers along their preferences.

I build on [Ivaldi and Martimort, 1994] duopoly model and index consumer type by a two-dimensional parameter without specifying the joint density.<sup>4</sup> Under the assumption that the utility function is quadratic and concave and the cost functions are linear (appropriate for publishing firm), each consumer can now be (endogenously) indexed by a one dimensional type. This “new” type combines the two-dimensional taste parameters and the the relative price differences (competition) together and acts as a “sufficient statistic” from the point of view of a seller. Let  $(\theta_1, \theta_2)$  be consumer type, where higher  $\theta_i$  reflects higher marginal utility from product  $i$ , ceteris paribus. Let the marginal prices be  $p_1$  and  $p_2$ . Then we can transform this type into two separate types  $z_i = t(\theta_1, \theta_2, p_{-i}), i = 1, 2$  such that from the point of view of seller  $i$ , knowing  $z_i$  is as good as knowing both  $\theta_1$  and  $\theta_2$ . So if the two ad choices are substitutes, then a consumer with highest  $\theta_i$  and lowest  $\theta_{-i}$  will have highest  $z_i$  (respectively lowest  $z_{-i}$ ) so will be treated as the highest (respectively lowest) valued customer by firm  $i$  (respectively firm  $-i$ ). The exact value of these two aggregates

<sup>2</sup> The hypothesis that the ad placed with the Yellow Pages directory are mutually independent was rejected.

<sup>3</sup> Bunching implies that two different types of consumers are allocated the same good.

<sup>4</sup> However, this joint density cannot be nonparametrically identified everywhere.

will depend on the prices, but for fixed prices the map  $(\theta_1, \theta_2) \mapsto (z_1, z_2)$  is one-to-one and onto. The sufficiency implies that the equilibrium can be determined by a monopoly principal-agent problem for seller  $i \in \{1, 2\}$  as if the consumer type is  $z_i$  and not  $(\theta_1, \theta_2)$ .

I apply the estimation procedure to a unique data on ad in two Yellow Pages directories in Center county, Pennsylvania. The data contains information about the menus of ad options and prices offered by two publishers Verizon and Ogden and the ad chosen by local business-units in the county. A working assumption of this paper is that ad is a final product for businesses. In equilibrium the allocation rules (mapping from consumer type  $z_i, i = 1, 2$  to the level of ad) are incentive compatible and hence monotonic. This means the equilibrium allocation rules can be inverted to identify (pseudo) aggregate types  $z_i$ 's, similar to how (pseudo) values are identified in independent private value auctions. See [Guerre, Perrigne, and Vuong, 2000]. I can identify the joint density of the consumer types for only those who buy from both. The truncated marginal density of  $z_i$ , however, can be identified from those who advertise in directory  $i$ ; the truncation is at the point where consumers don't advertise at all. I normalize the utility that a consumer with lowest value for both directories from not buying any ad to zero to fix the location of the utility. Then the common utility parameters and the firms' marginal costs can be identified. Then to recover the joint density of types, I use copula. Instead of choosing a family of copula in some ad hoc manner, I consider seven families of copula (these are some of the most widely used one parameter families in the literature) as competing models with an use one that is select by both Cramér-von Mises (goodness-of-fit) and [Vuong, 1989] (non-nested model selection) tests. This is in contrast to and better than the alternative where researchers choose parametric density because using copula requires fewer nontestable restrictions. The main assumption I need is that the joint density of consumer types have similar correlation at the two tails - so that I can use the correlation of ad choices for those who buy some ads and extend it to the area. However, the asymptotic distributions of these tests are nonstandard so I use the multiplier Bootstrap procedures in [Kojadinovic and Yan, 2011].<sup>5</sup>

The estimation results suggest: (a) the Joe copula provides the best fit for the joint density; (b) there is a substantial heterogeneity in how the two ad are valued; (c) competition is severe at the lower end of the market, which also has more mass, than the upper end of the market; (d) consumers treat the two ad as substitutes; and (e) counterfactual exercise shows that there is a loss of welfare, approximately 3.8% of sales revenue, due to asymmetric information. Since it is not clear if the sellers offer simultaneously or sequentially, I use only the best response (optimality) condition for price function pertaining to Ogden, the publisher who entered the market after Verizon. This condition

---

<sup>5</sup> Bootstrapping also corrects for estimation errors from prior steps in estimating the pseudo values and the utility and cost parameters. The first step is to estimate the densities of observed ad choices.

hold under both sequential and simultaneous move and hence ensures the identification does not depend on (unverifiable) timing assumption.

Furthermore, since I model both the product varieties/qualities and pricing rules as endogenous functions of unobserved consumers' type while allowing consumers to buy from both the sellers, this framework can potentially be useful to simulate mergers with endogenous product varieties as long as the assumptions about cost and utility functions are appropriate. This will require us to use the estimated joint density of types and simulate a multi product monopoly with multidimensional types as outlined in [Rochet and Choné, 1998] and compare the outcome with the data. This method should be viewed as an alternative to the using multiple discrete choice models of [Fan, 2013], when data from multiple markets are not available, but we still want to allow product varieties and prices to change endogenously with mergers. It is, however, difficult to implement [Rochet and Choné, 1998] with empirically estimated joint density of types, which is beyond the scope of this paper. See [Ekeland and Moreno-Bromberg, 2010] for more on computational difficulties in solving multidimensional principal-agent models. The paper is also related to papers that study identification of principal-agent models [Aryal, Perrigne, and Vuong, 2010; Perrigne and Vuong, 2011b], and the hedonic models [Ekeland, Heckman, and Nesheim, 2002, 2004]; both these strands of literature use demand and supply side for identification. I end this section by briefly discussing an important paper by [Perrigne and Vuong, 2011a], who consider the same market as I do. The main difference is that they focus only on Verizon and treat the market as being served by a monopoly publisher. Ignoring Ogden they can model consumer (unobserved) type by one dimensional parameter and allow more general base utility function than I can. They show how to identify the density of consumer type, which is similar to the identification of the marginal density in this paper, except that I am interested in joint density. Their cost identification is similar to my identification of marginal costs. In that respect these papers complement each other because together they illustrate how each assumption will affect the inference we can draw from the data.

The paper is organized as follows: Section 2 describes the data, the model is presented in Section 3 while identification and estimation is studied in Section 4. Estimation results are collected in Section 4.3 and Section 5 concludes. All the omitted proofs are collected in the appendix.

## 2. NONLINEAR PRICING IN YELLOW PAGES

I consider a market for ad in Yellow Pages directories placed by local business units in central Pennsylvania (State College and Bellefonte) for the year 2006. The market is served by two publishers, Verizon (henceforth, VZ), a utility provider, and Ogden (henceforth, OG), offer consumers a menu of options to choose from. Any phone number registered for business is treated as a consumer

and is recorded in the data. There are two parts to the data: the price schedules offered by the two publishers and the ad placed by each consumer with them. As a member of the Yellow Page Association, an umbrella organization of Yellow Page publishers, their price schedules are recorded with the association. I got the price information from the association. The association also ensures (as far as possible) that the publishers report the price data truthfully. Since I will use the supply side information, this is an important assumption for identification. The information on the ad placed by each business units were collected manually from the two directories. So for each business units, I know its business type (restaurant, dentist, etc) – the type of ad placed with each directory and the total outlay to each directory. It is a norm in the market to put names and addresses of each business-unit in the directory, so I know the names of all business-units even if they chose not to buy more ad.

As far as the market structure is concerned, VZ entered the market earlier than OG and is perceived to be the dominant player. Historically, the list of phone numbers was treated as proprietary asset of the phone provider, here VZ, until it was overturned, which subsequently allowed entry by other firms. In my case this led to the entry of OG. VZ's directory is slightly bigger and thicker, with three columns per page, with better quality paper than OG's directory which has only two columns. VZ distributes more than 215,400 copies while Ogden distributes only 73,000, but they cover the same market. Since it is not clear if a sequential (Stackelberg) or simultaneous duopoly is a better fit, I will use the optimality conditions that are valid under both.<sup>6</sup>

Both publishers offer different ad options that can be classified into three general categories: (i) listing; (ii) space listing; and (iii) display. VZ and OG both offer variations within each category. For example, VZ offers three font sizes to just list the names, address and phone number(s) of the business-unit. OG offers listing with only two font sizes. Listing with smallest font (standard listing) was offered for free by both to every one whose phone was registered as a business phone.<sup>7</sup> Listings account for 30% (resp. 53%) of the total ad sales in VZ (resp. OG). Space listing refers to an option where a space is allocated within the column under an appropriate business heading (such as Doctors, Salons, etcetera), and both VZ and OG offer five different variations and it accounts for 30% (resp. 26%) of the total ad sales in VZ (resp. OG). Finally, display refers to an option where a space (that could cover up to two pages) with colorful pictures is allocated for the buyer. VZ offers nine different variations and OG offers seven different variations within this category, which is also the most expensive of the three options. One can also choose different colors and sizes. VZ offers five color options – no color, one color, white background, white background plus one color and multiple colors including photos and OG offers four – no color, one color, white background plus one color and multiple colors including photos, Table A-1.

<sup>6</sup> I implicitly assume that if the game was Stackelberg VZ would be the leader.

<sup>7</sup> Thus, standard listing provides the exhaustive list of the names of all who can buy an ad.

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.

The unit of measurement is pica, which is approximately 1/6 of an inch. A standard listing in VZ is equal to 12 sq. pica, in OG is 9 sq. pica, and a full page in VZ is equal to 3,020 sq. pica, in OG is 1,845 sq. pica. From Table A-1 one can 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 (reap. \$6,324 in OG), which increase to \$32,395 (resp. \$9,435 in OG) with multiple colors; (ii) VZ's price is significantly higher than Ogden's across all the comparable advertising options, ( a half-page display without color costs \$10,093 in VZ and only \$3,372 in OG); (iii) the price difference between VZ and OG is smaller for the lower-end options, such as listing, than for the upper-end options, e.g., VZ's average price is 130% higher than OG's for the display option and this difference decreases drastically to 18% for space listing and to 17% for standard listing (no color); and (iv) for a given color category, both offer quantity discount: the price per sq. pica decreases with the ad size, e.g., the unit price per square pica for a double-page, a full-page, and half-page display ad with no color are \$5.68 (resp. \$3.43) \$6.13 (resp. \$3.68) and \$6.90 (resp. \$3.72), respectively for VZ (resp. OG); see Table 1. It is important for the empirical analysis that these "sticker" prices be the "real" prices paid by each consumer. These prices were provided by the Yellow Pages association, according to whom all members provide the actual prices that were charged to the advertisers. In the environment that I consider, where consumers can buy from both if they want, there is no reason why a publisher would announce prices that are different from the actual prices. However, there is no way for me to verify this claim and any empirical findings of this paper is contingent on this assumption.<sup>8</sup>

Similarly, from Table A-2 one can see that: (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 the business units choose listing and 14% choose display in VZ, while 94% and 3.8% choose listing and display in OG; (iii) around 54% of the business units advertise exclusively with VZ, whereas only 2% advertise exclusively with OG, and 12% advertise with both and the remaining choose the standard listing by default. The average prices paid in each directory by the firms purchasing from both

<sup>8</sup> One alternative could be to introduce measurement error in prices. This, however, requires some extra information that can shed light on signal to noise ratio. Such a data is not available— recall that there is no information even on consumer heterogeneity. Therefore any assumption I make about the measurement error would be not be based on the data and ad hoc.

