AN EMPIRICAL ANALYSIS OF COMPETITIVE NONLINEAR PRICING*

GAURAB ARYAL† AND MARIA F. GABRIELLI‡

ABSTRACT. We estimate a model of competitive nonlinear pricing, when there is asymmetric information between buyers and sellers about former's multidimensional preferences. We use a novel dataset on advertisements bought by *all* local businesses from two (duopoly) Yellow Pages directories in Central Pennsylvania. First, we study the identification of the joint distribution of preferences, (constant) marginal costs of publishing and utility parameters. Second, we find a significant welfare loss due to asymmetric information. Third using a merger simulation we estimate efficiency and distributional effects of (a lack of) competition. In particular, when we move from duopoly to a monopoly, we find that: (i) the producer surplus increases substantially; (ii) many lower type consumers are excluded; (iii) the product space under monopoly increases; and as a result (iv) consumer surplus for higher types increases. The total consumer surplus, however, decreases but it does not affect the inequality in the distribution of consumer surplus.

Keywords: Nonlinear Pricing, Competition, Identification, Multiproduct Monopoly.
JEL classification: C14, D22, D82, L11, L13.

1. INTRODUCTION

Second-degree price discrimination is widely used by profit maximizing seller(s) with market power. Nonlinear pricing is an example of second-degree price discrimination where price is not strictly proportional to the quantity, and it is often

Date: July 4, 2018.
* Previously this paper was circulated under the title "Competition and Nonlinear Pricing in Yellow Pages." We are thankful to Yao Huang for her help with an earlier version of the paper. We want to thank Herman Bierens, Vijay Krishna, Mark Roberts, Neil Wallace, Raimundo Soto and participants at numerous seminars and conferences. Special thanks are due to Isabelle Perrigne and Quang Vuong for their feedback.
† University of Virginia; e-mail: aryalg@virginia.edu.
‡ CONICET and Universidad Nacional de Cuyo. e-mail: florgabrielli@gmail.com.

used to increase profit(s) when faced with heterogeneous consumers (Figure 2). Under nonlinear pricing, sellers offer quantity-price pairs and each consumer chooses her preferred option and pays the associated price. For more on this topic see Mussa and Rosen [1978]; Maskin and Riley [1984]; Wilson [1993]; Rochet and Stole [2002], and Martimort and Stole [2009].

In this paper we consider nonlinear pricing for Yellow Pages advertisements in Central Pennsylvania. There are two directories one published by Verizon, which is also a utility company, and a publishing house called Ogden. The two publishers compete with each other and offer menu of ads and prices to local businesses. Once the ads are chosen, they distributes the directories across the region. To induce high-valued customers to choose bigger and more expensive ads, each seller offers discounts. Consequently the marginal prices are different from the average prices (see Figure 2), while possibly distorting the size of smaller ads. Our objective is to use a unique dataset from this market to shed light on the following questions: How big are these distortions? What are their effects on efficiency? Will the distortion increase or decrease if these directories were owned by a monopolist? Would the distribution of welfare worsen? What happens to these numbers if the monopolist were to sell only one directory?

Asymmetric information about consumers' preferences and the sellers' desire to maximize expected revenue affect both allocative efficiency and total welfare. But the extent of this inefficiency depend on the model primitives and the market structure, which makes this an inherently empirical question. Even though there are many papers that study the role of asymmetric information, and market power on welfare, they do it separately. With notable exceptions of Ivaldi and Martimort [1994] and Miravete and Röller [2004], most papers that estimate the welfare cost of asymmetric information assume a monopoly seller. As a consequence, our understanding of the role of market power in an environment with asymmetric information is limited, and that is the subject of our paper.

The role of competition on welfare depends crucially on the exact nature of the product. If consumers choose only one product, competition can improve the consumer welfare because consumers now have better outside options. If consumers can buy from multiple sellers then competition might shrink the product space, which in turn reduces welfare. We estimate these countervailing forces.

For that we use the *competing principal-agent* framework of Ivaldi and Martimort [1994], and then develop and estimate a model of duopoly nonlinear pricing with multidimensional preferences. Our model characterizes menus when Verizon and Ogden commonly know only the joint distribution of consumer preference and their publishing costs, but not the individual preferences of the consumers. Under this model the two publishers simultaneously choose the sizes of ads to offer and the price for each ad, and consumers observe both menus and self-select the best option(s). Our model can rationalize an key data feature that some consumers advertise with only one directory, while others will choose to advertise with both, and others still choose the free outside option of just printing one's name.

Our data contains *all* ads bought by *all* consumers in the market, and thus we know the complete list of all options sold by both publishers, which we use to construct the product spaces. One limitation is that we do not observe any other consumer characteristics. To circumvent this problem we use the supply side optimality conditions to identify the model primitives, which include the joint density of consumer preferences, the marginal cost of printing and common utility parameters.

For identification we use the monotonicity of the optimal nonlinear pricing: consumers with higher willingness to pay will buy bigger ads. Without perfect screening identification would be infeasible; see Aryal [2017]. Given that not all consumers advertise with both advertisers, and not all choose to advertise at all, we can nonparametrically identify only the truncated (marginal) distribution of each consumer type. To combine the two marginals, we use the Cramér-von Mises test and the Vuong [1989] test to select the best Copula that fits the data. We find that the Joe copula provides the best fit. We also use information about nonlinear pricing to identify other parameters. For instance, the largest ad is never distorted, which means the marginal utility is the same as marginal cost, which will be the same as the (observed) marginal price.

Using the estimated parameters first we conduct a counterfactual exercise to asses the welfare cost of asymmetric information. For each type we determine a product and price that equates marginal revenue and marginal cost. Comparing the welfare under this perfect information outcome with the welfare estimated from the data we estimate the cost of asymmetric information to be large, at around 20% of the consumer surplus under incomplete information. Because the quantity distortion is severe for the lower types, we find that complete information benefits the lower types more than the higher types, i.e., those who buy bigger and colorful ads.

In the next counterfactual exercise we study a hypothetical merger and its effects on welfare and the distribution of consumer surplus. We know that monopolies generate a deadweight loss because they sell at a price above marginal cost. But monopolies that can price discriminate remove some of the deadweight loss, but can also capture more of the consumer surplus. While reduction in the deadweight loss improves efficiency, it comes at the cost of lower consumer surplus. Competition also reduces the deadweight loss, but with asymmetric information competition can lead to a smaller product space (as we document in this paper).

We begin with our duopoly data and estimate the total producer surplus and distribution of consumer surplus. Then we solve for nonlinear pricing if the two directories were owned by one seller, and determine the counterfactual choices, prices, consumer and producer surplus corresponding to each type. To the best of our knowledge these exercises are new in the literature. However, because solving for optimal bundling of a multi-product monopolist with multidimensional asymmetric information is a notoriously difficult problem [Armstrong, 1996, 1999; Rochet and Choné, 1998; Basov, 2005], we follow Carroll [2017] and solve for the worst-case (profit) bound of the multi-product monopolist. Carroll [2017] show that the bound is achieved by selling two products separately, i.e., without bundling. While this method allows us to quantify the first-order effect of merger on total welfare and efficiency, its drawback is that we are silent about the benefit of bundling; see Bakos and Brynjolfsson [1999]; McAfee, McMillan, and Whinston [1989] for more.

We find that the total producer surplus increases considerably after the merger. This result shows that the prediction from Corts [1998] applies even to second-degree price discrimination. There are two channels through which lack of competition affects the profit. First, the monopolist increases the product line for Ogden. In fact the biggest ad offered by the monopolist doubles (from one page ad to two-pages ad as we see in the data for Verizon). Second, many more low-type consumers are excluded under monopoly than under duopoly. While consumer surplus for the high types increases, overall consumer surplus decreases by 17%.

We also consider a situation where after the merger (or due to exit) only Verizon is available. As expected, both the consumer surplus and the producer surplus decrease because not offering Ogden hurts consumers with a high valuation for Ogden. In fact the total consumer surplus decreases by 24%. On the other hand, comparing a monopolist who offers both Verizon and Ogden with a monopolist who only offers Verizon (the most profitable of the two) we find that the producer surplus decreases by 29 percent points. Interestingly increase in market power has no distributional effect (measured by *Gini* coefficient) on consumer surpluses.

Related Work. There is a large literature on price discrimination that is related to our paper. We refer the reader to Wilson [1993] and Tirole [1998]. Although the majority of the papers focus on a monopoly seller, there some theory papers that allow competition [Oren, Smith, and Wilson, 1983; Epstein and Peters, 1999; Armstrong and Vickers, 2001; Rochet and Stole, 2002; Armstrong, 2006; Martimort and Stole, 2002; Stole, 2007; Martimort and Stole, 2009; Zhou, 2017], and some empirical applications Ivaldi and Martimort [1994] and Miravete and Röller [2004]. Our paper is also related to [Crawford and Shum, 2006; Einav, Finkelstein, and Cullen, 2010; Einav, Finkelstein, Ryan, Schrimpf, and Cullen, 2012]; and Einav, Jenkis, and Levin [2012] that estimate the cost of asymmetric information under a monopoly.

Because we estimate consumer preferences or the "demand" from choices, our paper is also tangentially related to the vast literature on demand estimation based on discrete choices; see Berry [1994]; Berry, Levinsohn, and Pakes [1995]; Nevo [2000] and Train [2009]. There are, however, some important differences in our approaches. First and foremost our empirical analysis exploits continuity in options, whereas discrete choice models are, by definition, discrete. Second we allow for multidimensional asymmetric information, or unobserved (to the sellers and to the researchers) consumer heterogeneity, while these models assume perfect information. Third, using mechanism design literature we can endogenize the product spaces offered by each seller whereas most of the discrete choice literature treats the product spaces as exogenous, with the exception of Fan [2013]. Fourth, our method can naturally accommodate quantity discount, which means average prices are different from marginal prices, while the discrete choice literature considers only average prices.

Another paper that is related to our paper is Nevo, Turner, and Williams [2016], although the approaches are different. One major distinction between our papers is the data: while Nevo, Turner, and Williams [2016] have access to a very rich data on consumer demographic characteristics, the only information we know are the choices and payments. They estimate the demand without using the supply side.

In terms of the application, Rysman [2004] study the market for Yellow Page advertisements. While we focus on asymmetric information and (exogenous) consumer heterogeneity he uses two-sided market. In another paper, Busse and Rysman [2005] estimate the relationship between competition and prices in the Yellow Page industry, and like them we find that competition is severe at the lower quantity.

There are some empirical papers that use the monopoly principal-agent framework, e.g., Leslie [2004]; McManus [2006]; Cohen [2008] and Luo, Perrigne, and Vuong [2017]. In terms of modeling multidimensional preferences with competition, our paper has some similarity with Epple, Romano, and Sieg [2006] and Fu [2014], albeit in a different context. For example in Epple, Romano, and Sieg [2006] students differ in ability and income, and in Fu [2014] they differ in taste and ability.

Lastly, our identification strategy is related to that in auction [Guerre, Perrigne, and Vuong, 2000], price discrimination [Aryal, 2017; Aryal, Perrigne, and Vuong, 2016; Luo, Perrigne, and Vuong, 2017], and hedonic models [Ekeland, Heckman, and Nesheim, 2002, 2004; Heckman, Matzkin, and Nesheim, 2010].

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

2. Data

Our product is advertisements sold by two Yellow Page directories in central Pennsylvania (State College and Bellefonte), U.S., in 2006. One of the directories is published by Verizon (henceforth, VZ), which is a utility provider, and the other is published by Ogden (henceforth, OG). We have information on the ads bought by *everyone* with a telephone number registered as a "business phone number." So a dental clinic, or a salon who has a phone line for business is a potential consumer. For every consumer in this market we hand collect the ads they bought from either VZ or OG or from both, and with the price schedules determine their total payment.

Data Sources. To get the list of all consumers, we use the fact that there is a norm in this industry to publish the names and addresses of all business phone numbers by default and free of charge. We treat these free options (the smallest ad, known as the standard listing) as the outside options that are available to every consumer. We match each name with the ads from the two directories, and for every consumer we know whether or not the consumer placed ads with VZ or OG or both, and we also know the size, colors and prices for those ads.

To determine the prices we use the price schedules that VZ and OG have to send to the Yellow Page Association, which is a national trade organization of all Yellow Page publishers in the U.S. As members of this association, VZ and OG have to provide the association with their prices for all offered ad options. We use these

TABLE 1. Sales and Revenues by Ad Category.