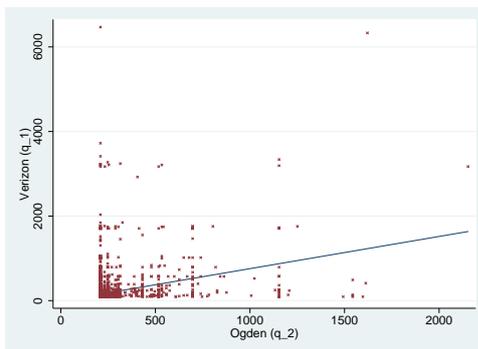


FIGURE 1. X-axis: Ogden ad; Y-axis: Verizon ad., in square pica

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.

One of the important implications that can be drawn from the choice of ad is that, within the quasi-linear preference environment, the (unobserved) preference heterogeneity of business units must be at least two dimensional. In other words we need marginal utilities to vary across all business units. This is because, if the utility is additively separable in price (quasi-linear) and the unobserved type was only one dimensional then that would mean both sellers have identical ranking of the consumers. In other words, a business unit is valued high by VZ if and only if it is valued high by OG, and as a result, the observed sales would coalesce around an increasing straight line and the ad sizes bought from VZ and OG would be highly correlated. This, however, is not what is observed in the data (see Figure 1), where the correlation between two ad is quite low at 0.25 increasing up to only 0.32 for the subset who buys from both. The correlation between the two ad 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. Together this motivates why I use unobserved types with at two dimensions, which leads to a model of competitive nonlinear pricing with multidimensional screening. There are other ways to rationalize the data that does not rely on multidimensional screening. For instance, one can construct a 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.). A optimizing consumer would then place ad in both 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, and moreover there is no information about the demographics, which makes modeling asymmetric stochastic matching technologies infeasible. Moreover one of the main objectives of this paper is to show how we can

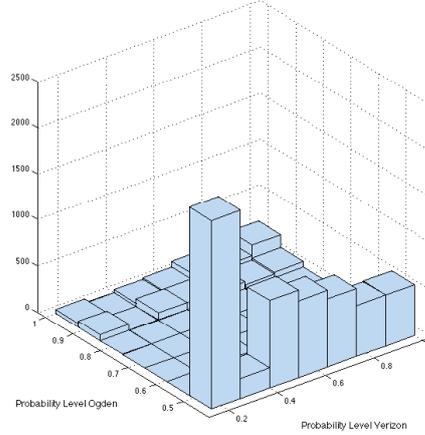


FIGURE 2. Rank plot of ad bought from Verizon and Ogden.

use principal-agent model to study markets when we do not have information on the consumers (besides sales) but we are still interested in unobserved consumer heterogeneity, hence my decision to use a multidimensional screening model with competition.

### 3. THE MODEL

In this section I present a model of competing nonlinear pricing that is based on [Ivaldi and Martimort \[1994\]](#). Let  $P1(VZ)$  and  $P2(OG)$  be the two sellers/principals. I will characterize the optimal nonlinear pricing, which consists a menu of quantities and prices, one for each seller. To solve the optimal price schedule for  $P1$ , I need to take a stand of whether the sellers move sequentially (Stackelberg) or simultaneously, but as discussed earlier it is not entirely clear what is the correct assumption. Therefore, I have decided to model only those supply side features that do depend on this assumption. It is known at least since [Borenstein and Rose \[1994\]](#); [Busse and Rysman \[2005\]](#), that that competition affects the equilibrium only through the price schedule and not through quantities (also known as allocation rules in mechanism design parlance). Hence, the optimal quantity functions are invariant to the timing assumption. Moreover, if the sellers moved sequentially then  $P2$  would be the second mover, which means I can only solve  $P2$ 's optimal price schedule (as a reaction function), but cannot solve  $P1$ 's optimal price schedule.

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

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

where  $T_i(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is the pricing function chosen by  $P_i$ . 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  $q_i$  defined as  $MU_i = \theta_i - b q_i + c q_{-i} - T'(q_i)$  increases with type  $\theta_i$  and increases or decreases with  $q_{-i}$  depending on whether the two are complementary, i.e.  $c > 0$  or substitutes, i.e.  $c \leq 0$ . If, however,  $c > 0$  then  $q_1$  and  $q_2$  should be positively-assortative given the concavity of the gross utility function. However, that is incompatible with the data ( Figures 1 - 2), so I restrict  $c \leq 0$ .

Equation (1) says that consumers have the same quasi-linear utility function but they differ in terms of their marginal utility for the two products. Similar functional form has also been used in hedonic models by [Ekeland, Heckman, and Nesheim, 2002, 2004]. The quasi-linearity assumption is simplifying assumption used in almost all mechanism design literature and discrete choice literature.<sup>9</sup> There are two advantages of using quadratic gross utility  $u(\mathbf{q}, \theta, A)$ . First, concavity of the utility function generates diversification (also known as “lover for variety” in international trade literature) where by there is an advantage in buying ads in both directory. Second, the marginal utility is linear in consumer types, which keeps the multidimensional screening model tractable. (Note that this framework is flexible enough to allow for any observed consumer heterogeneity as long as the asymmetric information is only with respect to  $(\theta_1, \theta_2)$ .)

The type  $(\theta_1, \theta_2)$  are private information of each consumer but the publishers commonly know that they are distributed as  $F(\cdot, \cdot)$ . They also know each other’s printing cost.

**Assumption 1.** *I make the following assumptions:*

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

Like in empirical auction, where each individual bidder is characterized by valuation  $v$  that summarizes the maximum willingness to pay, here it is assumed that all relevant heterogeneity is captured in  $(\theta_1, \theta_2)$ , which summarizes the value for placing an ad. In that sense, I do not model the determinants of demand for ad, which is an important topic of research left as a future project. And the

<sup>9</sup> In an automobile insurance market preferences are not quasi-linear. [Aryal and Perrigne, 2011] consider a model where consumers have private information about their risk (average number of accidents) and risk aversion, and [Aryal, Perrigne, and Vuong, 2010] show how to identify the joint density of consumer types nonparametrically.

facts that both publishers offer the same menu to all businesses and there is no consumer level information, there isn't much that I can do.<sup>10</sup> The functional form assumption for cost motivated by the observation in printing industry where the two main cost components are the fixed cost of printing machine and distribution and a fixed marginal cost of ink, paper and labor.

There are, however, two difficulties in finding the equilibrium nonlinear prices. First, without any restriction on space of pricing functions the model can be intractable. Second, the ad options offered in the data vary in both size and color, while the theoretical model treats either quantity or quality as a continuous one-dimensional variable. It turns out that the sellers do not use color to discriminate across buyers, which suggests that it is possible to combine size and color into a continuous one-dimensional variable: "quality-adjusted quantity." In the following subsection I explain how this can be achieved. In the process I also show that a function that is quadratic (concave) in quality-adjusted quantity approximates the raw price data very well, which means that if the method is plausible then while solving equilibrium prices we can restrict our attention to quadratic form (see Equation (2)).<sup>11</sup>

**3.1. Quality Adjusted Quantity.** To verify that such a transformation preserves the ranking from the perspective of the consumers, I use the following features. First, note that if both size and color were important then we should see the sellers using both dimensions to discriminate consumers, but once I control for the size the relative price is constant across colors. In particular, discounts are offered for large ad 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. Second, according to [Maskin and Riley, 1984], 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, which I call quality-adjusted quantity.

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

---

<sup>10</sup> If the distinctions, say, between restaurants and salons were important we expect the publishers to offer two separate menus of choices.

<sup>11</sup> In the supplementary note [Aryal, 2013] I derive a condition that is necessary and sufficient for the optimal pricing to be quadratic.

	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

using OLS:

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

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 (non-multicolored choices) 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:** This procedure is different from the approximation problem considered by Wilson [1993]. While he is interested in reasons why sellers might choose simpler pricing functions, such as two-part tariff, over continuous nonlinear pricing and finds that the former approximates the latter (in terms of profit) well, I am interested in narrowing the space of pricing functions that will make the problem tractable and at the same time preserve the fact that sellers don't use ad color to discriminate consumers.

**3.2. 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. As mentioned earlier, pricing functions are roughly quadratic, so I restrict  $T_1(\cdot)$  to be

$$T_1(q_1) = \begin{cases} \gamma_1 + \alpha_1 q_1 + \frac{\beta_1}{2} q_1^2 & \text{if } q_1 > q_{i0} \\ 0 & \text{if } q_1 \leq q_{i0}, \end{cases}\tag{3}$$

which is characterized by parameters  $\{\gamma_1, \alpha_1, \beta_1 : \gamma_1 > 0, \alpha_1 > 0, \beta_1 < 0\}$ . I assume that  $T_i(\cdot)$  is right differentiable at  $q_{i0}$  for  $i = 1, 2$ . In order to determine the participation/individual rationality (IR)

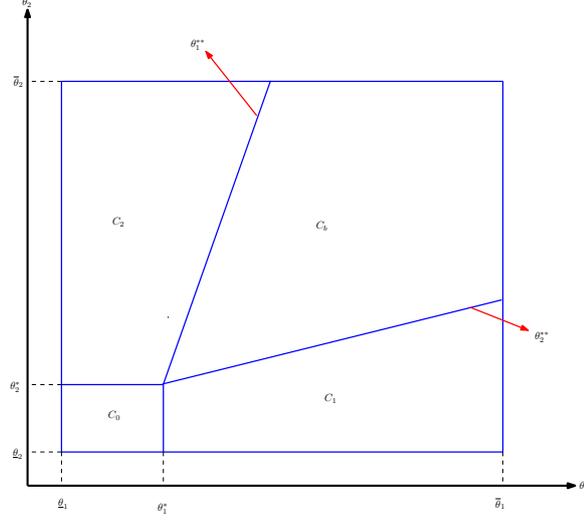


FIGURE 3. Partition of Consumer types:  $C_0$  types choose  $(q_{10}, q_{20})$ ;  $C_i$  types choose  $(q_i > q_{i0})$  and  $C_b$  types choose  $(q_i > q_{i0})$  for both  $i \in \{1, 2\}$ .

and incentive compatibility constraint (IC), I will use the consumer's first order conditions

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

to determine four types of consumers: those who do not participate and choose  $(q_{10}, q_{20})$  denoted as  $C_0$ , those who choose only from either  $P_1$  or  $P_2$  and are denoted, respectively as  $C_1$  and  $C_2$  and those who buy both  $q_1 > q_{10}$  and  $q_2 > q_{20}$ , denoted as  $C_b$ ; all in Figure 3. These four subsets are determined endogenously given  $T_1(\cdot)$  and  $T_2(\cdot)$ .