Verizon	Observations	Sales Share	Revenue ($)	Revenue Share
Standard Listing	2,152	31.54%	0	–
Listing	2,471	36.22%	648,128	11%
Space Listing	1,486	21.78%	1,132,531	19%
Display	714	10.46%	4,236,973	70%
Total	6,823	100.00%	6,017,632	100%
Ogden				
Standard Listing	5,910	86.62%	0	–
Listing	447	6.55%	88,932	11%
Space Listing	241	3.53%	143,656	17%
Display	225	3.30%	609,560	72%
Total	6,823	100%	842,148	100.00%

Notes: The table shows the sales and revenue by ad categories, for both Verizon and Ogden. Each row denotes the type of the ad, and the columns denote the share of that ad in total demand and their respective shares in revenue.

quotes to construct the total prices paid by each consumer to each publisher, unless if the consumer chooses the standard listing in which case the payment is zero.

Differentiated Directories. The two directories differ in both paper quality and size, but both offer a large number of options to choose from and are freely distributed over the same geographic area in Central Pennsylvania. The standard unit of measurement in this industry is called a pica, which is approximately $1/6$ of an inch. Even the free ad (the standard listing) is bigger in VZ than in OG (at 12 sq. picas vs. 9 sq. picas). This difference increases with the size of the ad. For example, a full page ad in VZ is $3,020$ sq. picas while it is only $1,860$ sq. picas in OG. The largest ad in VZ is a full two-page color ad whereas OG does not offer such an option; the largest ad offered by OG is only a one page ad.

The VZ directory is slightly bigger, thicker and with higher quality (glossier) paper than the OG directory. As a consequence, VZ can offer three columns in each page for advertisement while OG can offer only two columns. In terms of circulation, VZ distributes more than $215,400$ copies while Ogden distributes $73,000$, but their geographic coverage is similar. These features suggest that the two directories are differentiated "products" with possibly stronger demand, i.e., higher willingness to pay, for VZ ads than OG ads.

The advertisement options can be broadly divided into three categories: listing, space listing, and display. A space listing refers to an option where a space is allocated within the column under an appropriate business heading (such as Doctors, Salons, etc.). Both publishers offer different options within each category. The display ad is the most expensive option, and it refers to a listing option with a space (that could cover up to two pages) where the consumer can choose colorful pictures. VZ offers nine different variations within this category and OG offers six. On top of that, VZ offers five color options – no color, one color, white background, white background plus one color and multiple colors including photos– while OG offers the same options except the 'white background' option. Another difference is the font size, for example, VZ offers three font sizes to just list the names, address and phone number(s). OG offers listing with only two font sizes.

A large fraction of consumers opt for the free standard listing option, Table 1. The listing option accounts for 36.22% and 6.55% of the total ad sales in VZ and OG, respectively, and space listing accounts for 21.78% and 3.53% of the total ad sales in VZ and OG, respectively. From this table we can also see that the display option, which is the most expensive option, accounts for 70% or more of the revenue for both VZ and OG. Roughly 67% of consumers choose listing and 10% choose display in VZ, while 94% and 3.3% choose listing and display in OG, respectively.

To give an idea that indeed there are many options to choose from, we present a subset of the options in Appendix Table A-1. We find that: (i) for any size, color accounts for most of the differences in prices, e.g., a full-page display ad with no color costs $18,510 in VZ and $6,324 in OG, while for multiple colors the prices increase to $32,395 and $9,675, respectively; (ii) VZ's prices are significantly higher than Ogden's across all comparable advertising options, e.g., a half-page display without color costs $10,093 in VZ and only $3,372 in OG; (iii) the price differences between VZ and OG are smaller for the lower-end options, such as listing, than for the upper-end options such as display. For instance, VZ's average price is 130% higher than OG's for the display option while the difference in prices is only 18% for space listing and 17% for standard listing with no color; and (iv) for a given color category, both VZ and OG offer a quantity discount: the price per sq. pica decreases with the ad size. This last feature is an example of (implicit) quantity-discount in nonlinear pricing, which we discuss momentarily.

In our sample we find that 54% of consumers advertise exclusively with VZ, whereas only 2% advertise exclusively with OG; 12% advertise with both and the remaining choose the free (standard list) ad. The average prices paid in each directory by the firms purchasing from both directories are higher than those who purchase from only one directory, which may indicate a higher valuation of advertising among this group. A similar pattern is observed with respect to ad sizes. Our empirical exercise is to determine the joint distribution of consumer preferences, cost and utility parameters that match these salient features in our data.

Quality Adjusted Quantity. Ads differ in both sizes and color, but in order to keep the model tractable we combine the two attributes into one, which we shall call "quality-adjusted quantity." Heuristically, we estimate price gradients to determine the relative importance of sizes and colors and use this weight to express each ad in the data in terms of this new aggregate measure. We reduce the dimension of ads by projecting them on the "space" of the most expensive (multi-colored display) ad.

This projection is based on the following simplification. We consider a non-color listing ad of size 20 sq. pica, say, that costs $300, and ask, "If the consumer had chosen to spend $300 on a multi-colored display ad what would be the size of such an ad?" The answer to this question is the one-dimensional quality-adjusted quantity. To achieve this dimension reduction we estimate a quadratic single-index model on tariffs, for VZ and OG separately that accounts for the trade-off between size and color. Although this method has its own drawbacks, it nonetheless allows us to keep an already complicated model (competitive nonlinear pricing with multidimensional private information) tractable without sacrificing the main data features.

This aggregation won't work if the publishers use color as a separate tool for price-discrimination, because it must maintain the ordering of ads choices from the consumers' perspective. Indeed, we find that keeping the ad size fixed, the relative prices do not vary with colors. In particular, discounts are offered for large ads while no such discounts are observed for ads with multiple colors. Also the ratio of the (marginal) prices for two different colors are constant across different sizes. This is consistent with color not being used for screening.

Let q_{ij} be the size of an ad purchased by consumer j from seller $i \in \{VZ, OG\}$, and let q_{i0} denote the free outside option offered by seller i. To estimate the tradeoff between color and size for $q_{ij} > q_{i0}$ we estimate the following single-index model:

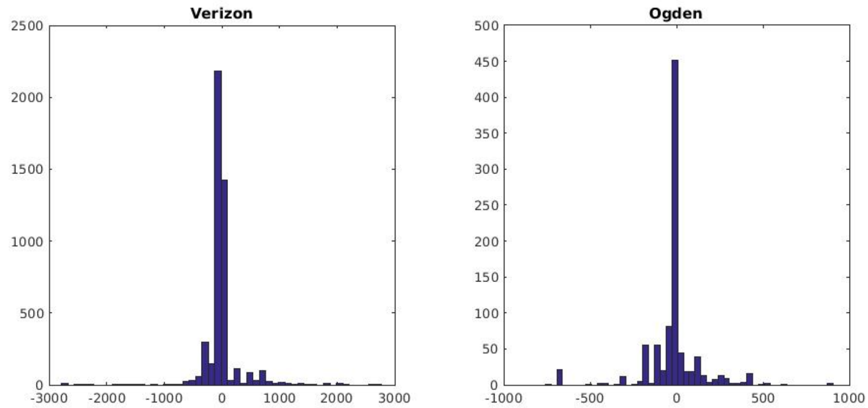
$$T_{ij} \quad = \quad d_i'\gamma_i + \tilde{T}_i(q_{ij} + \delta_i \times \texttt{color}_{ij}) + \varepsilon_{ij},$$

where d_i is a vector of two dummy variables for different sizes, T_{ij} denotes the payment made by consumer j to seller i, $\tilde{T}_i(\cdot)$ is an unknown function, $\texttt{color}_{ij} \in \{0, 1\}$ is a dummy variable for color which is equal to one if the ad is colored and zero otherwise so that the unknown parameter δ_i is the premium on color ads, and ε_{ij} is a classic (measurement) error.

Let $d_i = [\texttt{dsize}_i\ \texttt{dsmall}_i]$ be the two dummy variables, such that $\texttt{dsize}_i = 1$ if $i = VZ$ and the ad is 24 sq. pica or if $i = OG$ and the ad is 15 sq. pica, and $\texttt{dsmall}_i = 1$ if $i = VZ$ and the ad is smaller than 27 sq. pica or if $i = OG$ and the ad is smaller than 24 sq. pica. By allowing the intercepts to be different for the smaller sized ads, we intend to capture any nonlinearities in ad sizes, color and prices which seems to be more (empirically) important for size 24 sq. in VZ and 15 sq. in OG.[1]

[1] The purpose of using this specification is to aggregate the color and size and express them in terms of highest-quality. This requires finding a root of a quadratic function that is positive and smaller than the observed size to capture the idea that a full-page black and white ad, say, must equal to a multi-color ad that is smaller than a full-page. We find that the estimation for VZ is quite robust to different specifications, whereas the estimation for OG is sensitive to the exact functional form. This is possibly because a larger share of consumers choose VZ over OG. We choose this particular size dummies that gave us the most reasonable quality adjusted quantity predictions. We want to use the same specification for VZ and OG that is why we include the same set of dummies. That is the `dsmall` dummy collects the effect of the 3 smaller sizes offered by each firm and the `dsize` dummy differentiates the effect of the third smaller size offered which is 15 sq. picas for OG and 24 sq. picas

FIGURE 1. Histogram of Predicted Errors from Price Regressions



Notes: In this figure we show the histograms of predicted errors from Equation (1) for Verizon and Ogden, respectively.

In view of the theory model that lies ahead, we assume that $\tilde{T}_i$ is quadratic, which is rich enough to capture key data features (nonlinear pricing and quantity discounts). In other words, we estimate the following model:

$$\begin{aligned} T_{ij} &= \gamma_{0i} + \gamma_{1i}\mathtt{dsize}_{ij} + \gamma_{2i}\mathtt{dsmall}_{ij} + \alpha_{0i}(q_{ij} \times \mathtt{color}_{ij}) + \alpha_{1i}q_{ij} + \frac{\beta_i}{2}q_{ij}^2 + \varepsilon_{ij}, \\ &\equiv \gamma_i + \alpha_i q_{ij} + \frac{\beta_i}{2}q_{ij}^2 + \varepsilon_{ij}. \end{aligned} \tag{1}$$

By comparing the first and second line of equation (1) we see that the intercept (γ_i) is different for small size ads, for ads of size 24 sq. picas in VZ and 15 sq. picas in OG, and for large size ads. The coefficient on quantity (α) is also different if the ad is color or not. We present the least squares estimates in Table 2, and the quadratic price function provides a good fit to the data (see Figure 1).

Next, we take the payment T_{ij} for a non-multicolor ad q_{ij}, and define the quality-adjusted quantity to be the positive root of $T_{ij} - \hat{\gamma}_i - \hat{\alpha}_i\tilde{q} - \frac{\hat{\beta}_i}{2}\tilde{q}^2 = 0$, which is smaller than q_{ij}. To simplify notation, henceforth, we use q_{ij} to denote quality-adjusted quantity bought by consumer j from seller i; see Table (3).

for VZ. For our chosen specification, the roots of the quadratic polynomial were always positive and smaller than the corresponding non-adjusted quantity.

TABLE 2. Estimates of Tariff Function

	Verizon	Ogden
$\hat{\gamma}$	$85.48 - 136.36\texttt{dsize} - 68.46\texttt{dsmall}$	$303 + 100\texttt{dsize} - 223.24\texttt{dsmall}$
$\hat{\alpha}$	9.19	5.16
$\hat{\beta}$	-0.0010	-0.0034
R^2	0.99	0.96

Notes: Least squares estimates of (1). The dummies $\texttt{dsize}_{VZ} = 1$ when the ad is 24 sq. pica, $\texttt{dsize}_{OG} = 1$ if the ad is 15 sq. pica, and $\texttt{dsmall}_{VZ} = 1$ if the ad is smaller than 27 sq. pica while $\texttt{dsmall}_{OG} = 1$ if the ad is smaller than 24 sq. pica. All estimates are significant at 5%.

TABLE 3. Summary Statistic of Quality-Adjusted Quantity

	Min	25^{th} percentile	Median	Mean	Max
Verizon	5.18	17.38	31.09	98.47	6147
Ogden	7.99	8.74	15.00	123.27	1860

Notes: The table shows the key statistic for quality-adjusted quantity.

To get an idea of what quality-adjusted quantity means consider the following. A full page ad without any color in VZ measures 3,074 sq. pica, and its quality-adjusted quantity equivalent becomes 2,616 sq. pica, which means that for the same price of a black and white full page ad in VZ the consumer could have bought a smaller sized, 2,616 sq. pica, ad with full color. The decrease in size is the adjustment with respect to multicolor. Because the standard listings are free, this transformation does not affect them.

One of the features of nonlinear pricing is the quantity-discount. In Table 4 we present some evidence of this, but only for non-color ads. An ad that is 2.5% of a page costs \$10.84/sq. pica in VZ and \$10.65/sq. pica in OG, while the rate decreases by more than 43% in VZ to \$6.12 and by more than 67% in OG to \$3.42 for full page ads. Moreover, the quantity-discount is larger in OG than in VZ, which suggests stronger competition between the two for lower types than for higher type consumers.

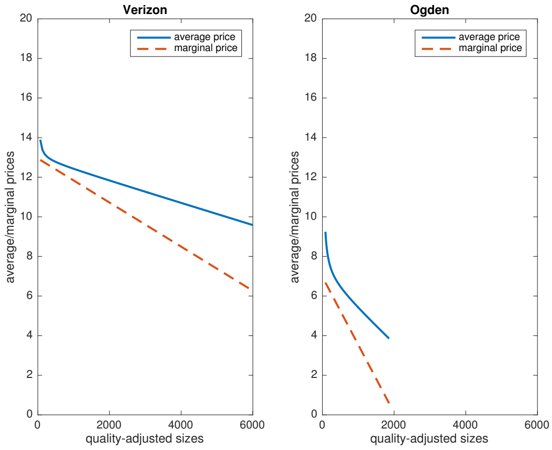
With quality-adjusted quantity, a better way to visualize this discount is to consider average and marginal prices for VZ and OG as shown in Figure 2. There are

TABLE 4. Quantity-Discounts

Ad Size	Verizon Price	Ogden Price
2.5	10.84	10.65
10	8.65	5.54
25	7.98	3.93
50	6.79	3.71
100	6.12	3.42

Notes: Ad size is expressed as a % of a full page non-colored ad. The other columns list the price, in $ per sq. pica for VZ and OG, respectively.

FIGURE 2. Average and Marginal Prices



Notes: The figure plots the average and marginal price that provides the best fit to the prices and ads, measured in terms of quality-adjusted quantity. The solid lines denote average prices and the dotted lines correspond to the marginal prices.

two salient features: a) marginal price and average price are different and decreasing which is evidence of quantity-discount – at the margin, larger ads cost less per size sq. pica; and b) OG not only offers smaller product line than VZ (Table 3), OG also gives larger quantity-discount than VZ. For our empirical analysis we propose and estimate a model that rationalizes these salient data features.[2]

We end with the following important observation (c.f. discussion in Section 3.1). This method of focusing on quadratic price function has some implications on the theory and the interpretation of the data. Although focusing on a simpler quadratic

[2] Conlon [2017] shows that in a market with a monopoly seller, as consumers become more heterogeneous, measured by the steepness of hazard function, the quantity-discount is steeper.

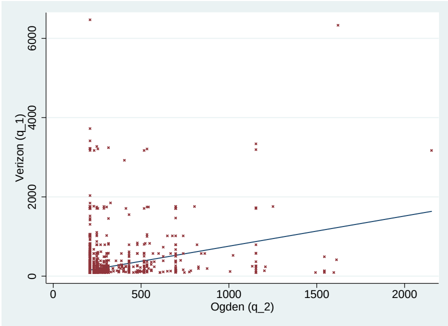
price function can be restrictive, especially if the sellers could improve their profits by choosing more complicated pricing functions, this assumption has two major benefits. First, as we discussed above it allows us to combine color and ad sizes into one measure. Second, focusing on simpler pricing function tremendously simplifies the theoretical model because common agency problems are difficult [Epstein and Peters, 1999; Martimort and Stole, 2002]. We revisit this point in the next section.

Two-Dimensional Preferences. In this section we will argue that to rationalize the choice data we must allow consumer (unobserved) preference heterogeneity to be at least two-dimensional. In other words, we must allow the intercept of marginal utility from VZ to be different than the intercept of the marginal utility from OG. Nonetheless, these intercepts or consumer "types" can be correlated.

To see why we need at least two-dimensional preference heterogeneity consider Figure 3 where we present a scatter plot of OG and VZ ads observed in the sample. If consumers' preferences could be indexed by a one-dimensional random variable then both sellers would necessarily use the same ordering of consumers: a consumer with high willingness to pay for VZ ad would also have high willingness to pay for OG ad, and vice versa. This, in turn, would mean that the observed ads bought would coalesce around an increasing line (one such line is given in the figure), and the correlation between VZ and OG ads would be high. But the ads are dispersed over the plane, and the correlation between VZ and OG ads is only 0.25.

In summary, this suggests that to capture the rich unobserved consumer heterogeneity we must allow a consumer to be high type for VZ but low type for OG, or vice versa. To that end, we model each consumer with a two-dimensional taste parameter, which is possibly correlated but only observed by the consumer. The Cramér-von Mises statistic for independence between the two sales is 1.66 (p-value ≈ 0), leading to a model of multidimensional screening with competition.

FIGURE 3. Scatter plot of Advertisement Choices



Notes: Scatterplot of ads purchased in the data. The x-axis denotes the OG ads, and the y-axis is the VZ ads. The line is a linear prediction of VZ given OG ads.

3. THE MODEL

Here we present a model of competing nonlinear pricing that builds on Ivaldi and Martimort [1994]. Let P_1 and P_2 be two sellers that stand for Verizon (VZ) and Ogden (OG), respectively. We assume that P_1 and P_2 have symmetric information and choose their menu simultaneously.[3]

Let $u(\mathbf{q};\theta,A)$ be the gross utility that a consumer of type $\theta \equiv (\theta_1,\theta_2)$ gets from choosing $\mathbf{q} = (q_1,q_2)$, where A is a vector of common utility parameters (defined shortly below). Let $T_i(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ be the pricing function chosen by P_i.

Assumption 1. *Let $U(\cdot;\theta)$ be the net utility when a $\theta-$type consumer chooses $\mathbf{q}$ such that*

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

The quasi-linearity is an important assumption to determine nonlinear pricing because it ensures that there is no income effect. The (net) marginal utility for the $\theta = (\theta_1,\theta_2)$ consumer from $q_i, i = 1,2$ is $MU_i = \theta_i - bq_i + cq_{-i} - T_i'(q_i)$. Therefore, θ_i denotes the intercept of the marginal utility from q_i, and so higher θ_i means a consumer's marginal utility from q_i is higher, so is the willingness to pay.

[3] [Busse and Rysman, 2005] show that competition affects the equilibrium only through the price schedules and not through quantities, so our estimation predominantly uses the allocation rule which is robust to sequential or simultaneous moves.

Assumption 2. (i) *Let* $A = \{b_1, b_2, c\}$ *and* $b_1 > 0, b_2 > 0, b_1 b_2 - c^2 > 0$ *and* $c \leq 0$.

(ii) $(\theta_1, \theta_2) \overset{i.i.d}{\sim} F(\cdot,\cdot)$, *with density* $f(\cdot,\cdot) > 0$ *on the support* $[\underline{\theta}_1, \bar{\theta}_1] \times [\underline{\theta}_2, \bar{\theta}_2]$.

(iii) *Cost function:* $C_i(q_i) = K_i + m_i \times q_i$ *with* $K_i \geq 0$ *and* $m_i > 0$ *for* $i = 1, 2$.

Assumption 2-(i) implies that the marginal utility of q_i decreases with q_i and the utility function is concave in q_1 and q_2. The quadratic utility also ensures that the marginal utility is linear in both θ_i and q_i. The assumption that $c \leq 0$ means that we assume the two ads q_1 and q_2 are weak substitutes. This assumption is driven by the data (Figure 3). Although all consumers have the same functional form for utility, they all have different (random) intercepts for marginal utility.

Assumption 2-(ii) implies that the consumers draw their private information θ independently and identically across all consumers. This rules out any correlation among consumers in terms of their valuation for ads. Implicitly, this assumption also means that consumers' values for ads are exogenously given and as a consequence the willingness to pay for an ad is independent of the composition of the ads placed by other consumers in the same business. Thus we follow the modeling assumptions used in the media economics literature [Anderson and Waldfogel, 2016; Berry and Waldfogel, 2016].

Assumption 2-(iii) implies that the production (printing) function exhibits constant returns to scale, that is characterized by a fixed cost K_i and marginal cost m_i for $i = 1, 2$. The fixed cost captures costs associated with the printing machine and the cost of distribution. The constant marginal cost is associated with the cost of printing such as ink, paper and labor.

3.1. **Price Functions .** Now we explain the sellers' problem of screening consumers. Publishers do not observe $\theta's$ but it is common knowledge among them that $\theta \overset{i.i.d}{\sim} F(\cdot)$, and that the consumers can see both menus before making their choices. We also assume that the cost parameters are common knowledge among them. Then they simultaneously choose menus that maximize their expected profits.

The revelation principle [Myerson, 1981] implies that, without loss of generality, we can restrict ourselves to only direct mechanisms. A direct mechanism consists of two functions, an allocation rule $q_i : [\underline{\theta}, \overline{\theta}] \to Q_i$ that specifies the quantity $q_i(\tilde{\theta})$ to a consumer who reports her type to be $\tilde{\theta}$, and a payment function $T_i : [\underline{\theta}, \overline{\theta}] \to \mathbb{R}$ that specifies the price $T_i(\tilde{\theta}) \equiv T_i(q_i(\tilde{\theta}))$ charged to such a consumer. For the seller P_i, the pair $\{q_i(\cdot), T_i(\cdot)\}$ is said to be feasible if they satisfy consumers' incentive compatibility and the participation constraints for each type, and maximize the expected profit of P_i given the choices of the other seller $\{q_{-i}(\cdot), T_{-i}(\cdot)\}$, where $i = 1, 2$.

However, the application of the revelation principle with competing sellers is not straightforward [see Epstein and Peters, 1999; Martimort and Stole, 2009]. Furthermore, without additional functional form assumptions it is hard to guarantee uniqueness of the equilibrium. In fact, full characterization of equilibria in a model of competing principal (also known as the problem of common agency) is an open question. One such functional form assumption can be on the price function $T_i(\cdot)$. In view of the discussion about quality-adjusted-quantity (1), we assume that

$$T_i(q_i) = \begin{cases} \gamma_i + \alpha_i q_i + \frac{\beta_i}{2} q_i^2 & \text{if } \ q_i > q_{i0} \\ 0 & \text{if } \ q_i \leq q_{i0}, \end{cases} \tag{2}$$

and $T_i(\cdot)$ is right differentiable at $q_{i0} > 0$, which is the free outside option (the standard listing) for $i = 1, 2$. This means that P_i's problem of choosing a function $T_i(\cdot)$ simplifies to choosing three parameters $\gamma_i > 0, \alpha_i > 0, \beta_i < 0$. This assumption narrows the space to search over. As a consequence we can apply the revelation principle, which in turn simplifies the model.

It is important to note that without any additional assumption(s) on $F(\theta_1, \theta_2)$ it is not necessarily the case that $P_2's$ best response to quadratic $T_1(\cdot)$ is also quadratic. Similarly, $P_1's$ best response to quadratic $T_2(\cdot)$ need not be quadratic. Therefore, our equilibrium concept is not a Nash equilibrium in the strict sense but rather a rationalizable equilibrium where both players believe (falsely) that their opponents are using quadratic pricing rules.[4] In this regard we closely follow Ivaldi and Martimort [1994] although they impose parametric assumption on the joint distribution.

[4] We want to thank an anonymous referee for making this important observation. The optimal allocation rules, however, are mutually best responses to each other. See Propositions 1 and 2.

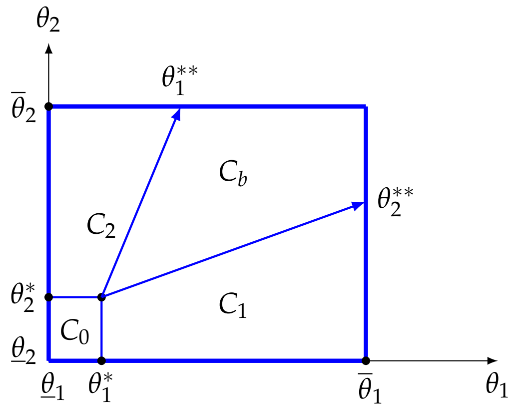
Although our choice of quadratic prices is motivated by our desire to reduce product's attributes and to simplify the equilibrium of competing pricing, there is some precedence in economics literature on using simpler strategy space for simplicity and tractability. Moreover, Wilson [1993], McAfee [2002] and Chu, Leslie, and Sorensen [2011] find that simple pricing strategies are often nearly optimal, suggesting that in our case the optimal price functions could be approximately quadratic, although there is no way to verify this claim. Similarly, there are papers that study the problem of approximating complex strategies using simpler alternatives. For example, Rogerson [2003] shows that in the standard principal-agent model, the principal can capture most of the gains from offering optimal continuous menu of contracts just by offering simpler (discrete) alternatives, and Carroll [2015] shows the robustness of linear contracts, albeit in a different environment than ours.