Consider the set  $C_0$ , where the types do not participate. Then for all  $(\theta_1, \theta_2)$  the net marginal utility  $MU_i(\cdot, \cdot; \theta_1, \theta_2) \leq 0$  when evaluated at the pair  $(q_{10}, q_{20})$  for  $i = 1, 2$ .  $MU_i(q_{10}, q_{20}; \theta) \leq 0$  for  $i = 1, 2, \theta \in C_0$ . 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  $(\theta_1^*, \theta_2^*)$  the marginal type who choose  $(q_{10}, q_{20})$ , i.e.

$$\theta_1^* = \alpha_1 + (b_1 + \beta_1) q_{10} - c q_{20} \quad (5)$$

$$\theta_2^* = T_2'(q_{20}) + b_2 q_{20} - c q_{10}. \quad (6)$$

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

$$\theta_2^{**} = \left( b_2 - \frac{c^2}{b_1 + \beta_1} \right) q_{20} + T_2'(q_{20}) + \frac{c \alpha_1}{b_1 + \beta_1} - \frac{c}{b_1 + \beta_1} \theta_1. \quad (7)$$

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

$$\theta_1^{**} = \left( b_1 + \beta_1 - \frac{c^2}{b_2} \right) q_{10} + \alpha_1 + \frac{c}{b_2} T_2'(q_2) - \frac{c}{b_2} \theta_2. \quad (8)$$

$C_b$  is determined by the two first-order conditions as in Equation (4) that can be simplified as

$$\theta_1 - b_1 q_1 + c q_2 = \alpha_1 + \beta_1 q_1, \quad (9)$$

$$\theta_2 - b_2 q_2 + c q_1 = T_2'(q_2), \quad (10)$$

In this subsection I solve for optimal nonlinear pricing (best response) for the follower  $P2$  after observing  $T_1(\cdot)$  given in (3).<sup>12</sup> For those consumers who buy  $q_2 > q_{20}$ , the corresponding  $q_1$  can be determined from (9) as

$$q_1 = \begin{cases} \frac{\theta_1 - \alpha_1 + c q_2}{b_1 + \beta_1}, & \theta_1 > \theta_1^* \\ q_{10}, & \theta_1 \leq \theta_1^* \end{cases} \quad (11)$$

Substituting (11) in (10) gives the necessary condition for optimal  $q_2$  to be optimal for type  $(\theta_1, \theta_2)$  consumer, i.e.

$$\theta_2 + \frac{c \theta_1}{b_1 + \beta_1} = \frac{c \alpha_1}{b_1 + \beta_1} + \left( b_2 - \frac{c^2}{b_1 + \beta_1} \right) q_2 + T_2'(q_2). \quad (12)$$

Notice that the unobserved types appear only in the LHS of (12) and given  $T_1(\cdot)$  they can be treated as exogenous from the point of view of  $P2$ . This suggests that the LHS can be treated as an aggregated one-dimensional type (sufficient statistic). Let  $z_2 = \theta_2 + \frac{c \theta_1}{b_1 + \beta_1}$  be such a new type, then a type  $(\theta_1, \theta_2)$

<sup>12</sup> An alternative way to solve the problem is to use [Rochet and Choné, 1998], which is explored in the supplement [Aryal, 2013].

consumer who is also now a type  $z_2$  chooses an optimal  $q_2$  that solves

$$z_2 = \frac{c\alpha_1}{b_1 + \beta_1} + \left( b_2 - \frac{c^2}{b_1 + \beta_1} \right) q_2 + T_2'(q_2). \quad (13)$$

For  $P2$ ,  $z_2$  is unobserved and is exogenously determined, and hence can be treated as the (unobserved) preference of a firm for  $q_2$ . It aggregates  $(\theta_1, \theta_2)$  in the sense that it captures the taste for  $q_2$ : it is increasing in  $\theta_2$  and decreasing in  $\theta_1$  (those who value  $q_1$  more like  $q_2$  less) and decreases with  $\beta_1$  ( $q_2$  is weakly increasing price for VZ, ceteris paribus). Therefore,  $z_2$  aggregates  $(\theta_1, \theta_2)$  and acts as sufficient statistic in the sense that a mechanism that depends on  $z_2$  will do as good as a mechanism that depends on  $(\theta_1, \theta_2)$ . This also means that there will be pooling at equilibrium, i.e. two different types with same  $z_2$  will be allocated the same good, but pooling is inevitable because the seller has only one dimensional instrument  $q_2 \in \mathbb{R}$ , while consumers have two dimensional types. The only important question is how should the types  $\theta_1$  and  $\theta_2$  be pooled, and argument above shows that  $z_2$  is one such way.

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

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

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

$$\max_{T_2(\cdot), q_2(\cdot), z_2^0} \left\{ \mathbb{E}\Pi_2 = \int_{z_2^0}^{\bar{z}_2} \left( T_2(q_2(z_2)) - m_2 q_2(z_2) \right) g_2(z_2) dz_2 - K_2 - m_2 q_{20} G_2(z_2^0) \right\}, \quad (14)$$

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 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  $\theta_1 \geq \theta_1^*$ . To determine the IC constraint note that  $z_2$ 's choice of  $q_2$  depends on her choice of  $q_1$ , (9) & (10), which in turn depends on  $q_2$ , and so on, a difficult task in general. However, the structure of our problem can be used to simplify the solution. In particular, I can write  $z_2$ 's net utility from  $(q_1, q_2)$  (denoted by  $W_2(\theta_1, z_2)$ ) as a sum of the net utility that  $z_2$  gets from  $(q_1, q_{20})$  (denoted by  $w_2(\theta_1, z_2)$ ) and any additional utility from choosing  $q_2 > q_{20}$  (denoted by  $s_2(q_2, z_2)$ ), such that  $s_2(q_{20}, z_2) = 0$ . To wit, note that using the definitions

$$\begin{aligned} w_2(\theta_1, z_2) &:= \max_{q_1 \geq q_{10}} \left[ u\left(q_1, q_{20}; \theta_1, z_2 - \frac{c\theta_1}{b_1 + \beta_1}\right) - T_1(q_1) \right]; \\ s_2(q_2, z_2) &:= \max_{q_2 \geq q_{20}} \left\{ \left( z_2 - \frac{c\alpha_1 - c^2 q_2}{b_1 + \beta_1} \right) (q_2 - q_{20}) - \frac{b_2}{2} (q_2^2 - q_{20}^2) - T_2(q_2) \right\}, \end{aligned}$$

the net utility can be written as

$$\begin{aligned} W_2(\theta_1, z_2) &:= \max_{q_1 \geq q_{10}, q_2 \geq q_{20}} \left[ u \left( q_1, q_2; \theta_1, z_2 - \frac{c\theta_1}{b_1 + \beta_1} \right) - T_1(q_1) - T_2(q_2) \right] \\ &= w_2(\theta_1, z_2) + s_2(q_2, z_2). \end{aligned}$$

Then the IC constraint becomes  $s_2(q_2(z_2); z_2) \geq s_2(q_2(\bar{z}_2); z_2)$  for all  $z_2, \bar{z}_2 \in [z_2^0, \bar{z}_2]$ . Moreover,  $s_2(\cdot)$  is continuous, convex and satisfies the envelope conditions

$$s_2'(z_2) = q_2(z_2) - q_{20} \quad \forall z_2 \in (z_2^0, \bar{z}_2], \quad (15)$$

and

$$T(z_2) = \left( z_2 - \frac{c\alpha_1 - c^2 q_2}{b_1 + \beta_1} \right) (q_2 - q_{20}) - \frac{b_2}{2} (q_2^2 - q_{20}^2) - s_2(z_2). \quad (16)$$

From (15) and (16),  $P2'$ s can be viewed as choosing  $s_2(z_2)$ , the rent function. In an influential paper [Rochet \[1987\]](#) showed:

**Lemma 1.** *The global IC constraint is satisfied if, and only if, (a)  $s_2(z_2) = \int_{z_2^0}^{z_2} (q_2(t) - q_{20}) dt + s_2^+$ ,  $\forall z_2 \in [z_2^0, \bar{z}_2]$ , where  $s_2^+ \equiv \lim_{z_2 \downarrow z_2^0} s_2(z_2)$ ; and (b)  $s_2(\cdot)$  is convex or equivalently  $q_2'(z_2) > 0$ .*

From (15) it follows that the global IC is satisfied if and only if  $q_2(\cdot)$  is strictly increasing in  $z_2$ . The participation (IR) constraint becomes  $W_2(\theta_1, z_2) = w_2(\theta_1, z_2) + s_2(z_2) \geq \max\{w_2(\theta_1, z_2), 0\}$ , which is equivalent to  $s_2(z_2) \geq 0$ . Then,  $P2'$ s optimization becomes

$$\begin{aligned} \max_{q_2(\cdot), z_2^0, s_2^+} \mathbb{E}\Pi_2 &= \int_{z_2^0}^{\bar{z}_2} \left\{ \left( z_2 - \frac{c\alpha_1 - c^2 q_2(z_2)}{\beta_1 + b_1} \right) (q_2(z_2) - q_{20}) - \frac{b_2}{2} (q_2^2(z_2) - q_{20}^2) - m_2 q_2(z_2) \right. \\ &\quad \left. - s_2^+ - (q_2(z_2) - q_{20}) \frac{1 - G_2(z_2)}{g_2(z_2)} \right\} g_2(z_2) dz_2 - K_2 - m_2 q_{20} G_2(z_2^0), \end{aligned}$$

subject to the  $q_2'(z_2) > 0$  (IC) and  $s_2(z_2) \geq 0$  (IR) for all  $z_2 \in [z_2, \bar{z}_2]$ . To solve the above problem, I consider the relaxed problem where the constraints are verified ex post. Since  $s_2(\cdot)$  is convex  $s_2(z_2^0) = 0$  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. Then the existence and uniqueness of the solution can be guaranteed:

**Theorem 1.** *Under the maintained assumptions on preferences and cost and the full support assumption of  $g_2(\cdot)$  there exists a unique solution to the problem (14).*

The proof is based on the result by [Rochet and Choné \[1998\]](#) is given in the supplementary material [[Aryal, 2013](#)]. The solution to this problem is formalized below:

**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

$$q_2(z_2) = \frac{z_2 - \frac{1-G_2(z_2)}{g_2(z_2)} - m_2 - \frac{c^2 q_{20} + c\alpha_1}{b_1 + \beta_1}}{b_2 - \frac{2c^2}{b_1 + \beta_1}}, \forall z_2 \in (z_2^0, \bar{z}_2] \quad (17)$$

such that  $q_2(z_2) = q_{20}$  otherwise, and  $z_2^0$  solves

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

(2)  $T_2(q)$  must satisfy (16), it must also satisfy the Ramsey rule:

$$\frac{T_2'(q_2(z_2)) - m_2}{T_2(q_2(z_2))} = \frac{1 - G_2(z_2)}{g_2(z_2)} \frac{1}{\frac{\partial s_2(q_2(z_2))}{\partial q_2}}. \quad (18)$$

*Proof.* The proof is standard in the literature, for instance see [Stole, 2007]. Only the main steps are highlighted here. The first step determines that the expected profit function is concave in  $q_2$  and super modular in  $(q_2, z_2)$ . Let  $I$  be the integrand of the expected profit function:

$$\begin{aligned} \frac{\partial I}{\partial q_2} &= \left( \left( z_2 - \frac{c\alpha_1 - c^2 q_2}{\beta_1 + b_1} \right) + \frac{c^2}{b_1 + \beta_1} (q_2 - q_{20}) - b_2 q_2 - \frac{1 - G_2(z_2)}{g_2(z_2)} - m_2 \right) g(z_2), \\ \frac{\partial^2 I}{\partial q_2^2} &= - \left( b_2 - \frac{2c^2}{b_1 + \beta_1} \right) g_2(z_2), \\ \frac{\partial^2 I}{\partial q_2 \partial z_2} &= \left( \left( 1 - \frac{\partial}{\partial z_2} \frac{1 - G_2(z_2)}{g_2(z_2)} \right) - \left( b_2 - \frac{2c^2}{b_1 + \beta_1} \right) q_2'(\cdot) \right) g(z_2) = 0. \end{aligned}$$

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  $I$ :

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

which gives

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

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

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

And since  $q_2(z_2^0) = q_{20}$ ,  $z_2^0$  solves

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

Since  $T_2'(q_2) = z_2 - \frac{c\alpha_1}{b_1 + \beta_1} - (b_2 - \frac{2c^2}{b_1 + \beta_1})q_2 + \frac{c^2}{b_1 + \beta_1}q_{20}$ , which follows from differentiating (16), the Ramsey equation (18) follows immediately.  $\square$

Equation (18) 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.