Similar idea of appromixating, Markov Perfect Equilibrium [Ericson and Pakes, 1995] by a simpler alternative, Oblivious Equilibrium [Weintraub, Benkard, and Roy, 2008] and Partially Oblivious Equilibrium [Benkard, Jeziorski, and Weintraub, 2015], has been fruitfully used in the empirical industrial organization literature. Many empirical applications necessarily make some modeling simplifications that help reduce complexity of the equilibrium. These simplifications range from discretizing the state space and/or to reducing the number of firms [Benkard, 2004; Gawrisankaran and Town, 1997; Collard-Wexler, 2013] to using functional approximation of the value functions [Sweeting, 2013; Fowlie, Reguant, and Ryan, 2016], among many others. There is also another line of research that introduces psychological/behavioral constraints into a standard industrial organization framework to explain patterns in the data better; see, for example, Ellison [2006] and Al-Najjar, Baliga, and Besanko [2008].

Next, we consider the incentive compatibility (IC) and participation (IR) constraints. Consider θ's first order conditions:

$$\begin{aligned}(\theta_1 - b_1 q_1 + c q_2 - T_1'(q_1))(q_1 - q_{10}) &= 0; \\ (\theta_2 - b_2 q_2 + c q_1 - T_2'(q_2))(q_2 - q_{20}) &= 0. \qquad (3)\end{aligned}$$

FIGURE 4. Consumer types and Choices



Notes: A schematic representation of the four subsets of consumer types. C_0 denotes consumer types that choose (q_{10}, q_{20}); C_1 and C_2 denote types that buy only from VZ and OG, respectively; and C_b denotes types that buy from both VZ and OG.

These conditions determine four subsets of consumers: C_0 which denotes consumers who choose the outside option (q_{10}, q_{20}); C_1 and C_2 which denote consumers who buy an ad from only P_1 or P_2, respectively; and C_b which denotes consumers who choose from both sellers (Figure 4).

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

$$\begin{aligned}\theta_1 - b_1 q_{10} + c q_{20} &\leq \alpha_1 + \beta_1 q_{10};\\ \theta_2 - b_2 q_{20} + c q_{10} &\leq \alpha_2 + \beta_2 q_{20}.\end{aligned}$$

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

$$\begin{aligned}\theta_1^* &= \alpha_1 + (b_1 + \beta_1) q_{10} - c q_{20};\\ \theta_2^* &= \alpha_2 + (b_2 + \beta_2) q_{20} - c q_{10}.\end{aligned}$$

So all consumers with type $(\theta_1, \theta_2) \ll (\theta_1^*, \theta_2^*)$ find it optimal to choose (q_{10}, q_{20}).

Case (2): C_1. Here consumers choose $q_1 > q_{10}$ and $q_2 = q_{20}$. The types must satisfy the following conditions:

$$\theta_1 - b_1 q_1 + c q_{20} = \alpha_1 + \beta_1 q_1;$$

$$\theta_2 - b_2 q_{20} + c q_1 \leq \alpha_2 + \beta_2 q_{20}.$$

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

$$\theta_2^{**} = \left(b_2 - \frac{c^2}{b_1 + \beta_1}\right) q_{20} + (\alpha_2 + \beta_2 q_{20}) + \frac{c\alpha_1}{b_1 + \beta_1} - \frac{c}{b_1 + \beta_1}\theta_1. \tag{4}$$

Case (3): C_2. This is the counterpart of C_1 and is determined in the same way. Let θ_1^{**} be the threshold type such that any type with $\theta_1 \leq \theta_1^{**}$ buys q_{10}, then:

$$\theta_1^{**} = \left(b_1 + \beta_1 - \frac{c^2}{b_2}\right) q_{10} + \alpha_1 + \frac{c}{b_2}(\alpha_2 + \beta_2 q_2) - \frac{c}{b_2}\theta_2.$$

Case (4): C_b. This corresponds to the "interior solution" where consumers choose positive amounts of both. This set is determined by the two first-order conditions (see equation 3) as follows,

$$\theta_1 - b_1 q_1 + c q_2 = \alpha_1 + \beta_1 q_1; \tag{5}$$

$$\theta_2 - b_2 q_2 + c q_1 = \alpha_2 + \beta_2 q_2. \tag{6}$$

Nonlinear Pricing. Note that, even though consumers have two dimensional types (θ_1, θ_2), each seller has only a one dimensional instrument q. This means that there will be bunching, i.e., more than one type of consumer will be allocated the same q. Then P_2's objective is to determine the most profitable way to bunch, while preserving the incentive compatibility across types that are not bunched. Next, using the consumer optimality conditions, we show that the two-dimensional screening problem can be transformed into a one-dimensional screening problem using a sufficient-statistic. Thus, P_2's problem becomes a canonical one-dimensional screening problem, with respect to this aggregated type, which is easier to solve.

For those consumers who buy $q_2 > q_{20}$, the corresponding q_1 is given by (5) as

$$q_1 = \begin{cases} \frac{\theta_1 - \alpha_1 + cq_2}{b_1+\beta_1}, & \theta_1 > \theta^* \\ q_{10}, & \theta_1 \le \theta_1^*. \end{cases} \tag{7}$$

As we mentioned in the previous section q_{10} is the square pica of the outside option that is available for free. Substituting (7) in (6) gives the necessary condition for q_2 to be optimal for the type (θ_1, θ_2) consumer, i.e.,

$$\theta_2 + \frac{c\theta_1}{b_1+\beta_1} = \frac{c\alpha_1}{b_1+\beta_1} + \left(b_2 - \frac{c^2}{b_1+\beta_1}\right) q_2 + (\alpha_2 + \beta_2 q_2). \tag{8}$$

Notice that consumer's type (θ_1, θ_2) appear only in the LHS of (8). So, given $T_1(\cdot)$, P_2 can take the linear combination of (θ_1, θ_2) as exogenous. Letting $z_2 := \theta_2 + \frac{c\theta_1}{b_1+\beta_1}$ we can re-write (8) in terms of a one-dimensional variable:

$$z_2 = \frac{c\alpha_1}{b_1+\beta_1} + \left(b_2 - \frac{c^2}{b_1+\beta_1}\right) q_2 + (\alpha_2 + \beta_2 q_2). \tag{9}$$

Therefore z_2 aggregates the consumer type (θ_1, θ_2) into one dimension from the point of view of P_2. This aggregation preserves some desirable properties. For instance, z_2 increases with θ_2, given $c \le 0$ it does not increase with θ_1, and fixing (θ_1, θ_2) it increases with β_2. The last part is consistent with the fact that if P_1 gives a lower discount (i.e., higher β_2) the willingness to pay for q_2 should increase.

This means that all consumers with the same z_2 will buy the same q_2 even though they might have different (θ_1, θ_2). Said differently, z_2 provides a way to optimally bunch consumers, so that a pricing mechanism that depends only on the aggregated value z_2 will do as good as a mechanism that depends on knowing θ_1 and θ_2 separately. Thus we can re-cast P_2's problem in terms of only $z_2 \in [\underline{z}_2, \overline{z}_2]$. Let $G_2(\cdot)$ be the distribution of z_2 with its density given by

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

Now, P_2's optimization problem can be written in terms of z_2 as

$$\max_{T_2(\cdot),q_2(\cdot),z_2^0} \left\{ \mathbb{E}\Pi_2 = \int_{z_2^0}^{\bar{z}_2} \Big(T_2\left(q_2(z_2)\right) - m_2 q_2(z_2) \Big) g_2(z_2)\mathrm{d}z_2 - K_2 - m_2 q_{20} G_2(z_2^0) \right\}, \quad (10)$$

subject to the appropriate IC and IR constraints (see below). The threshold z_2^0 corresponds to the types that choose q_{20}, i.e., the subsets C_1 and C_0 in Figure 4 are such that if $\theta_1 \le \theta_1^*$ then $z_2^0 = \theta_2^* + \frac{c\theta_1^*}{b_1+\beta_1}$ and if $\theta_1 \ge \theta_1^*$ then $z_2^0 = \theta_2^{**} + \frac{c\theta_1}{b_1+\beta_1}$.

The utility from q_2 depends on q_1, which in turn depends on q_2, (see (5) and (6)), meaning that determining the incentive compatibility constraints for P_2 needs some additional considerations. Let $W_2(\theta_1, z_2)$ be z_2's indirect utility from (q_1, q_2), i.e.,

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

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

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

And let $s_2(z_2)$ be such that

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

After some simplification one can see that

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

which means one can decompose the indirect utility from $(q_1, q_2(z_2))$ into the sum of the indirect utility from (q_1, q_{20}) and the additional utility from choosing $q_2(z_2) > q_{20}$. Since P_2 can only affect $s_2(\cdot)$ it is the only relevant "utility function" that P_2 cares about. Thus the IC constraints can be expressed as follows:

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

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

$$
\begin{aligned}
s_2'(z_2) &= q_2(z_2) - q_{20}, \qquad \forall z_2 \in (z_2^0, \bar{z}_2], &(11)\\
T(z_2) &= \left(z_2 - \frac{c\alpha_1 - c^2 q_2(z_2)}{b_1+\beta_1}\right)(q_2(z_2)-q_{20}) - \frac{b_2}{2}(q_2^2(z_2) - q_{20}^2) - s_2(z_2). &(12)
\end{aligned}
$$

From (11) and (12), we see that without loss of generality P_2 can be viewed as choosing $s_2(z_2)$ as the rent function and charging $T_2(\cdot)$. Rochet [1987] showed that the global IC constraint is satisfied if and only if: (i) $s_2(z_2) = \int_{z_2^0}^{z_2}(q_2(t) - q_{20})\mathrm{d}t + s^+, \forall z_2 \in [z_2^0, \bar{z}_2]$, where $s_2^+ \equiv \lim_{z_2 \downarrow z_2^0} s_2(z_2)$; and (ii) $s_2(\cdot)$ is increasing, or equivalently from (11), the allocation function is strictly increasing in z_2, i.e., $q_2'(z_2) > 0$.

Similarly, the participation or IR constraint becomes $W_2(\theta_1, z_2) = w_2(\theta_1, z_2) + s_2(z_2) \geq \max\{w_2(\theta_1, z_2), 0\}$, or equivalently $s_2(z_2) \geq 0$. Then substituting these in (10), P_2's problem becomes

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

subject to the following IC and IR constraints, respectively

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

To solve the problem, we follow the literature and ignore the second order IC constraint and verify ex-post that the solution satisfies the constraint. Since $s_2(\cdot)$ is increasing, $s_2(z_2^0) = 0$ implies $s_2(z_2) > 0$ (IR) for all $z_2 \in (z_2^0, \bar{z}_2]$. It is immediate to see that $s_2^+ = 0$ is optimal. This is a one-dimensional screening problem for which a unique equilibrium exists; see Rochet and Choné [1998].

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

$$q_2(z_2) = \frac{z_2 - \frac{1-G_2(z_2)}{g_2(z_2)} - m_2 - \frac{c^2 q_{20} + c\alpha_1}{b_1+\beta_1}}{b_2 - \frac{2c^2}{b_1+\beta_1}}, \forall z_2 \in (z_2^0, \bar{z}_2] \tag{13}$$

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

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

The proof is in Appendix B. Following the same steps as above, we can determine the optimal allocation function $q_1(\cdot)$, using $z_1 = \theta_1 + \frac{c\theta_2}{b_2+\beta_2} \sim g_1(\cdot)$.

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

$$q_1(z_1) = \begin{cases} \frac{z_1 - \frac{1-G_1(z_1)}{g_1(z_1)} - m_1 - \frac{c\alpha_2 + c^2 q_{10}}{b_2+\beta_2}}{b_1 - \frac{2c^2}{b_2+\beta_2}}, & z_1 \in (z_1^0, \bar{z}_1] \\ q_{10}, & z_1 \in [\underline{z}_1, z_1^0]. \end{cases} \tag{14}$$

4. IDENTIFICATION

In this section we study the identification of the parameters of the model, which include utility parameters $\{b_1, b_2, c\}$, cost parameters $\{m_1, m_2, K_1, K_2\}$, and the joint distribution of types $F(\cdot, \cdot)$. Our observables include the parameters from the price functions $\{\alpha_i, \beta_i, \gamma_i : i = 1, 2\}$, and advertisements $\{q_{1j}, q_{2j}\}$ placed by consumer $j = 1, \ldots, J$ with the two Yellow Pages directories.

The outcome of the firms' optimization does not depend on the fixed cost, so we cannot identify K_1 and K_2. Given the nonlinearity and non separability of our model and the fact that some consumers choose (q_{10}, q_{20}) we do not have information about the lowest type of consumers. We fix the lower bound of the support of θ_2 and normalize the utility of the lowest types as follows.

Assumption 3. *Normalization: Let* $\underline{\theta}_2 = 0$ *and* $u(q_{10}, q_{20}; \underline{\theta}_1, \underline{\theta}_2) = 0$.

Although we are interested in the joint density $f(\theta_1, \theta_2)$, we begin by considering the identification of $g(z_1, z_2)$ the joint density of (z_1, z_2) because our model treats the z_1 and z_2 as the primitives and the allocation rules are expressed in terms of z and not θ. Once we identify $g(\cdot, \cdot)$ we can identify $f(\cdot, \cdot)$.

Let $H(\cdot, \cdot)$ be the conditional joint distribution of (q_1, q_2) given $q_i > q_{i0}$, and let $H_i(q_i)$ be the corresponding marginal distribution of q_i given $q_i > q_{i0}$, for $i = 1, 2$. If we focus only on $[z_i^0, \overline{z}_i]$ where $q_i \geq q_{i0}$ then the incentive compatibility constraint implies that the equilibrium allocation rule $q_i(\cdot) : [z_i^0, \overline{z}] \mapsto [q_{i0}, \overline{q}_i]$ is monotonic, hence it can be inverted to provide a (inverse) mapping $\mathcal{Z}_i(\cdot) \equiv q_i^{-1}(\cdot)$, where $\overline{q}_i := \max_j q_{ij}$ is the largest ad sold by P_i. Then $z_{ij} = z_i(q_{ij})$ is the type that choses $q_{ij}(> q_{i0})$, so, if $z_i^0 := \mathcal{Z}_i(q_{i0})$,

$$\begin{aligned}
H_i(q) &:= \int_{q_{j0}}^{\overline{q}_j} H(q, \xi) d\xi = \Pr[q_i \leq q | q_i \geq q_{i0}] \\
&= \Pr[z_i \leq \mathcal{Z}_i(q) | z_i > \mathcal{Z}_i(q_{i0})] = \frac{G_i(z) - G_i(z_i^0)}{1 - G_i(z_i^0)}; \qquad (15) \\
h_i(q) = H_i'(q) &= \frac{g_i(z)}{1 - G_i(z_i^0)} \times \mathcal{Z}_i'(q). \qquad (16)
\end{aligned}$$

From (15) and (16) we get

$$\frac{1 - G_i(z_i)}{g_i(z_i)} = \frac{1 - H_i(q)}{h_i(q)} \mathcal{Z}_i'(q),$$

which together with (14) and (13) give

$$q_i = \frac{z_i - \frac{1 - H_i(q_i)}{h_i(q_i)} \mathcal{Z}_i'(q_i) - m_i - \frac{c\alpha_{-i} + c^2 q_{i0}}{b_{-i} + \beta_{-i}}}{b_i - \frac{2c^2}{b_{-i} + \beta_{-i}}},$$

which can be inverted to give

$$z_i = q_i \left(b_i - \frac{2c^2}{b_{-i} + \beta_{-i}} \right) + m_i + \frac{c\alpha_{-i} + c^2 q_{i0}}{b_{-i} + \beta_{-i}} + \frac{(1 - H_i(q_i))}{h_i(q_i)} \mathcal{Z}_i'(q_i), \qquad (17)$$

thereby identifying z_i conditional on identifying $\{\mathcal{Z}_i'(\cdot), m_i, b_1, b_2, c\}$. Identification of $\mathcal{Z}_i'(\cdot)$ follows from differentiating (12) twice, i.e., for $i = 1, 2$

$$\mathcal{Z}_i'(q_i) = T_i''(q_i) + b_i - \frac{2c^2}{(b_{-i} + \beta_{-i})}. \tag{18}$$

Therefore, for those in C_b, (14) and (13) can be inverted to identify $G(\cdot, \cdot | z_1 \geq z_1^0, z_2 \geq z_2^0)$ from

$$\begin{pmatrix} z_{1j} \\ z_{2j} \end{pmatrix} = \begin{pmatrix} q_1^{-1}(q_{1j}) \\ q_2^{-1}(q_{2j}) \end{pmatrix} = \begin{pmatrix} q_{1j}\left(b_1 - \frac{2c^2}{b_2+\beta_2}\right) + m_1 + \frac{c\alpha_2 + c^2 q_{10}}{b_2+\beta_2} + \frac{(1-H_1(q_{1j}))}{h_1(q_{1j})}\mathcal{Z}_1'(q_{1j}) \\ q_{2j}\left(b_2 - \frac{2c^2}{b_1+\beta_1}\right) + m_2 + \frac{c\alpha_1 + c^2 q_{20}}{b_1+\beta_1} + \frac{(1-H_2(q_{2j}))}{h_2(q_{2j})}\mathcal{Z}_2'(q_{2j}) \end{pmatrix}, \tag{19}$$

with $i = 1, 2; j = 1, \ldots, 6328$. Henceforth, we use (z_1, z_2), to mean one of these combinations: $(z_1, z_2), (z_1, z_2^0), (z_1^0, z_2)$ and (z_1^0, z_2^0), depending on whether (z_1, z_2) are in C_b, C_1, C_2 and C_0 (see Figure 4), respectively.

If every consumer had bought ads with both directories then we would observe joint choices (q_{1j}, q_{2j}) for each consumer j, which identifies (z_{1j}, z_{2j}), and (z_{1j}, z_{2j}) with $\{c, \beta_1, \beta_2\}$ would identify $(\theta_{1j}, \theta_{2j})$. However, less than 20% of consumers buy from both directories. Most consumer buy from only one directory, and even among them, most buy from VZ. In view of that we invert one choice at a time to identify the marginal densities of of z_1 and z_2 separately, and use a parametric Copula to estimate a joint density.

There are many Copulas to choose from, but the theory of nonlinear pricing is silent about the correlation between θ_1 and θ_2. So, instead of using an ad-hoc method to choose one family we non-nested model selection and goodness-of-fit tests to select the Copula that provides the best fit to the data. In particular, we use both the Cramer-von-Misses goodness-of-fit test and the Vuong [1989] test to choose a parametric Copula, from 10 most widely used Copulas. We further elaborate on this process in subsection 5.2.

Cost Parameters. To identify cost parameters we use the properties of nonlinear pricing: it is never optimal for the seller to distort the output meant for the highest type, even though the outputs for other types are distorted. This property is referred to as "no-distortion-on-top" and it helps to identify marginal costs. The idea is simple and intuitive: "no-distortion-on-top" implies that the quantity offered to the highest type maximizes social welfare. This means that the marginal benefit from $\overline{q}_i$ is equal to the marginal cost m_i which in turn is equal to marginal prices at $\overline{q}_i$, i.e., $T_i'(\overline{q}_i) = \alpha_i + \beta_i\overline{q}_i, i = 1,2$. We can identify $\overline{q}_i := \max_j\{q_{ij}\}, i = 1,2$, and we can estimate $T_i'(\max_j q_{ij})$ for $i = 1,2$, therefore we can identify $\{m_1, m_2\}$.

Utility Parameters. To identify the utility parameters we make use of the curvature of the utility function and focus on consumers with extreme choices, i.e., those who choose large q_1 but q_{20} and vice versa. Concavity of the utility function implies that the parameters b_1, b_2 and c govern the "love-for-variety," penalizing extreme choices. In particular, larger b_1 and b_2 push consumers to advertise with both Yellow pages. If some choose unequal q_1 and q_2, say, then it means small b_1 and b_2.

For further intuition, suppose $b_1 = b_2 = 0$, so the utility function is linear and the consumers care only about the total quantity of ads $(q_1 + q_2)$, but not the composition. Suppose $q_2 = 0$. Now, if we increase $b_1 > 0$, the marginal utility from q_1 falls and q_2 starts to become important as consumers start choosing $q_2 > 0$. So, the more concave the utility the lower the likelihood of observing unequal q_1 and q_2 choices.

This constraint is the most binding, and therefore the most informative, for the highest type $\overline{z}_i$ with the skewed (the largest q_i and smallest q_{-i}) ad choices. The value of b_1 must be small enough to rationalize these choices. The formal argument is slightly involved because the highest types and c are unknown.

Because for $i = 1,2$, $\overline{z}_i$ has an interior solution, the choice $\overline{q}_i$ must equate marginal utility and marginal price, i.e., $\overline{\theta}_i - b_i\overline{q}_i + cq_{j0} = \alpha_i + \beta_i\overline{q}_i \Rightarrow \overline{\theta}_i = \alpha_i + (b_i + \beta_i)\overline{q}_i - cq_{j0}, \quad j \neq i$. This identifies $\overline{\theta}_1$ conditional on identifying b_1 and c and $\overline{\theta}_2$ conditional on identifying b_2 and c. Evaluating $q_2(z_2)$ in (13) at $\overline{z}_2$ gives

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

From the normalization assumption we get $\underline{\theta}_2 = 0$ and

$$\underline{\theta}_1 = \frac{b_1}{2} q_{10} + \frac{b_2}{2}\frac{q_{20}^2}{q_{10}} - cq_{20},$$

which, when substituted in the previous equation for $\overline{z}_2 = \overline{\theta}_2 + \frac{c\underline{\theta}_1}{b_1+\beta_1}$, gives

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

$$\alpha_2 + (b_2 + \beta_2)\overline{q}_2 - cq_{10} = \overline{q}_2 b_2 + m_2 + \frac{-2c^2(\overline{q}_2 - q_{20}) + c(\alpha_1 - \frac{b_1}{2} q_{10} + \frac{b_2}{2}\frac{q_{20}^2}{q_{10}})}{b_1 + \beta_1}$$

$$2c^2(\overline{q}_2 - q_{20}) - c(\alpha_1 - \frac{b_1}{2} q_{10} + \frac{b_2}{2}\frac{q_{20}^2}{q_{10}} + (b_1 + \beta_1) q_{10}) - (\alpha_2 + \beta_2\overline{q}_2 - m_2)(b_1 + \beta_1) = 0,$$

identifying c as the negative root of the quadratic equation.

Next, to identify b_1 and b_2, for any $q_i < \overline{q}_i$ we rewrite the optimal allocation rule

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

so that for any two $q_i \neq \tilde{q}_i$ we get

$$b_i + \beta_i = \frac{\alpha_i + \beta_i q_i - m_i}{\frac{1-H_i(q_i)}{h_i(q_i)}} + \frac{2c^2}{b_{-i} + \beta_{-i}}; \quad b_i + \beta_i = \frac{\alpha_i + \beta_i \tilde{q}_i - m_i}{\frac{1-H_i(\tilde{q}_i)}{h_i(\tilde{q}_i)}} + \frac{2c^2}{b_{-i} + \beta_{-i}}.$$

Equating these two equations identifies $b_i, i = 1, 2$ as

$$b_i = \frac{1}{2}\left(\frac{\alpha_i + \beta_i q_i - m_i}{\frac{1-H_i(q_i)}{h_i(q_i)}} + \frac{\alpha_i + \beta_i \tilde{q}_i - m_i}{\frac{1-H_i(\tilde{q}_i)}{h_i(\tilde{q}_i)}}\right) - \beta_i.$$

Thus, we first identify $(\overline{q}_1, \overline{q}_2, m_1, m_2)$ and (b_1, b_2), then c and, finally $(\overline{\theta}_2, \overline{\theta}_1, \underline{\theta}_1)$.