Now, I characterize the (optimal) nonlinear pricing schedule for  $P1$  given the best response (reaction) function of  $P2$ . As mentioned earlier, I will only solve optimal quantity function  $q_1(\cdot)$ . Firms with type  $\theta_1 \geq \theta_1^{**}$  choose  $q_1 > q_{10}$  while those with  $\theta_1 \leq \theta_1^{**}$  choose  $q_{10}$ . Using (11) and  $T_2(\cdot)$ , the allocation rule  $q_2$  is a function of  $q_1$ :

$$q_2(q_1; \theta_1, \theta_2) = \begin{cases} \frac{\theta_2 - \alpha_2 + cq_1}{b_2 + \beta_2}, & \theta_2 > \theta_2^{**} \\ q_{20}, & \theta_2 \leq \theta_2^{**} \end{cases} \quad (19)$$

Following the same arguments as with  $P2$ , let  $z_1 = \theta_1 + \frac{c\theta_2}{b_2 + \beta_2}$  be the new aggregator such that the optimal  $q_1$  solves

$$z_1 = b_1q_1 + c \left[ \frac{\alpha_2 - cq_1}{b_2 + \beta_2} \right] q_1 + \alpha_1 + \beta_1q_1. \quad (20)$$

and the threshold type

$$z_1^0 = \begin{cases} \theta_1^* + \frac{c\theta_2^*}{b_2 + \beta_2} & \theta_2 \leq \theta_2^* \\ \theta_1^{**} + \frac{c\theta_2}{b_2 + \beta_2} & \theta_2 \geq \theta_2^* \end{cases}$$

such that any type  $z_1 \leq z_1^0$  buys  $q_{10}$ . Similarly, let  $W_1(z_1, \theta_2)$ ,  $w_1(\cdot, \cdot)$  and  $s_1(q_1, z_1)$  be  $P1$ 's counterpart such that  $W_1(z_1, \theta_2) = w_2(z_1, \theta_2) + s_1(q_1, z_1)$ , where

$$s_1(q_1, z_1) := \max_{\{q_1 \geq q_{10}\}} \left( z_1 + \frac{c^2q_1 - c\alpha_2}{b_2 + \beta_2} \right) (q_1 - q_{10}) - \frac{b_1}{2}(q_1^2 - q_{10}^2) - \gamma_1 - \alpha_1q_1 - \frac{\beta_1}{2}q_1^2.$$

The function  $s_1(q_1(z_1), z_1) \equiv s_1(z_1)$  is the relevant rent function for  $P1$  from which I get

$$T_1(q_1) = \left( z_1 + \frac{c^2q_1 - c\alpha_2}{b_2 + \beta_2} \right) (q_1 - q_{10}) - \frac{b_1}{2}(q_1^2 - q_{10}^2) - \int_{z_1^0}^{z_1} (q_1(t) - q_{10}) dt - s_1^+ \quad (21)$$

that allows us to re-write  $P1$ 's problem as

$$\begin{aligned} \max_{q_1(\cdot), z_1^0, s_1^+} \mathbb{E}\Pi_1 = & \int_{z_1^0}^{\bar{z}_1} \left[ \left( z_1 + \frac{c^2 q_1 - c \left( \zeta_2 + l_2 m_2 + \frac{c \alpha_1 (l_2 - 1)}{b_1 + \beta_1} \right)}{b_2 l_2 - \frac{c^2 (l_2 - 1)}{b_1 + \beta_1}} \right) (q_1 - q_{10}) \right. \\ & \left. - \frac{b_1}{2} (q_1^2 - q_{10}^2) - (q_1 - q_{10}) \tilde{G}_1(z_1) - K_1 - m_1 q_1 \right] g_1(z_1) dz_1 - m_1 G_1(z_1^0) q_{10}, \end{aligned}$$

subject to  $q_1'(\cdot) > 0$  (IC) and  $s_1(\cdot) \geq 0$  (IR) constraints for all  $z_1 \in [z_1^0, \bar{z}_1]$ .

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

$$q_1(z_1) = \begin{cases} \frac{z_1 - \tilde{G}_1(z_1) - m_1 - \frac{c \alpha_2 + c^2 q_{10}}{b_2 + \beta_2}}{b_1 - \frac{2c^2}{b_2 + \beta_2}}, & \forall z_1 \in (z_1^0, \bar{z}_1] \\ q_{10}, & \forall z_1 \in [z_1^0, z_1^0] \end{cases} \quad (22)$$

The optimal quantity is determined for a particular price schedule, for which the proof is straightforward and is omitted.

#### 4. IDENTIFICATION AND ESTIMATION

**4.1. Identification.** In this section I study the identification problem of the model, which concerns the possibility of drawing inferences about the model structure outlined above from the observed data on ads bought and the two price schedules. Failure to identify the model structure implies that the data lacks sufficient information to distinguish between alternative structures. The model primitives are the joint distribution of types  $F(\cdot, \cdot)$  and the set of utility and cost parameters  $X = [m_1, m_2, K_1, K_2, b_1, b_2, c]$ . The data provides information on the price functions  $\{\alpha_i, \beta_i, \gamma_i : i = 1, 2\}$  and the ad bought by  $J$  business-units  $\{q_{1j}, q_{2j}\}_{j=1}^J$ . A structure is a set of hypothesis that implies a unique distribution consistent with the data. Two structures  $\{F(\cdot, \cdot), X\} \neq \{\tilde{F}(\cdot, \cdot), \tilde{X}\}$  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. Given our environment, I assume that the joint distribution  $F(\cdot, \cdot)$  is defined on  $\Theta := [\underline{\theta}_1, \bar{\theta}_1] \times [\underline{\theta}_2, \bar{\theta}_2]$  such that  $F(\cdot, \cdot)$  is absolutely continuous with continuously differentiable and nowhere vanishing density  $f(\cdot, \cdot)$  and  $X$  is such that (i)  $b_1 b_2 - c^2 > 0$ ; (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 concave utility and convex optimization problem.

For a fixed parameter set  $X$ , consider the demand system

$$\mathbf{q}(\theta; X) = (q_1(\theta_1, \theta_2; X), q_2(\theta_1, \theta_2; X)) : \Theta \rightarrow \mathbb{R}_+^2,$$

which can be re-written as a function of  $z$ 's, i.e.

$$\mathbf{q}(z; X) = (q_1(z_1; X), q_2(z_2; X)).$$

The objective is to show that the system is invertible in  $(z_1, z_2)$ , and since  $(z_1, z_2) \mapsto (\theta_1, \theta_2)$  is one-to-one, we can identify  $(\theta_1, \theta_2)$  and hence  $F(\cdot, \cdot)$ . However, this is not possible because only few choose  $q_1 > q_{10}$  and  $q_2 > q_{20}$ . Therefore, I focus only on  $[z_i^0, \bar{z}_i]$  where  $q_i \geq q_{i0}$ , and use the fact that  $q_i(z_i)$  is monotonic (because of incentive compatibility), pseudo  $\hat{z}_i$  can be identified along with its conditional marginal density, for  $i = 1, 2$ . Under some intuitive normalization of utility from outside option, I show the parameters  $X$  can be identified. This means one cannot identify the joint density of  $\theta$  from the two nonparametric (conditional) densities of  $\theta_1$  and  $\theta_2$ . Since  $(\theta_1, \theta_2) \in C_b$  choose  $q_i > q_{i0}$ , the correlation between  $q_1$  and  $q_2$  will be informative about the correlation between  $\theta_1$  and  $\theta_2$ . So, I use (empirical) Copula with unspecified marginal distributions to determine the joint density of  $(\theta_1, \theta_2)$ . To choose the right copula I use both goodness-of-fit and non-nested model selection tests.<sup>13</sup>

The optimal nonlinear pricing functions do not depend on the fixed costs  $K_1$  and  $K_2$ , so they cannot be identified. Incentive compatibility implies that the equilibrium allocation rule  $q_i(\cdot) : [z_i^0, \bar{z}_i] \mapsto [q_{i0}, \bar{q}_i]$  is monotonic, hence can be inverted to provide a (inverse) mapping  $z_i(\cdot) \equiv q_i^{-1}(\cdot)$ , from sales (data) to type. Then  $z_{ij} = z_i(q_{ij})$  is the consumer  $j$ 's type who chose  $q_{ij} > q_{i0}$  from  $P_i$ . Let  $H(\cdot, \cdot)$  be the conditional joint distribution of  $(q_1, q_2)$  given  $q_i > q_{i0}$  and let

$$H_i(q_i) = \int H(q_i, q_j) dq_j := \Pr[q_i \leq q | q_i > q_{i0}] = \Pr[z_i \leq z_i(q) | z_i > z_i(q_{i0})],$$

for  $i = 1, 2, j \in \{1, 2\}, j \neq i$ , be the corresponding marginals. From the data we can identify (estimate) both the joint  $H(\cdot, \cdot)$  and the two marginals  $H_i(q_i), i = 1, 2$ . Since  $H_i(q) = \frac{G_i(z_i) - G_i(z_i^0)}{1 - G_i(z_i^0)}$  and therefore  $h_i(q) = \frac{\partial H_i(q)}{\partial q_i} = \frac{g_i(z_i) z_i'(q)}{(1 - G_i(z_i^0))}$ , we get