5. Estimation

We observe (q_{1j}, q_{2j}) for consumers $j = 1, 2, ..., J = 6823$. The optimal allocation functions define the econometric model for $i = 1, 2$ and $j = 1, \ldots, J$,

$$q_{ij} = \begin{cases} [z_{ij} - \frac{1-G_i(z_{ij})}{g_i(z_{ij})} - m_i - \frac{c\alpha_{-i} + c^2 q_{i0}}{b_{-i}+\beta_{-i}}] \Big/ [b_i - \frac{2c^2}{b_{-i}+\beta_{-i}}], & z_{ij} \in (z_i^0, \overline{z}_i] \\ q_{i0}, & z_{ij} \in [\underline{z}_i, z_i^0]. \end{cases} \tag{20}$$

Let N_1^* and N_2^* be the number of consumers whose ads are strictly larger than q_{10} and q_{20}, respectively. In (20) z_{ij} plays the role of the "error" in the usual regression models. However, the model also depends on $\frac{1-G_i(\cdot)}{g_i(\cdot)}$, which is unknown, but from the identification arguments we know it can be replaced with $\frac{1-H_i(q_{ij})}{h_i(q_{ij})}\mathcal{Z}_i'(q_{ij})$, where $\mathcal{Z}_i'(q_{ij})$ is identified from equation (18). As $\frac{1-H_i(q_{ij})}{h_i(q_{ij})}$ can be estimated nonparametrically (described below), we can replace it with its estimate $\frac{1-\hat{H}_i(q_{ij})}{\hat{h}_i(q_{ij})}$. Besides these two equations, one for each seller, we have the following vector of conditions for estimation:

$$s(\psi) := \begin{pmatrix} m_1 - \alpha_1 + \beta_1\overline{q}_1 \\ m_2 - \alpha_2 - \beta_2\overline{q}_2 \\ 2(b_1+\beta_1) - \frac{\alpha_1+\beta_1 q_1 - m_1}{\frac{1-H_1(q_1)}{h_1(q_1)}} - \frac{\alpha_1+\beta_1\tilde{q}_1 - m_1}{\frac{1-H_1(\tilde{q}_1)}{h_1(\tilde{q}_1)}} - \beta_1 \\ \overline{q}_2 c q_{10} - \alpha_2 - (b_2+\beta_2)\overline{q}_2 + \overline{\theta}_2 \\ \overline{\theta}_1 - \overline{q}_1(b_1-\beta_1) - \alpha_1 + cq_{20} \\ (b_2+\beta_2)\overline{q}_2 - cq_{10} - \overline{\theta}_1 + \alpha_2 \\ (\overline{\theta}_2 - \overline{q}_2 b_2 - m_2)(b_1+\beta_1) - c^2 q_{20} - c\alpha_1 + c\underline{\theta}_1 + 2c^2\overline{q}_2 \\ (\underline{\theta}_1 + cq_{20})2q_{10} - b_1 q_{10}^2 - b_2 q_{20}^2 \end{pmatrix} = 0 \tag{21}$$

Therefore the estimation procedure consists of the following two steps: (1) Estimate the inverse hazard function $(1-H_i(\cdot))/(h_i(\cdot))$ for $i = 1,2$ using a local polynomial estimator; (2) plug-in these estimates in (20) and (21), and estimate the parameters using the nonlinear least squares method.

5.1. **Distributions and Densities of q_1 and q_2.** We use local polynomial estimators (LPE) [Fan and Gijbels, 1996] to estimate the marginal distributions and densities of observed ad choices $\{H_i(\cdot), h_i(\cdot) : i = 1,2\}$ nonparametrically. We choose LPE over the widely used Parzen-Rosenblatt Kernel estimator (KDE) because: (i) KDE 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; Devrôye, 1997].

LPE is also suitable to estimate data with thin tails, which is particularly useful for our data. LPE is an intuitive method based on the local polynomial technique so it easy to implement and does not require pre-binning or any other transformation of the data, while still being fully boundary adaptive and automatic. In particular, LPE is a weighted least squares estimator where we weight a polynomial function, we use a polynomial of order two, locally by a Kernel. The basic estimation steps are given in Appendix C. For more on LPE see Cattaneo, Jansson, and Ma [2017a].

5.2. **Estimating the Joint Distribution of** (θ_1, θ_2) **.** Once we have estimates $\hat{z}_{ij}$ we can again use LPE to estimate the truncated marginal densities. We are also interested in the joint distribution $F(\cdot,\cdot)$ of (θ_1, θ_2), for which we need the joint distribution $G(z_1, z_2)$, and so far we only have estimates for the two (truncated) marginals. The consumers who buy the standard listing do not provide any information about either their $z's$ or the dependence between their z_1 and z_2. We know that the observed choices (q_1, q_2) are correlated, although the correlation is not high, and given that the two goods are (weak) substitutes the only way to rationalize that correlation is through the correlation between θ_1 and θ_2. However, without further parametric assumptions, we can only identify the Fréchet-Hoeffding bounds.

Under an additional assumption on the tail dependence, however, we can use the theory of empirical Copula to estimate the joint distribution. From Sklar's Theorem [Nelson, 1999] we know that there is a unique Copula $C : [0,1]^2 \to [0,1]$ such that

$$G(z_1, z_2) = C(z_1, z_2) := C(G_1(z_1), G_2(z_2)).$$

Thus estimating $G(\cdot,\cdot)$ is the same as estimating $C(\cdot,\cdot)$. To that end we consider parametric a Copula, $C(\cdot,\cdot;\kappa)$ so that the Copula is known up to the parameter $\kappa \in \Gamma \subset \mathbb{R}$, where Γ is the set of parameter values. This parameter κ governs the dependence between z_1 and z_2. We use the subsample of consumers who buy ads from both VZ and OG to estimate this dependence, and we use a testing procedure to choose the right family of Copula, i.e., whether it is a Gaussian Copula, or t-Copula or Frank Copula, etcetera.

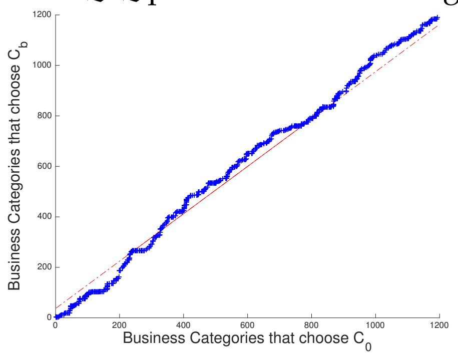
To identify the dependence we make the assumption that the correlation between (θ_1, θ_2) among those consumers who buy ads in both directories is equal to the correlation between (θ_1, θ_2) among those who choose free listings in both directories. Under this assumption we can use the dependence at the upper tail to extrapolate the dependence at the lower tail. Although restrictive, this tail dependence assumption shows the limit of identification without having access to rich data on consumer covariates and without making additional assumptions about exogeneity of those covariates and the functional form of how those covariates affect (z_1, z_2).

Although we cannot control for possible selection because of our limited data, we know the business category (e.g., doctor, hair dresser, plumber) a consumer belongs to. With this limited information we can at least check whether some businesses, say doctors, are more or less likely to advertise with both VZ and OG while others, say, hair dressers, choose not to advertise at all.

To that end, we begin with the relevant subsample of consumers –those who only buy the free standard listing or those who advertise in both directories– and enumerate each business category with a unique number. In total there are 1,200 different business categories in this subsample. Then we use the Wilconxon-Mann-Whitney rank-sum test to test the null hypothesis that it is equally likely that a randomly selected business type will choose free ads or will choose ads from both directories. We estimate the test statistic to be 0.5420, and using Bootstrapped p-values we cannot reject the null of equality. In Figure 5 we present the quantile-quantile plot of the distribution of business categories. Although both these results are only suggestive evidences of a lack of selection, lacking any other data, we interpret these as consistent with our tail-dependence assumption.

The second problem is the selection of the family of Copula. Once we know the Copula family we can use pseudo MLE to estimate the parameter κ. To choose the Copula family we implement the following procedure: (i) We begin with 10 of the most widely used families of Copulas; (ii) we pick two families and implement the

FIGURE 5. Q-Q plot of Business Categories.



Notes: This is a quantile-quantile plot of business categories. Those who only buy the outside option (the free standard listing) on the x-axis and those who choose $(q_1 > q_{10}, q_2 > q_{20})$, are on the y-axis. Each number on the axes denotes a unique business category.

non-nested model selection test of Vuong [1989] and Cramér-von Mises goodness-of-fit tests of Fermanian, Radulović, and Wegkamp [2004] and Genest, Rémillard, and Beaudoin [2009];[5] (iii) we repeat step (ii) for all pairs of Copulas and count the number of times a family is selected according to each of the two criteria; and (iv) we choose the Copula that is selected the maximum number of times according to both criteria. Once the family is selected we estimate the parameter κ using pseudo MLE; see Genest, Ghoudi, and Rivest [1995] and Genest, Quessy, and Rémillard [2006].

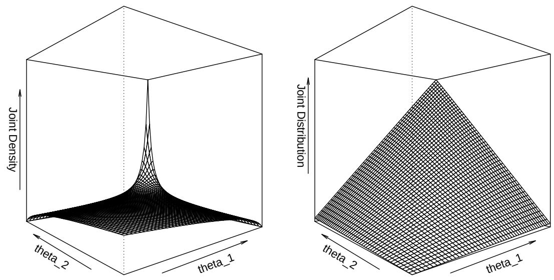
Based on 10,000 Bootstrap simulations for each comparison we choose the Joe Copula as the best family, and estimate $\hat{\kappa} = 1.214$ with a bootstrapped standard error of (0.013). The Joe Copula is an example of an Archimedean Copula [see Nelson, 1999]. From the estimate of (z_1, z_2) we determine (θ_1, θ_2) and its joint CDF

$$\hat{F}(\theta_1, \theta_2) = 1 - \left[(1 - \hat{G}_1(z_1(\theta)))^{\hat{\kappa}} + (1 - \hat{G}_2(z_2(\theta)))^{\hat{\kappa}} - (1 - \hat{G}_1(z_1(\theta)))^{\hat{\kappa}} (1 - \hat{G}_2(z_2(\theta)))^{\hat{\kappa}} \right]^{\frac{1}{\hat{\kappa}}}.$$

In Figure 6 we present the joint distribution and its corresponding joint density by simulating draws from them.

[5] The asymptotic distribution of the goodness-of-fit test statistic is not distribution-free [Genest and Rémillard, 2004], and because we use the estimates $(\hat{z}_1, \hat{z}_2)$, to compute the critical values for goodness-of-fit tests and the Vuong-test we use the bootstrap procedures from Genest and Rémillard [2004]; Genest and Rémillard [2008]; Kojadinovic and Holmes [2009]; Kojadinovic and Yan [2011] and in Clarke [2007], respectively. We use the R package called copula by Kojadinovic and Yan [2017].

FIGURE 6. Estimated Joint PDF and CDF of (θ_1, θ_2)



5.3. **Estimation Results.** Table 5 presents the results of our estimation procedure as well as the lower and upper ends of the (bootstrapped) confidence interval for each estimator. The estimated gross utility function becomes

$$\hat{u}(q_1, q_2, \theta_1, \theta_2) = \theta_1 q_1 - \frac{0.09}{2} q_1^2 + \theta_2 q_2 - \frac{0.17}{2} q_2^2 - 0.0003 \times q_1 \times q_2.$$

The estimated parameter $\hat{c}$ is negative, $\hat{c} = -0.0003 < 0$, which means the two directories can be treated as substitutes, although the rate of substitution is weak. The marginal cost of printing for VZ is $\hat{m}_1 = 9.7$ which is more than twice the one for OG, $\hat{m}_2 = 3.5$, capturing the differences in paper size and quality. The support of $\hat{F}(\cdot, \cdot)$ is estimated to be $[1.28, \;\; 576.1] \times [0, \;\; 319]$ reflecting that the consumers prefer VZ to OG.

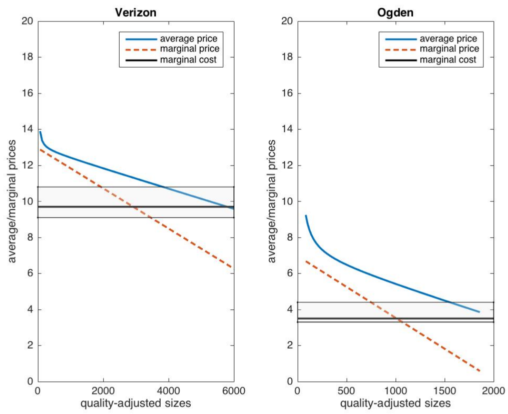
Armstrong [1996] showed that with multidimensional private information it is always optimal for the seller to price the goods in such a way that some positive fraction of consumers are not served. The threshold type z_i^0 then depends on the

TABLE 5. Estimated Parameters

Parameters	Estimates	95% Confidence Interval
c	-0.0003	$[-0.0004, -0.000002]$
b_1	0.09	$[0.07, 0.12]$
b_2	0.17	$[0.1, 0.2]$
m_1	9.7	$[9.1, 10.8]$
m_2	3.5	$[3.3, 4.4]$

Notes: Estimates of utility and cost parameters. The first column is the list of parameters, the second column is the estimates and the third column is the Bootstrapped 95% confidence interval of the estimates.