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

**Identification of Cost Parameters.** To identify the marginal cost, I use the fact that there is no distortion on the top. In other words, the highest type gets the quantity that maximizes the social welfare. Consider  $P_2$ : the data identifies  $\bar{q}_2 = \max\{q_{2i}; i = 1, \dots, J\}$ , but monotonicity implies that the type that buys  $\bar{q}_2$  is  $\bar{z}_2$ , i.e.,  $q_2(\bar{z}_2) = \bar{q}_2$ . Then the pricing function gives  $T_2'(q_2(\bar{z}_2)) = T_2'(\bar{q}_2) = \alpha_2 + \beta_2 \bar{q}_2$ . Substituting this in the Ramsey rule (18) gives  $\alpha_2 + \beta_2 \bar{q}_2 - m_2 = 0$ , which identifies  $m_2$ . Because of the sequentiality of the game, the same argument cannot be applied to identify  $m_1$ . Differentiating  $T_1(\cdot)$  in Equation (21) with respect to  $q_1$  and solving for  $z_1'(q_1)$  gives  $z_1'(q_1) = T_1''(q_1) + b_1 - \frac{2c^2}{b_1 + \beta_1}$ ,

<sup>13</sup>As is clear, one can allow for observed consumer heterogeneity  $Z$  to estimate  $\hat{f}(\theta_1, \theta_2 | Z)$ .

which can be used in the optimal allocation rule

$$T'_i(q_i) = m_i + \frac{1 - H_i(q_i)}{h_i(q_i)} z'_i(q_i). \quad (23)$$

for  $q_1(\cdot)$  to get

$$T'_1(q_1) = m_1 + \frac{1 - H_1(q_1)}{h_1(q_1)} \left( T''_1(q_1) + b_1 - \frac{2c^2}{b_2 + \beta_2} \right), \quad \forall q_1 \in [q_{10}, \bar{q}_1].$$

When the last equation is evaluated at  $\bar{q}_1 := q_1(\bar{z}_1) = \max\{q_{1j}; j = 1, 2, \dots, J\}$  gives

$$m_1 + \underbrace{\frac{1 - H_1(\bar{q}_1)}{h_1(\bar{q}_1)}}_{=0} \left( \beta_1 + b_1 - \frac{2c^2}{b_2 + \beta_2} \right) = \alpha_1 + \beta_1 \bar{q}_1 \Rightarrow m_1 = \alpha_1 + \beta_1 \bar{q}_1. \quad (24)$$

**Identification of  $b_1, b_2, c$  and the support.** The utility function is concave and the parameters  $b_1$  and  $b_2$  induce “love for variety,” ceteris paribus. To see this consider the extreme case when the utility is 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  $q_2 > 0$ . Intuitively, as the utility 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. The  $\bar{z}_1$  – type’s optimality condition (marginal utility equals marginal price)

$$\bar{\theta}_1 - b_1 \bar{q}_1 + cq_{20} = \alpha_1 + \beta_1 \bar{q}_1$$

identifies  $c$  as a function of  $\bar{\theta}_1$  and  $b_1$ . Similarly, the optimality for the  $\bar{z}_2$  – type

$$\bar{\theta}_2 - b_2 \bar{q}_2 + cq_{10} = \alpha_2 + \beta_2 \bar{q}_2 \Rightarrow b_2 = \frac{cq_{10} + \bar{\theta}_2 - \alpha_2}{\bar{q}_2} - \beta_2$$

identifies  $b_2$  as a function of  $\bar{\theta}_2$ . Therefore,  $c$  and  $b_2$  are identified from  $\{\bar{\theta}_1, \bar{\theta}_2, b_1\}$ . For any  $q_i < \bar{q}_i$ , I rewrite (23) 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  $q_1 \neq \bar{q}_1$  gives

$$b_1 + \beta_1 = \frac{\alpha_1 + \beta_1 q_1 - m_1}{\frac{1 - H_1(q_1)}{h_1(q_1)}} + \frac{2c^2}{b_2 + \beta_2}, \quad \text{and} \quad b_1 + \beta_1 = \frac{\alpha_1 + \beta_1 \bar{q}_1 - m_1}{\frac{1 - H_1(\bar{q}_1)}{h_1(\bar{q}_1)}} + \frac{2c^2}{b_2 + \beta_2}.$$

Solving these two equations identifies  $b_1$  as

$$b_1 = \frac{1}{2} \left( \frac{\alpha_1 + \beta_1 q_1 - m_1}{\frac{1-H_1(q_1)}{h_1(q_1)}} + \frac{\alpha_1 + \beta_1 \tilde{q}_1 - m_1}{\frac{1-H_1(\tilde{q}_1)}{h_1(\tilde{q}_1)}} \right) - \beta_1. \quad (25)$$

Then evaluating  $q_2(z_2)$  in (17) at  $(\bar{z}_2)$  gives  $\bar{z}_2 = \bar{q}_2 \left( b_2 - \frac{2c^2}{b_1 + \beta_1} \right) + m_2 + \frac{c^2 q_{20} + c\alpha_1}{b_1 + \beta_1}$ , which together with the definition of  $\bar{z}_2$  identifies  $\bar{\theta}_2$  as

$$\bar{\theta}_2 = \bar{q}_2 b_2 + m_2 + \frac{c^2 q_{20} + c\alpha_1 - c\bar{\theta}_1 - 2c^2 \bar{q}_2}{b_1 + \beta_1}. \quad (26)$$

Since some consumers choose  $(q_{10}, q_{20})$ , normalization is important to determine the support. Although there are many possible normalizations that works, I assume that the type space is  $\Theta = [\underline{\theta}_1, \bar{\theta}_1] \times [0, \bar{\theta}_2]$  and normalize the utility from that the lowest types  $(\underline{\theta}_1, \underline{\theta}_2)$  get from the lowest quantities  $(q_{10}, q_{20})$  as zero.

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

Together the assumption determine  $\underline{\theta}_1$  as

$$\underline{\theta}_1 = \frac{b_1}{2} q_{10} + \frac{b_2}{2} \frac{q_{20}^2}{q_{10}} - c q_{20}. \quad (27)$$

**Identification of Marginal Densities.** Since implementability implies that the equilibrium allocation rules  $(q_1(\cdot), q_2(\cdot))$  are monotonic in  $z_1$  and  $z_2$ , respectively, these functions can be inverted to express conditional distribution of types as a function of observed demand distribution. That is using the ad placed with VZ (resp. OG) I can identify the conditional marginal distribution of  $z_1$  (resp.  $z_2$ ) given that  $z_1 \geq z_1^0$  (resp.  $z_2 \geq z_2^0$ ).

Recall that the relationship between type  $z_{ij}$  and ad  $q_{ij}$  is as follows  $\tilde{G}_i(z_{ij}) := \frac{G_i(z_i) - G_i(z_i^0)}{1 - G_i(z_i^0)} = \frac{1 - H_i(q_{ij})}{h_i(q_{ij})} m_i$ , which can then be used to recover  $(z_{1j}, z_{2j})$  from the consumption bundle  $(q_{1j}, q_{2j})$ . It is important to note, however, that the transformation is unique only for some subset. For instance, for  $C_b$  it must be the case that both types are greater than the threshold, so I can invert (11) and (17) to recover

$$\begin{pmatrix} z_{1j} \\ z_{2j} \end{pmatrix} = \begin{pmatrix} q_1^{-1}(q_{1j}) \\ q_2^{-1}(q_{2j}) \end{pmatrix} = \begin{pmatrix} q_{1j} \left( b_1 - \frac{2c^2}{b_2 + \beta_2} \right) + m_1 + \frac{c\alpha_2 + c^2 q_{10}}{b_2 + \beta_2} + \frac{(1 - H_1(q_{1j}))}{h_1(q_{1j})} m_1 \\ q_{2j} \left( b_2 - \frac{2c^2}{b_1 + \beta_1} \right) + m_2 + \frac{c\alpha_1 + c^2 q_{20}}{b_1 + \beta_1} + \frac{(1 - H_2(q_{2j}))}{h_2(q_{2j})} m_2 \end{pmatrix} \quad (28)$$

and hence the conditional joint distribution  $G(\cdot, \cdot | z_1 \geq z_1^0, z_2 \geq z_2^0)$ . Now consider  $C_1$  (resp.  $C_2$ ), which includes the types that buy only from VZ i.e.  $z_2 \leq z_2^0$ . In this region, I can invert only the allocation corresponding to VZ (resp. OG) i.e. Equation (11) (resp. 17), to recover the corresponding

$z_{1j}$  (resp.  $z_{2j}$ ). On the other hand, I can only recover the proportion of firms with  $(z_1, z_2) \leq (z_1^0, z_2^0)$ , i.e.  $\Pr(z_1 = z_1^0, z_2 = z_2^0) = \Pr(q_1 = q_{10}, q_2 = q_{20})$ .<sup>15</sup>

**4.2. Estimation.** The estimation is based on the equilibrium strategies of Section 2. More specifically, I observe  $(q_1, q_2)_j$ ,  $j = 1, 2, \dots, 6328$ . I assume that these purchases are the outcomes of the model equilibrium in (11) and (17). I define our econometric model accordingly as

$$q_{ij}(z_{ij}) = [z_{ij} - \tilde{G}_i(z_{ij}) - m_i - \frac{c\alpha_{-i} + c^2 q_{i0}}{b_{-i} + \beta_{-i}}] / [b_i - \frac{2c^2}{b_{-i} + \beta_{-i}}], \quad (29)$$

for all  $z_{ij} \in (z_i^0, \bar{z}_i]$  and  $q_{ij}(z_{ij}) = q_{i0}$  otherwise, where  $i = 1, 2$ , indexes VZ or OG and  $J = 6328$ . The pair  $(z_{1j}, z_{2j})$  is the source of randomness in the econometric model. Besides the above two optimal purchase equations, I have five structural equations defining the optimal price schedules, which give additional restrictions on the structural parameters.

I assume that every firm  $j$  draws  $(\theta_{1j}, \theta_{2j})$  independently from  $F(\cdot, \cdot)$ . Given the choice of price functions by the two publishers, every  $(\theta_{1j}, \theta_{2j})$  determines a pair  $(z_{1j}, z_{2j})$ , distributed with  $G(\cdot, \cdot)$ . The estimation procedure consists of several steps. In the first step, the quantity sold by each publisher is separately used to estimate the nonparametrically inverse hazard rate  $(1 - H_i(\cdot)) / (h_i(\cdot))$  for  $i = 1, 2$  using standard kernel density estimator. In the second step, I use the estimated inverse hazard rate, along with (29) and the five structural equations mentioned above to estimate the utility and cost parameters.

**4.2.1. Estimating the Density of ad.** 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  $h_i(\cdot)$  one can use the empirical distribution and kernel density estimator, respectively:

$$\hat{H}_i(q) = \frac{1}{N_i^*} \sum_{j=1}^{N_i^*} \mathbb{1}(q_{ij} \leq q), \text{ for } q \in [q_{i0}, \bar{q}_i]; \quad \hat{h}_i(q; \zeta) = \frac{1}{N_i^*} \sum_{j=1}^{N_i^*} \frac{1}{\zeta} K\left(\frac{q - q_{ij}}{\zeta}\right), \quad (30)$$

where  $\zeta$  is a bandwidth,  $K(\cdot)$  is a kernel. Among other things, however, 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 suffers 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; Devr oye, 1997]. Although various solutions have been proposed to address those shortcomings, I use adaptive kernel density

<sup>15</sup> Henceforth, I 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.

estimator based on linear diffusion processes as proposed by [Botev, Grotowski, and Kroese, 2010]. The main idea is to view the kernel density estimator as a transition density of a diffusion process, which leads to a simple kernel density estimator with substantially reduced asymptotic bias and mean squared error, and an improved plug-in bandwidth selection method. A key observation is that if the kernel in Equation (30) is Gaussian with location  $q_j$  and scale  $\zeta = \sqrt{t}$ , i.e. (suppressing the index  $i$  for the VZ and OG)  $K(q, q_j; \zeta) = (2\pi t)^{-1/2} \exp(-(q - q_j)^2/2t)$  then it is the unique solution to the diffusion partial differential equation

$$\frac{\partial}{\partial t} \hat{h}(q; t) = \frac{1}{2} \frac{\partial^2}{\partial q^2} \hat{h}(q; t), t > 0,$$

with  $q \in [q_0, \bar{q}]$  and initial condition  $\hat{h}(q; 0) = \frac{1}{N} \sum_{i=1}^N \delta(q - q_i)$  (the empirical density) and the boundary condition  $\frac{\partial}{\partial t} \hat{h}(q; t) \Big|_{q=q_0} = \frac{\partial}{\partial t} \hat{h}(q; t) \Big|_{q=\bar{q}} = 0$ . This means that a solution to a linear diffusion process (with proper boundary condition) is a valid (nonparametric) kernel density estimator that is locally adaptive to boundary conditions. So I follow [Botev, Grotowski, and Kroese, 2010] for estimation of the densities, ensuring that the estimated inverse hazard function is well defined, i.e.  $\lim_{q_i \rightarrow \bar{q}_i} \frac{1 - \hat{H}_i(q_i)}{\hat{h}_i(q)} = 0$  for  $i = 1, 2$ . Once  $\{\hat{H}_i(\cdot), \hat{h}_i(\cdot); i = 1, 2\}$  are estimated the parameter set  $X$  can be estimated. See Appendix (A) for more on bandwidth selection.

4.2.2. *Estimating the Parameters  $X$ .* Now, I outline the steps to estimate the parameters : (i) fix  $\underline{\theta}_2 = 0$  and estimate  $m_1$  and  $m_2$ ; (ii) choose any two values of  $q_{1j}$  and  $\bar{q}_{1j}$  and use (25) to estimate  $b_1$ ; (iii) estimate parameters  $X$  by solving

$$\min_{X \in \mathcal{X}} s_N(X)' s_N(X),$$

where

$$s(X) = \begin{pmatrix} cq_{10} - \alpha_2 - (b_2 + \beta_2)\bar{q}_2 + \bar{\theta}_2 \\ \bar{\theta}_1 = \bar{q}_1(b_1 + \beta_1) + \alpha_1 - cq_{20} \\ (b_2 + \beta_2)\bar{q}_2 - cq_{10} - \bar{\theta}_1 + \alpha_2 \\ (\bar{\theta}_2 - \bar{q}_2 b_2 - m_2)(b_1 + \beta_1) - c^2 q_{20} - c\alpha_1 + c\bar{\theta}_1 + 2c^2 \bar{q}_2 \\ (\bar{\theta}_1 + cq_{20})2q_{10} - b_1 q_{10}^2 - b_2 q_{20}^2 \end{pmatrix}$$

are the equations that identify the parameters. Since the parameters are identified there is a unique solution to the above minimization problem. After estimating  $X$ , the pseudo types  $z_1$  and  $z_2$  can be recovered and estimate the marginal (conditional) distributions and densities  $\{\hat{G}_i^*(\cdot), \hat{g}_i^*(\cdot)\}$  using the same diffusion method outlined above. The basic consistency result is stated below without the proof, which is a straightforward extension of the consistency results by [Guerre, Perrigne, and Vuong, 2000; Perrigne and Vuong, 2011a].

**Lemma 2.** *Suppose all the assumptions mentioned so far are valid. Then:*

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

Recall that  $\hat{q}_i$  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 conditional density given  $z_i > z_i^0$ . Next, I address the estimation of joint density of types.

4.2.3. *The Joint Density.* I am interested in estimating the joint cdf  $F(\cdot, \cdot)$  and the joint density  $f(\cdot, \cdot)$ . It is helpful to recall the data generating process: a firm  $j \in \{1, 2, \dots, N\}$  draws  $(\theta_{1j}, \theta_{2j})$  *i.i.d.* from  $F(\cdot, \cdot)$ , and for a given pair  $(T_1(\cdot), T_2(\cdot))$  it is transformed into  $(z_{1j}, z_{2j})$  and in equilibrium  $j$  chooses  $q_{1j}(z_{1j})$  and  $q_{2j}(z_{2j})$  from VZ and OG, respectively. Since the types are independent the observed ad are not independent and because not all consumers buy from both sellers the joint density of  $(z_1, z_2)$  cannot be (nonparametrically) identified everywhere, therefore neither can  $f(\cdot, \cdot)$  be (nonparametrically) identified everywhere. Therefore without further assumption  $f(\cdot, \cdot)$  cannot be identified.

What interests me is combining the two (conditional) marginals and then extending it to the whole support, for which I propose to use copula. Although the marginals are censored, they are nonparametric, and hence it is convenient to adopt a parametric form for the dependence function  $C_\kappa(\cdot, \cdot)$  (defined later) while keeping marginals unspecified. But there is no guidance as to what the parameter  $\kappa$  be as there are many families of copulas such as Gaussian,  $t$ -copula, etc. Since it is not entirely obvious what family is appropriate, choosing one without due diligence with respect to the data will defy the whole purpose of nonparametric identification of the conditional densities. So I propose to use the classic goodness-of-fit test and Vuong's non-nested model selection test to find the "best" family. In essence the method estimates the dependence between types using data on those who buy from both and uses this dependence to select the family that provides the best global fit.

A function  $C : [0, 1]^2 \rightarrow [0, 1]$  is a two-dimensional copula if  $C(\cdot, \cdot)$  is the joint distribution of random variable in  $[0, 1]^2$  with uniform marginals. 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 as mentioned earlier cannot be determined (nonparametrically) from the data. I assume that  $C(\cdot, \cdot)$  is known up to one parameter, i.e. it belongs to the class

$\mathcal{C}_0 = \{C_\kappa : \kappa \in \Gamma \subset \mathbb{R}\}$ , where  $\Gamma$  is the parameter set that contains  $\kappa$ . Some of the widely used parametric families are Clayton, Archimedean, Gaussian copulas.<sup>16</sup> If the copula family was known, i.e. if the null  $H_0 : C \in \mathcal{C}_0$ , were known to be true, then the parameter  $\kappa$  could be estimated either by maximizing the joint likelihood function or by matching some measure of dependence such as Kendall's  $\tau$ , or Spearman's  $\rho$ . However, I do not know the null, in other words I do not know if the copula is Clayton or Gaussian.

Since the marginal distribution of  $z_i$  is unspecified, I 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 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  $\hat{u}_{ij} := r_{ij}(J+1) = \hat{G}_i(z_{ij}) \times J/(J+1)$ .<sup>17</sup> The first sensible thing to do is to check if  $z_1$  and  $z_2$  are independent, even though it was verified that  $q_1$  and  $q_2$  are not independent for robustness. If they are independent then the joint distribution is simply product of two marginals. So, I test (the null)  $H_0 : \forall (u_1, u_2) \in [0, 1]^2, C(u_1, u_2) = u_1 u_2$  against (the alternative)  $H_A : \exists (u_1, u_2) \in [0, 1]^2, C(u_1, u_2) \neq u_1 u_2$ .

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 I compute the classic Cramér- von Mises statistic

$$T_N = \int_{[0,1]^2} J \{C_J(u_1, u_2) - u_1 u_2\}^2 du_1 du_2$$

to test for independence. There are two issues that complicate implementation of the test. First, the asymptotic distribution of  $T_J$  under the null is not distribution free, [Genest and Rémillard, 2004], and second, the distribution is also affected by the first-step errors from estimating the pseudo  $z_1$  and  $z_2$ . So, I compute the critical values using Bootstrap procedure as outlined in [Genest and Rémillard, 2004], [Kojadinovic and Holmes, 2009]. The test statistic is estimated to be  $\hat{T}_J = 1.66467$  with the  $p$ -value equal to 0.0005. Therefore I conclude that  $z_1$  and  $z_2$  are not independent.

The empirical copula  $C_J(\cdot, \cdot)$  is a consistent estimator of  $C(\cdot, \cdot)$ , [Fermanian, Radulović, and Wegkamp, 2004]), so a natural goodness-of-fit test would use some form of distance between the estimate of the candidate family  $C_{\kappa_J}$  and  $C_J(\cdot, \cdot)$  (under  $H_0$  that it is true). Let

$$\mathbb{C}(u_1, u_2) = \sqrt{N} \{C_J(u_1, u_2) - C_{\kappa_J}(u_1, u_2)\} \quad (31)$$

be an empirical process, and for a given family, let  $\hat{\kappa}$  be the value of the parameter that maximizes the pseudo-log-likelihood, i.e.  $\hat{\kappa} = \arg \max_{\kappa \in \Gamma} \left\{ l(\kappa) := \sum_{j=1}^N \log [C_\kappa(\hat{U}_{1j}, \hat{U}_{2j})] \right\}$ , as defined by [Genest,

<sup>16</sup> For example, bi-variate Clayton copula is of the form  $C(a, b) = (\max\{a^{-\kappa} + b^{-\kappa} - 1; 0\})^{-1/\kappa}$ .

<sup>17</sup> 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  $[0, 1]^2$ .

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-ml and  $p$ - values of the Cramér-von Mises statistic are computed using 10,000 Bootstrap replications and the rank in Vuong test.

Ghoudi, and Rivest, 1995; Genest, Quessy, and Rémillard, 2006]. Then, [Genest, Rémillard, and Beaudoin, 2009] show that the Cramér-von Mises statistic

$$\mathbb{T}_J = \int_{[0,1]^2} \mathbb{C}_J(u_1, u_2)^2 d\mathbb{C}_J(u_1, u_2) = \sum_{i=1}^J \left\{ \mathbb{C}_J(\hat{u}_{1j}, \hat{u}_{2j}) - C_{\kappa_j}(\hat{u}_{1j}, \hat{u}_{2j}) \right\}^2$$

can be used as a goodness-of-fit criteria and the test is consistent. To characterize the asymptotic distribution of the test I use the following weak convergence result for (31) from [Fermanian, Radulović, and Wegkamp, 2004]. Let  $C_\kappa^{[j]} = \frac{\partial C_\kappa}{\partial u_j}$  and  $\eta_\kappa$  be a  $C_\kappa$ - Brownian bridge.<sup>18</sup>

**Theorem 2.** *Let  $C_\kappa$  have partial derivatives. Then the empirical process (31)  $\mathbb{C}(u_1, u_2)$  converges weakly in  $l^\infty([0, 1]^2)$  to the tight, centered Gaussian process*

$$\mathbb{C}(u_1, u_2) = \eta_\kappa(u_1, u_2) - \eta_\kappa(u_1, 1)C_\kappa^{[1]}(u_1, u_2) - \eta_\kappa(1, u_2)C_\kappa^{[2]}(u_1, u_2), \quad u_1, u_2 \in [0, 1].$$

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  $\hat{\kappa}$ .<sup>19</sup> Therefore, the approximate  $p$ - values can only be obtained from a Bootstrap procedures outlined by [Genest and Rémillard, 2008]. In practice this is a slow method so I use the multiplier central limit theorem to determine the large sample distribution of the test statistic; see [Kojadinovic and Yan, 2011]. I 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 Bootstrap replications. The results are reported in first column of Table 3. It is evident then that only the Joe copula provides the best fit.