FIGURE 7. Average Prices, Marginal Prices and Marginal Costs



Notes: The figure plots the average and marginal price that provides the best fit to the prices and ads, measured in terms of quality-adjusted quantity along with the estimated constant marginal costs. The shaded regions are the 95% CI.

density of consumer types, 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. In this case, competition between VZ and OG at the lower end is stronger than at the upper end.

We see this pattern reflected in our data: the difference in average price per pica widens as we move from lower category to higher, see Table 4. The fact that VZ's prices are consistently higher across comparable categories than of OG's suggests that VZ enjoys a higher brand effect. In Figure 7 we present average and marginal prices from Figure 2, but this time with the estimated marginal costs and their confidence interval. The figure gives a sense of markups for both VZ and OG.

5.4. **Cost of Asymmetric Information.** In this section we answer the following question. What is the welfare cost of asymmetric information between the sellers and the buyers? Suppose both sellers can observe (θ_1, θ_2) and offer socially optimal quantities that equate marginal utility and marginal costs. The difference in welfare under this counterfactual from that in the data is a measure of the cost of asymmetric information.

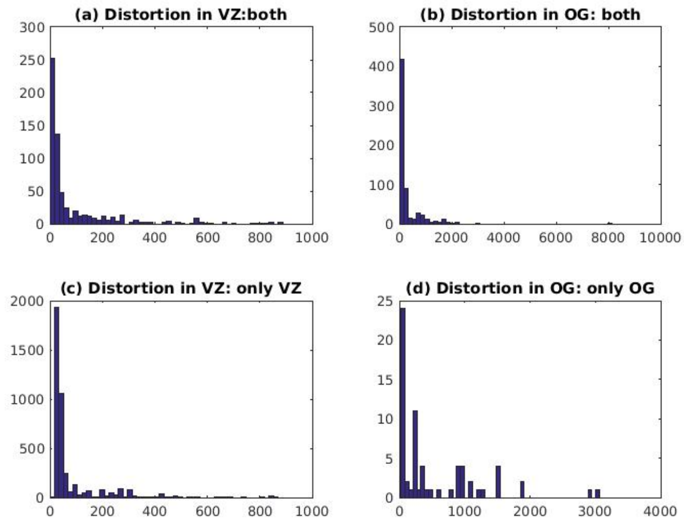
We begin with the subset of consumers in C_b. For each of these consumers we know their type (θ_1, θ_2) and for each of those types we determine quantities (q_1, q_2) that (simultaneously) equate marginal utilities with marginal costs. Figure 8 presents the histogram of distortion in quantity, which is defined to be the difference between socially optimal quantities and what is observed in the data. As we can see, among those who buy positive amount from both publishers, the distortion is more pronounced in the allocation of VZ, Figure 8-(a), than of OG, Figure 8-(b). We also consider the consumer who buy only from VZ/OG, and determine their socially optimal VZ/OG, while fixing $q_2 = q_{20}$ or $q_1 = q_{10}$. The histogram of this distortion is Figure 8-(c) and (d). Together these results show that asymmetric information leads to distortion but also has different effects on different types of consumers.

We also calculate the changes in welfare due to asymmetric information for each of these three types of consumers separately.[6] The corresponding picture is presented in Figure 9. Note the dispersion in the welfare cost of asymmetric information, and how that varies across different groups of consumers. Figure 9 also illustrates heterogeneous effects of asymmetric information among consumers. Aggregating the total effect, we find that without asymmetric information, the welfare would have increased by 20% of the total consumer surplus.

5.5. **Competition and Welfare.** Next, we consider the role of competition and asymmetric information on consumer surplus and producer surplus. For that we have to solve for nonlinear pricing under monopoly, i.e., when VZ and OG merge and maximize their joint profit, and compare the outcomes with the data. To the best

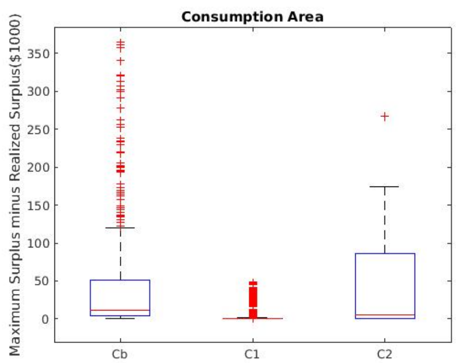
[6] Quasi-linear preferences imply that the welfare calculation does not depend on the prices.

FIGURE 8. Distortion in Quantity



Notes: The figure shows histograms of the difference in maximum surplus (under first-best quantity) and the surplus in the data. amount of distortions in quantities, where distortion is defined as the difference between the socially optimal quantities and what is observed in the data. Figure (a) and (b) correspond to consumers who buy from both. Figure (c) corresponds to the distortion in VZ for consumers who buy only from VZ, and (d) corresponds to those who buy only from OG.

FIGURE 9. Difference in Total Surplus



Notes: This is a box-plot for the difference between maximum surplus (under first-best quantity) and the surplus in the data, expressed in $1000. C_b, C_1 and C_2 refer to the sets in Figure 4.

of our knowledge, this is the first attempt to quantify the effect of competition on welfare with multidimensional asymmetric information. Solving optimal nonlinear pricing in a multiproduct monopoly setting is, however, a difficult problem, [Armstrong, 1996; Rochet and Choné, 1998; Basov, 2005]. Even finding a numerical solution is hard and is limited (θ_1, θ_2) being uniformly random variables; see Ekeland and Moreno-Bromberg [2010].

In view of these difficulties, we consider a slightly simpler, but more robust (in the max-min sense) problem studied by Carroll [2017], namely, we determine the optimal nonlinear pricing scheme that maximizes the worst-case expected profit for the multiproduct monopolist. In particular, we solve the following problem:

$$\max_{\{q(\theta_1,\theta_2),T(q(\theta_1,\theta_2))\}} \left\{ \min_{\{\tilde{F}(\cdot)\in\mathcal{F}\}} \left\{ \int_{\Theta} [T(q(\theta_1,\theta_2)) - \sum_{i=1}^{2} (m_i \times q_i(\theta_1,\theta_2)] d\tilde{F}(\theta_1,\theta_2) \right\} \right\}, \tag{22}$$

where $\mathcal{F}$ is the set of all regular joint distributions of (θ_1, θ_2) with the estimated marginal distributions $\hat{F}_1(\theta_1)$ and $\hat{F}_2(\theta_2)$, subject to the usual individual rationality and incentive compatibility constraints. The solution provides a worst-case bound for the monopoly's profit and upper bound on consumer surplus.

Before we proceed, two considerations are in order. First is the problem of nonzero cross price elasticity between VZ and OG. We know from Figalli, Kim, and McCann [2011] that the optimization problem (22) need not be convex. Given that our estimate $\hat{c}$ is very close to zero (see Table 5), we assume zero cross-price elasticity by setting $c = 0$ for this exercise. Second is the problem of the outside option. In the data the listings are free, which lose revenue for the sellers. So for merger we assume that the monopoly will no longer offer the free listing option. This exercise provides a lower bound on producer surplus and upper bound on consumer surplus.

Carroll [2017] shows that the maximization problem (22) can be decomposed into two parts: using $\hat{F}_1(\theta)$ to solve for monopoly nonlinear pricing for VZ and $\hat{F}_2(\cdot)$ to solve for monopoly nonlinear pricing for OG separately. In other words, we can replace the $\min\{\cdot\}$ and the joint CDF $\tilde{F}(\cdot,\cdot)$ with the marginals. So (22) simplifies to a separable problem

$$\max_{\{q_1(\cdot),q_2(\cdot),T_1(\cdot),T_2(\cdot)\}} \left\{ \iint_{\Theta} \left[\sum_{i=1}^{2} T_i(q_i(\theta_i)) - \hat{m}_i q_i(\theta_i) \right] \hat{f}_1(\theta_1)\hat{f}_2(\theta_2) d\theta_1 d\theta_2 \right\}$$

subject to the respective IR and IC constraints. We can easily solve for each dimension separately [Jullien, 2000] and determine the optimal allocation and price functions for $i \in \{1, 2\}$ to be, respectively,

$$q_i(\theta_i) = \frac{1}{b_i}\left(\theta_i - \frac{1-\hat{F}_i(\theta_i)}{\hat{f}_i(\theta_i)} - \hat{m}_i\right); \tag{23}$$

$$T_i(\theta_i) = \theta_i q - \int_{\underline{\theta}}^{\theta} q_i(t)dt - \frac{\hat{b}_1}{2}\left(q_i(\theta)\right)^2. \tag{24}$$

The profit from selling VZ and OG separately, but optimally, provides the lowest bound on the joint monopoly profit, and given that the utility function is additively separable in payment, it provides an upper bound on the consumer surplus. Given how few consumers choose OG over VZ, another interesting counterfactual would be to consider a situation where after the merge, the monopoly does not offer the OG directory. Even though this will hurt consumers with high θ_2, because the probability of such consumer is so low, it might have a smaller over-all effect on consumers. Thus it would be interesting to determine the welfare with and without OG and compare it to the observed welfare.

Given the separability of the problem we also consider the other plausible counterfactual scenario where after merger the firm shuts down the least profitable product. In our case VZ is clearly more profitable than OG, so this allows us to determine welfare when only VZ is sold. We can calculate the welfare only from VZ by setting $q_2 = 0$ and $T_2 = 0$ for all consumers.

The steps involved for this empirical exercise are as follows. Using the estimated parameters, we determine the optimal nonlinear pricing for VZ and OG from (23) and (24). Then, for each (θ_1, θ_2) identified from the data we calculate the consumer surplus (CS) under duopoly, and under monopoly. To determine the CS when only VZ is available, we repeat the second step but with q_2 and T_2 equal to zero for all consumers. To determine producer surplus, we simply use the transfer from the solution and subtract the total variable cost for each consumer. We are interested

TABLE 6. Welfare under Merger Simulations

		Duopoly	Merger (both)	Merger (only VZ)
Product Space	$[\underline{q}_1, \ \overline{q}_1]$	$[5.19, \ 6,147]$	$[0, \ 6,124]$	$[0, \ 6124]$
	$[\underline{q}_2, \ \overline{q}_2]$	$[7.18, \ 1,860]$	$[0, \ 4,494]$	
Consumer Surplus	$25p$	207.3	0	0
	Median	782	7.54	3.64
	$75p$	1,694	47.8	47
	$90p - 10p$	4,146	7,138	3,638
Prodcuer Surplus	$25p$	−78	0	0
	Median	53	33	16.6
	$75p$	219.9	261	261
	$90p - 10p$	818	24,063	4063
	Total (in MM)	1.72	13.3	9.39

Notes: This table presents CS and PS estimated from the data (column 1), merger when both VZ and OG are offered (column 2) and when only VZ is offered (column 3). Here p stands for percentile, so $10p$ is the 10^{th} percentile, and so on.

in the distribution of CS and PS, so we compare and contrast merger with duopoly data and summarize our findings in Table 6.

In the first column we present the estimates under duopoly (data), in the second column we present results from a merger when both directories are offered, and in the last column we present results when only VZ is offered. The first effect we see is that under monopoly the product space of OG increases. The size of the biggest ad increases from 1,860 sq. picas to 4,494 sq. picas, whereas the maximum size for VZ does not change. This is consistent with the fact that under competition the largest advertisement option with OG was a one page ad, whereas in VZ it was a two-page ad, so under monopoly the firm will offer a two-page ad. The smallest ad decreases to zero, because the (free) outside option is not offered under the merger.

Next consider the distribution of CS. When we move from duopoly to merger, the percentiles of CS shrink. For instance under duopoly the 25^{th} percentile of CS is 207.3 whereas under the merger it decreases to zero, because under monopoly many lower-type consumers are excluded as it is not always optimal to serve every type of consumer. The CS for high-types increases, the difference between 90^{th} percentile and 10^{th} percentile increases, which is primarily due to the fact that the monopolist offers larger OG ads. Interestingly although the total consumer surplus under monopoly is 17% lower than under duopoly, the inequality in the distribution of CS (measured by the *Gini* coefficients) do not change.

Consider now the distribution of PS, in Table 6. Under duopoly, the PS of the 25^{th} percentile is equal to $-\$78$ because of the free listing option, whereas after the merger this number is zero which is consistent with exclusion of at least 25% of consumers. When both directories are offered, the total PS increases from \$1.72 million to \$13.3 million. But when only VZ is offered, PS decreases by 29 percentage points (from \$13.3 to \$9.39 million), and the CS decreases by 24%, suggesting that it is socially inefficient to only offer VZ (the last row of Table 6).