I also consider Vuong's non-nested model selection test à la [Vuong, 1989]. To implement the test, I consider any two families (from the seven families) and give +1 to the one that is selected by this

<sup>18</sup> A Brownian bridge is a tight centered Gaussian process on  $[0, 1]^2$  with covariance function  $\mathbb{E}[\eta_\kappa(u_1, u_2)\eta_\kappa(u'_1, u'_2)] = C_\kappa(u_1 \wedge u'_1, u_2 \wedge u'_2) - C_\kappa(u_1, u_2)C_\kappa(u'_1, u'_2)$ ,  $u_1, u_2, u'_1, u'_2 \in [0, 1]$  and  $a \wedge b = \min\{a, b\}$ .

<sup>19</sup>Using pseudo-log-likelihood is just one of many ways to estimate the parameters in the literature. For robustness, I also estimated the parameters that maximize Kendall's tau and Spearman's Tho, and reach the same conclusion.

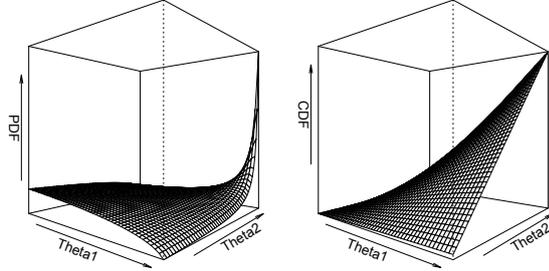


FIGURE 4. Estimated Joint pdf and cdf of recovered  $(\hat{\theta}_1, \hat{\theta}_2)$

test and  $-1$  otherwise. I follow the bootstrap procedure of [Clarke, 2007] to compute the  $p$ - values. For example, between Frank and Gaussian if Vuong's test selects Frank, it gets  $+1$  and Gaussian gets  $-1$ . I repeat this pair-wise test for all such pairs and add all the scores and present that in Table 3. The family with the highest score is the one selected. Again I find that the Joe copula is selected as the best model. Therefore, one can conclude that the best copula family is Joe and with  $\hat{\kappa} = 1.206497$  (*s.e.* = 0.0137). Then the estimated joint density of  $(z_1, z_2)$  becomes<sup>20</sup>

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

Then the estimated joint density of  $(\theta_1, \theta_2)$ , Figure 4, evaluated at  $z_j(\theta) \equiv z_j(\theta_1, \theta_2)$  becomes

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

**4.3. 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 ad 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 that of OG at

<sup>20</sup>A 2-dimensional copula  $C$  is called Archimedean if it has the representation  $C(u_1, u_2) = \phi(\phi^{-1}(u_1) + \phi^{-1}(u_2))$ ,  $(u_1, u_2) \in [0, 1]^2$ , where  $\phi(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is an Archimedean generator, i.e. continuous, strictly decreasing on  $\{\phi > 0\}$  and satisfying  $\phi(0) = 1$  and  $\lim_{t \rightarrow \infty} \phi(t) = 0$ . For the Joe copula,  $\phi(t) = 1 - (1 - \exp(-t))^{1/\kappa}$ .

$\hat{m}_2 = 3.145$ , which captures the differences in the 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 also means that competition between VZ and OG at the lower end is severe than at the upper end. This is reflected in the differences in prices: the difference in average price per pica widens as I 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.

Advertising is a business-to-business activity, so the demand or the willingness to pay for different sizes might depend on its usefulness in creating more demand via exposure. For example, a single doctor in a market might have less value for an ad than a market with few doctors. It is also conceivable that the value might be low if there are many doctors, in other words the value for ad might be inverse U-shaped as a function of competition. Although ad are treated as final consumption commodity, on account of complexity of the problem, once  $(\hat{\theta}_{1j}, \hat{\theta}_{2j})$ , is obtained, this question can be addressed by running a simple OLS regression on some measure of level of competition. I estimate the following model:

$$\hat{\theta}_{ij} = a_{i0} + a_{i1}\#C_{ij} + a_{i2}(\#C_{ij})^2 + a_{i3}avg(q_{ij}) + a_{i4}std.dev(q_{ij}) + a_{i5}National_{ij} + a_{i6}Guide_{ij} + \epsilon_{ij},$$

where  $\hat{\theta}_{ij}$  is firm  $j$ 's (pseudo) marginal willingness to pay for ad with  $P_i$ ,  $\#C_{ij}$  is the number of firms with the same sub-heading as  $j$  who advertise with  $P_i$  under the same heading as  $j$ , likewise  $(\#C_{ij})^2$  is its square,  $avg(\cdot)$  and  $std.dev(\cdot)$  are the average and standard deviation of ad sizes bought by firms in that industry.  $National_{ij}$  is a dummy if the firm is a national brand and  $Guide_{ij}$  is a dummy if  $j$  opts for guide option. Recall the Guide option provides additional advertising space by listing specialities and it covers attorneys, dentists, physicians, insurance companies, etc. The result for both VZ and OG is presented in Table 4 where the standard errors are reported in parenthesis and (\*\*) and (\*) denote estimates that are significant at 5% and 10% confident level, respectively.

As can be seen the direct effect of number of similar business units on the willingness to pay is positive but insignificant while the effect decreases significantly as shown by the coefficient of the Square number of firms. This suggests that the effect of competition decreases with the level

	$\hat{\theta}_1$	$\hat{\theta}_2$
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%.

of competitors.<sup>21</sup> The estimates suggest that whether or not a firm has national presence affects  $\hat{\theta}_1$  negatively but  $\hat{\theta}_2$  positively. The model does not explain either the demand pattern of firms with national brand or the brand effect of the publisher, because the demand side is captured by reduced form parameters. Explicitly modeling the demand side is important but beyond the scope of this paper.

4.3.1. *Counterfactual: Cost of Asymmetric Information.* What is the welfare cost of asymmetric information in this market? In this section I want to quantify the loss of welfare due to asymmetric information. We know that asymmetric information leads to second best outcome causing some welfare loss, but we want to quantify this loss and also see how this loss is shared across different consumers. To quantify the loss, I assume that the publishers play Stackelberg duopoly game with perfect information where P1 moves first and P2 moves second, find the social welfare under this game and compare it with the data. Since the sellers know the type of each consumer this leads to first degree price discrimination where they allocate quantity that equates (residual) marginal utility with marginal cost. I maintain the assumption that both sellers offer  $q_{10}$  and  $q_{20}$  for free.

In the second stage, a  $(\theta_1, \theta_2)$ - consumer who buys  $\tilde{q}_1$  from VZ and pays  $t_1$  gets gross utility  $D(q_2; \tilde{q}_1; \theta) = u(q_2, \tilde{q}_1; \theta) - t_1$  if she buys  $q_2$  from OG. For such  $q_2$  the maximum price she is willing to pay (and what will be charged by OG) is

$$t_2(q_2) = \theta_2(q_2 - q_{20}) - \frac{b_2}{2}(q_2^2 - q_{20}^2) + c\tilde{q}_1(q_2 - q_{20}). \quad (32)$$

OG will make a take-it-or-leave-it offer of  $q_2$  at  $t_2$  that maximizes the profit  $t_2(q_2) - m_2q_2$ . From Equation (32), 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

<sup>21</sup> The limitation of this regression is the fact that only few consumers buy from both and for most either one or both of the types do not vary, which can induce correlation in these types.

Qt: Incomplete Info.	Complete Info.	# Obs	$\Delta$ 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	\$1,900,000
(572, 843)	(575, 851)	1	\$2,200,000
(1709, 697)	(1711, 706)	6	\$15,000,000
(1709, 1154)	(1711, 1160)	3	\$16,000,000
(3171, 2153)	(3173, 2153)	1	\$55,000,000
(6330, 1621)	(6330, 1624)	1	\$210,000,000

TABLE 5. Welfare cost of asymmetric information: Comparison of welfare under incomplete Information and a counterfactual of complete information.

VZ can charge for any  $q_1$  is

$$\begin{aligned}
t_1(q_1) &= \theta_1(q_1 - q_{10}) + \theta_2(q_2(q_1) - q_2(q_{10})) - \frac{b_1}{2}(q_1^2 - q_{10}^2) \\
&- \frac{b_2}{2}(q_2(q_1)^2 - q_2(q_{10})^2) + c(q_1q_2(q_1) - q_{10}q_2(q_{10})) - (t_2(q_1) - t_2(q_{10})).
\end{aligned}$$

where  $t_2(q_{10})$  can be determined by evaluating (32) at  $q_{10}$ . Then,  $q_1 = (\theta_1 - m_1)/b_1$  maximizes the profit  $t_1(q_1) - m_1q_1$  and the corresponding  $q_2$  (as a function of  $q_1$ ) is  $q_2 = [b_1(\theta_2 - m_2) + c(\theta_1 - m_1)]/[b_1b_2]$ . 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  $\int_{q_{20}}^{q_2} D_2(q_1^*, y)dy - K_2 - m_2q_2$ . 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. Given the quasi-linear utility, I find that VZ gains \$2,651,052,914 while OG gains \$48,330,062 and the firms will 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 be equal to its valuation for  $(q_{10}, q_{20})$ , which is increasing in type. See Table 5, which 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.

## 5. CONCLUSION

In this paper I use data from one market, where two sellers use nonlinear pricing to sell their differentiated product to (potentially heterogeneous) consumers, to recover and estimate the unobserved consumer heterogeneity, utility function and the cost function. To achieve that I rely a model of competitive principal-agent, where consumers have private information about their multidimensional preferences. Using aggregation method to (endogenously) generate a one dimensional sufficient statistic for the consumer (unobserved) preferences, I characterize the equilibrium allocation rule that is monotonic in consumer types and maps model parameters to observed sales data. This allocation rule plays the same role in identification as does the monotonic bidding strategy in a first price sealed bid auction. Then I implement the model on the data on ad bought from two Yellow Pages directories in central Pennsylvania. Normalization of the outside option for the lowest types is sufficient to identify utility, cost parameters and the truncated marginal densities of the sufficient statistics. Then I use empirical copula method to combine the two marginal densities into a joint density, and to choose the right family of copula family I compare seven most widely used families and choose the one that is selected according to both Cramér-von Mises goodness-of-fit test and Vuong's non-nested model selection test.

The estimates reasonably rationalize the observed data and suggest that the consumers are sufficiently heterogeneous in terms of the valuation for the two advertising choices. Interestingly, the estimates suggest that the competition between the two publishers is higher at the lower end of the market, where the density puts more mass. This explains why the per unit prices diverge as we move up the size of ad. In summary, the paper shows how one can use supply side to estimate demand when information on the consumers side is not available to estimate the underlying consumer heterogeneity.

Estimation of the joint density of unobserved consumer types, under competition, is a prerequisite to quantify the effect of competition on welfare, when we want to allow the product varieties to be endogenous. The estimated joint density of types can be used to simulate a merger by solving the multidimensional screening problem for a multi-product monopolist à la [Rochet and Choné, 1998]. Having said that, solving [Rochet and Choné, 1998] is a difficult problem, and only recently some progress has been made in this direction, see [Ekeland and Moreno-Bromberg, 2010]. This suggests that we can combine this paper with these two papers to study mergers and consumer heterogeneity, a lacuna in the literature that is worth exploring. I leave this new line of enquiry as a future research question.

## APPENDIX A. BANDWIDTH SELECTION

In this section I present the procedure followed to estimate the kernel density and is taken from [Botev, Grotowski, and Kroese, 2010]. Given  $N$  IID realizations  $Y = \{Y_1, \dots, Y_N\}$  from an unknown continuous density  $\tilde{f}(\cdot)$ . The Gaussian kernel density estimator is defined as  $\hat{f}(y; t) = \frac{1}{N} \sum_{j=1}^N \left( \frac{1}{\sqrt{2\pi t}} e^{-(y-Y_j)^2/2t} \right)$ . Asymptotically optimal value of  $\xi$  minimizes the Asymptotic Mean Integrated Squares of Error and is given by  $*t = \left( \frac{1}{2N\sqrt{\pi} \|\tilde{f}''\|^2} \right)^{2/5}$ . Since the optimal  $*t$  depends on the functional  $\|\tilde{f}''\|^2$  and using the estimator of this functional gives the following plug-in method to select optimal bandwidth  $*\hat{t} = \left( \frac{8+\sqrt{2}}{24} \frac{3}{N\sqrt{\pi/2} \|\tilde{f}^{(3)}\|^2} \right)^{2/7}$ , but this requires estimating  $\tilde{f}^{(3)}$  and requires us to solve for a fixed point of an infinite sequence; see [Wand and Jones, 1995] for a solution. For the kernel density based on diffusion process the following algorithm can be used: (i) From the data estimate the Gaussian kernel density using  $*\hat{t}$ ; (ii) Estimate  $\|L\tilde{f}\|^2$  via the plug-in estimator from step 1 using  $*t$ , where  $L(\cdot) := \frac{1}{2} \frac{\partial}{\partial y} (a(y) \frac{\partial}{\partial y} (\cdot))$  is a differential operator with  $a(y) = \tilde{f}(y)^\iota, \iota \in [0, 1]$ ; and (iii) Use estimate from step 2 with the variance  $\sigma$  to get the optimal bandwidth:  $t^* = \left( \frac{\mathbb{E}_{\tilde{f}}[\sigma^{-1}(Y)]}{2N\sqrt{\pi} \|L\tilde{f}\|^2} \right)$ .

## APPENDIX B. TABLES

VZ Picas	VZ Percentage	OG Picas	OG Percentage	Color Category 1		Color Category 2		Color Category 3		Color Category 4	
Listing				VZ Price	OG Price						
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%
<hr/>				
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

- ARMSTRONG, M. (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, 64, 51–75. [3, 29](#)
- ARMSTRONG, M., AND J. VICKERS (2001): "Competitive Price Discrimination," *The RAND Journal of Economics*, 32, 579–605. [3](#)
- ARYAL, G. (2013): "Supplement: An Empirical Analysis of Competitive Nonlinear Pricing," *Mimeo.*, [Link to the Note. 11, 14, 16](#)
- ARYAL, G., AND I. PERRIGNE (2011): "An Insurance Model with Bi-Dimensional Screening," *Working Paper, ANU. 10*
- ARYAL, G., I. PERRIGNE, AND Q. H. VUONG (2010): "Identification of Insurance Models with Multidimensional Screening," *Working Paper, ANU. 5, 10*
- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890. [1](#)
- BERRY, S. T. (1994): "Estimation of Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25(2), 242–262. [1](#)
- BORENSTEIN, S. (1991): "Selling Costs and Switching Costs: Explaining Retail Gasoline Margins," *The RAND Journal of Economics*, 22, 354–369. [3](#)
- BORENSTEIN, S., AND N. ROSE (1994): "Competition and Price Dispersion in the U.S. Airline Industry," *Journal of Political Economy*, 102, 653–683. [3, 9](#)
- BOTEV, Z. I., J. F. GROTHOWSKI, AND D. P. KROESE (2010): "Kernel Density Estimation Via Diffusion," *The Annals of Statistics*, 38(5), 2916–2957. [24, 33](#)

- BUSSE, M., AND M. RYSMAN (2005): "Competition and Price Discrimination Competition and Price Discrimination Competition and Price Discrimination in Yellow Pages Advertising," *The RAND Journal of Economics*, 36, 378–390. [3](#), [9](#)
- CLARKE, K. A. (2007): "A Simple Distribution-Free Test for Nonnested Model Selection," *Political Analysis*, 15(3), 347–363. [28](#)
- COHEN, A. (2008): "Package Size And Price Discrimination In The Paper Towel Market," *International Journal of Industrial Organization*, 26(2), 502–516. [3](#)
- CRAWFORD, G., AND M. SHUM (2006): "Monopoly Quality Degradation in Cable Television," *Journal of Law and Economics*, 49, 181–209. [2](#)
- DEVROÛYE, L. (1997): "Universal Smoothing Factor Selection in Density Estimation: Theory and Practice," *Test*, 6, 223–320. [23](#)
- 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. [2](#)
- EINAV, L., A. FINKELSTEIN, S. RYAN, P. SCHRIMPE, AND M. CULLEN (2012): "Selection on Moral Hazard in Health Insurance," *American Economic Review*, Forthcoming. [2](#)
- EINAV, L., M. JENKIS, AND J. LEVIN (2012): "Contract Pricing in Consumer Credit Markets," *Econometrica*, 80(4), 1387–1432. [2](#)
- EKELAND, I., J. HECKMAN, AND L. NESHEIM (2002): "Identifying Hedonic Models," *American Economic Review*, 92(2), 304–309. [5](#), [10](#)
- (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112, 60–109. [5](#), [10](#)
- EKELAND, I., AND S. MORENO-BROMBERG (2010): "An Algorithm for Computing Solutions of Variational Problems with Global Convexity Constraints," *Numerische Mathematik*, 115, 45–69. [5](#), [32](#)
- EPSTEIN, L. G., AND M. PETERS (1999): "A Revelation Principle for Competing Mechanisms," *Journal of Economic Theory*, 88, 119–160. [2](#)
- FAN, Y. (2013): "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market," *American Economic Review*, 103(5), 1598–1628. [5](#)
- FERMANIAN, J. D., D. RADULOVIĆ, AND M. H. WEGKAMP (2004): "Weak Convergence of Empirical Copula Processes," *Bernoulli*, 95, 119–152. [26](#), [27](#)
- GENEST, C., K. GHOUDI, AND L.-P. RIVEST (1995): "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrika*, 82(3), 543–552. [26](#)
- 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. [27](#)

- GENEST, C., AND B. RÉMILLARD (2004): "Tests for Independence and Randomness Based on Empirical Copula Process," *Test*, 13(2), 335–369. [26](#)
- 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. [27](#)
- 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. [27](#)
- GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–574. [4](#), [24](#)
- IVALDI, M., AND D. MARTIMORT (1994): "Competition under Nonlinear Pricing," *Annales d'Économie et de Statistique*, 34, 71–114. [3](#), [9](#)
- JONES, M. C., J. S. MARRON, AND S. J. SHEATHER (1996): "A brief survey of bandwidth selection for density estimation," *Journal of American Statistical Association*, 91, 401–407. [23](#)
- 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. [26](#)
- 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. [4](#), [27](#)
- LESLIE, P. J. (2004): "Price Discrimination in Broadway Theatre," *The RAND Journal of Economics*, 35, 520–541. [3](#)
- LOTT, J. R., AND R. D. ROBERTS (1991): "A Guide to the Pitfalls of Identifying Price Discrimination," *Economic Enquiry*, 29(1), 14–23. [3](#)
- MARRON, J. S., AND M. P. WAND (1992): "Exact Mean Integrated Error," *The Annals of Statistics*, 10, 712–736. [23](#)
- MARTIMORT, D., AND L. A. STOLE (2002): "The Revelation and Delegation Principles in Common Agency Games," *Econometrica*, 70(4), 1659–1673. [2](#)
- MASKIN, E., AND J. RILEY (1984): "Monopoly with Incomplete Information," *The RAND Journal of Economics*, 15, 171–196. [2](#), [11](#)
- MCMANUS, B. (2006): "Nonlinear Pricing in an Oligopoly Market: The Case of Speciality Coffee," *The RAND Journal of Economics*. [3](#)
- MUSSA, M., AND S. ROSEN (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317. [2](#)
- NELSON, R. B. (1999): *An Introduction to Copulas*. Springer-Verlag New York, Inc. [25](#)

- OREN, S. S., S. A. SMITH, AND R. WILSON (1983): "Competitive Nonlinear Tariffs," *Journal of Economic Theory*, 29(1), 49–71. [3](#)
- PERRIGNE, I., AND Q. VUONG (2011a): "Nonlinear Pricing in Yellow Pages," *Mimeo*. [5](#), [24](#)
- PERRIGNE, I., AND Q. H. VUONG (2011b): "Nonparametric Identification of a Contract Model with Adverse Selection and Moral Hazard," *Econometrica*, 79(5), 1499–1539. [5](#)
- ROCHET, J.-C. (1987): "A necessary and sufficient condition for rationalizability in a quasi-linear context," *Journal of Mathematical Economics*, 16(2), 191–200. [16](#)
- ROCHET, J.-C., AND P. CHONÉ (1998): "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 66(4), 783–826. [3](#), [5](#), [14](#), [16](#), [32](#)
- SHEPARD, A. (1991): "Price Discrimination and Retail Price Discrimination and Retail Configuration," *Journal of Political Economy*, 99, 30–53. [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. Elsevier. [3](#), [17](#)
- TERRELL, G. R., AND D. W. SCOTT (1992): "Variable Kernel Density Estimation," *The Annals of Statistics*, 20, 1236–1265. [23](#)
- VERBOVEN, F. (2002): "Quality-based Price Discrimination and Tax Incidence—the Market for Gasoline and Diesel Cars in Europe," *The RAND Journal of Economics*, 33, 275–297. [3](#)
- VUONG, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57(2), 307–333. [4](#), [27](#)
- WAND, M., AND M. C. JONES (1995): *Kernel Smoothing*. Chapman and Hall, London. [33](#)
- WILSON, R. (1993): *Nonlinear Pricing*. Oxford University Press. [2](#), [12](#)
- YANG, H., AND L. YE (2008): "Nonlinear Pricing, Market Coverage, and Competition," *Theoretical Economics*, 3, 123–153. [3](#)