6. Conclusion

In this paper we estimate a model of competitive nonlinear pricing using a novel dataset on advertisements placed with two Yellow Pages directories. We use a model of competing principals, where the two sellers use nonlinear prices to attract consumers who are heterogenous with respect to their value for the ads in the two directories. We show that the model parameters can be identified from first-order conditions that characterize the equilibrium allocations and consumption patterns.

Our estimates can rationalize why prices set by two publishers are similar for smaller ads but diverge for larger ads. That is to say we can rationalize the observation that the two sellers compete more strongly for the lower end of the market than for the upper end, which would not be possible under a linear pricing scheme. Our estimates are consistent with heterogenous preferences and asymmetric information in the Yellow Pages ad market. In a counterfactual exercise, we find that the welfare cost of asymmetric information is approximately 20% of the consumer surplus.

Using a merger simulation we find that under monopoly product space increases, producer surplus increases, and consumer surplus for low types decreases whereas the consumer surplus for high type increases. When aggregated we find that under a merger the total consumer surplus falls and producer surplus increases, but the distribution of consumer surplus does not change.

APPENDIX A. TABLES

				No color		One color		White Background		White Backg. + 1 color		Multicolor	
VZ Sq. Picas	VZ %	OG Sq. Picas	OG %	VZ Price	OG Price	VZ Price	OG Price	VZ Price	OG Price (n.a.)	VZ Price	OG Price	VZ Price	OG Price
Listing													
		9	0.48		$0								
12	0.40	12	0.65	$0	$134		$147						
		15	0.81		$240		$278						
18	0.60			$151									
27	0.89			$290		$492							
36	1.19			$492		$845							
Space Listing													
		46	2.47		$490		$528						
54	1.79			$504									
72	2.38	92	4.95	$781	$587		$650						
108	3.58			$1,134		$1,789				$2,873			
		138	7.42		$1,008		$1,096						
144	4.77			$1,436		$2,243		$2,293					
		150	8.06		$1,243								
216	7.15			$2,079				$2,239		$3,326			
		276	14.84		$1,741								
Display													
174	5.76			$1,638						$2,609		$2,873	
208	6.90			$1,915		$2,861						$3,326	
		226	12.15		$1,358		$1,638				$1,868		$1,902
355	11.77			$3,074		$4,612						$5,381	
		468	25.16		$2,202		$2,734				$3,135		$3,173
564	18.66			$4,474		$6,703				$7,145		$8,190	
		607	32.63								$3,054		$3,568
762	25.24			$5,872						$9,388		$10,256	
		923	49.62		$3,372								$7,176
1137	37.66			$8,719								$14,579	
		1235	66.40				$3,612				$4,660		$5,324
1485	49.18			$10,093								$17,640	
		1860	100.00		$6,324								$9,675
3020	100.00			$18,510		$27,770				$29,610		$32,395	
6066	200.89											$60,380	

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

APPENDIX B. PROOF OF PROPOSITION 1

Proof. Since the proof is standard in the literature [Stole, 2007] we will highlight only the main steps. The first step is to show that $\mathbb{E}\Pi_2$ is concave in q_2, and super modular in (q_2, z_2). Let I be the integrand of the expected profit function. Then,

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

$$\frac{\partial^2 I}{\partial q_2^2} = -(b_2 - \frac{2c^2}{b_1+\beta_1}) g_2(z_2),$$

$$\frac{\partial^2 I}{\partial q_2 \partial z_2} = \left(\left(1 - \frac{\partial}{\partial z_2}\frac{1-G(z_2)}{g(z_2)}\right) - \left(b_2 - \frac{2c^2}{b_1+\beta_1}\right) q_2'(\cdot)\right) g(z_2) = 0.$$

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. $\frac{\partial^2 I}{\partial q_2 \partial z_2} \geq 0$. The optimal allocation q_2 can be determined by simple point-wise maximization of I:

$$\frac{c^2}{b_1+\beta_1}(q_2-q_{20})+(z_2-\frac{c\alpha_1-c^2q_2}{b_1+\beta_1})-b_2q_2-\frac{1-G_2(z_2)}{g_2(z_2)}-m_2=0,$$

which yields:

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

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

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

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

APPENDIX C. LOCAL POLYNOMIAL ESTIMATION

We present the steps to estimate the distributions and densities of ad choices, $\{\hat{H}_i(q),\hat{h}_i(q)\}$, and consumer types $\{\hat{G}_i(z),\hat{g}_i(z)\}$ for $i=1,2$. The steps are the same for $i=1,2$ and for ad choices and consumer types, so we suppress the index.

(1) Using the observed choices $\{q_j : j=1,\ldots,N^*\}$ define the empirical CDF as $\tilde{H}(q)=\frac{1}{N^*}\sum_{j=1}^{N^*}\mathbb{1}(q_j\leq q)$, for every $q\in[\min_j q_j,\max_j q_j]$.

(2) For every q in the range solve the following (weighted) least squares problem

$$\min_{\{\lambda_0,\lambda_1,\lambda_2\}}\sum_{j=1}^{N^*}\left(\tilde{H}(q_j)-\underbrace{\left\{\lambda_0+(q_j-q)\lambda_1-(q_j-q)^2\lambda_2\right\}}_{\texttt{polynomial of order 2}}\right)^2\underbrace{K\left(\frac{q_j-q}{h}\right)}_{\texttt{weight}},$$

where $K(\cdot/h)$ is the Epanechnikov kernel function and h is the optimal bandwidth that minimizes the mean squared error for each grid point.

(3) The estimated constant λ_0 is the LPE estimator of the CDF, i.e., $\hat{H}(q)=\hat{\lambda}_0(q)$, while the estimate of the first-derivative of the polynomial λ_1 is the LPE estimator of PDF, i.e., $\hat{h}(q)=\hat{\lambda}_1(q)$.

To choose the optimal bandwidth, and to implement these estimation steps we use the R package called lpdensity by Cattaneo, Jansson, and Ma [2017b].

The estimates of the marginal distributions and densities satisfy consistency properties. In particular, application of widely available consistency properties of non-parametric estimators, such as those in [Guerre, Perrigne, and Vuong, 2000] yield the following results.

Lemma 1. *Let $\hat{\overline{q}}_i$ be the sample estimate of the largest size of ads offered by publisher i, and $\hat{z}_i$ be the plug-in estimator for consumer type for $i = 1, 2$ defined in (19). Suppose all the assumptions mentioned so far are valid. Then:*

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

For the proof we suggest the reader consult Guerre, Perrigne, and Vuong [2000]. For completeness we point out that Lemma 1-(1) and (2) follow from their Proposition 2; wheras Lemma 1-(3) and (4) follow from their Propositions 1 and 3, respectively. Finally we note that, in a similar way as described above, to estimate the distributions and densities of consumer type we can replace q with $\hat{z}$.

References

AL-NAJJAR, N., S. BALIGA, AND D. BESANKO (2008): "Market Forces meet Behavioral Biases: Cost Misallocation and Irrational Pricing," *RAND Journal of Economics*, 39(1), 214–317. 19

ANDERSON, S. P., AND J. WALDFOGEL (2016): "Preference Externalities in Media Markets," in *Handbook of Media Economics*, ed. by S. P. Anderson, J. Waldfogel, and D. Stromberg, vol. 1A, chap. 1. Elsevier. 17

ARMSTRONG, M. (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, 64, 51–75. 4, 34, 38

——— (1999): "Price Discriminiation by Many-Product Firm," *Review of Economic Studies*, 66(1), 151–168. 4

——— (2006): "Recent Developments in the Economics of Price Discrimination," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by R. Blundell, W. K. Newey, and T. Persson, vol. 2, pp. 97–141. New York: Cambridge University Press. 5

ARMSTRONG, M., AND J. VICKERS (2001): "Competitive Price Discrimination," *The RAND Journal of Ecoomics*, 32, 579–605. 5

ARYAL, G. (2017): "Identifying a Model of Screening with Multidimensional Consumer Heterogeneity," *Working Paper*. 3, 7

ARYAL, G., I. PERRIGNE, AND Q. H. VUONG (2016): "Identification of Insurance Models with Multidimensional Screening," *Working Paper*. 7

BAKOS, Y., AND E. BRYNJOLFSSON (1999): "Bundling Information Goods: Pricing, Profits, and Efficiency," *Management Science*, 45(12), 1613–1630. 5

BASOV, S. (2005): *Multidimensional Screening*. Springer-Verlag, Berlin. 5, 38

BENKARD, C. L. (2004): "A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft," *Review of Economic Studies*, 71(3), 581–611. 19

BENKARD, C. L., P. JEZIORSKI, AND G. Y. WEINTRAUB (2015): "Oblivious Equilibrium for Concentrated Industries," *RAND Journal of Economics*, 46(4), 671–708. 19

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890. 6

BERRY, S. T. (1994): "Estimation of Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25(2), 242–262. 6

BERRY, S. T., AND J. WALDFOGEL (2016): "Empirical Modeling for Economics of the Media: Consumer and Advertiser Demand, Firm Supply and Firm Entry Models for Media Markets," in *Handbook of Media Economics*, ed. by S. P. Anderson, J. Waldfogel, and D. Stromberg, vol. 1A, chap. 3. Elsevier. 17

BUSSE, M., AND M. RYSMAN (2005): "Competition and Price Discrimination in Yellow Pages Advertising," *The RAND Journal of Econmics*, 36, 378–390. 6, 16

CARROLL, G. (2015): "Robustness and Linear Contracts," *American Economic Review*, 105(2), 365–563. 19

——— (2017): "Robustness and Separation in Multidimensional Screening," *Econometrica*, 85(2), 453–488. 5, 38

CATTANEO, M. D., M. JANSSON, AND X. MA (2017a): "Simple Local Polynomial Density Estimators," *Mimeo*. 31

——— (2017b): "lpdensity: Local Polynomial Density Estimation and Inference," *Mimeo*. 44

CHU, S., P. LESLIE, AND A. SORENSEN (2011): "Bundle-Size Pricing as an Approximation to Mixed Bundling," *American Economic Review*, 101(1), 263–303. 19

CLARKE, K. A. (2007): "A Simple Distribution-Free Test for Nonnested Model Selection," *Politica; Analysis*, 15(3), 347–363. 33

COHEN, A. (2008): "Package Size And Price Discrimination In The Paper Towel Market," *International Journal of Industrial Organization*, 26(2), 502–516. 6

COLLARD-WEXLER, A. (2013): "Demand Fluctuation in the Ready-Mix Concrete Industry," *Econometrica*, 81(3), 1003–1037. 19

CONLON, J. R. (2017): "Does Buyer Heterogeneity Steepen or Flatten Quantity Discounts?," *RAND Journal of Economics*, 48(4), 1027–1043. 14

CORTS, K. S. (1998): "Third-Degree Price Discrimination in Oligopoly: All-Out Competition and Strategic Commitment," *RAND Journal of Economics*, 29, 306–323. 5

CRAWFORD, G., AND M. SHUM (2006): "Monopoly Quality Degradation in Cable Television," *Journal of Law and Economics*, 49, 181–209. 6

DEVRÔYE, L. (1997): "Universal Smoothing Factor Selection in Density Estimation: Theory and Practice," *Test*, 6, 223–320. 30

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. 6

EINAV, L., A. FINKELSTEIN, S. RYAN, P. SCHRIMPF, AND M. CULLEN (2012): "Selection on Moral Hazard in Health Insurance," *American Economic Reveiw*, Forthcoming. 6

EINAV, L., M. JENKIS, AND J. LEVIN (2012): "Contract Pricing in Consumer Credit Markets," *Econometrica*, 80(4), 1387–1432. 6

EKELAND, I., J. HECKMAN, AND L. NESHEIM (2002): "Identifying Hedonic Models," *American Economic Reveiw*, 92(2), 304–309. 7

——— (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112, 60–109. 7

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

ELLISON, G. (2006): "Bounded Rationality in Industrial Organization," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by R. Blundell, W. K. Newey, and T. Persson, vol. 2, pp. 142–174. New York: Cambridge University Press. 19

EPPLE, D., R. ROMANO, AND H. SIEG (2006): "Admission Tuition, And Financial Aid Policies in the Market for Higher Education," *Econometrica*, 74(4), 885–928. 7

EPSTEIN, L. G., AND M. PETERS (1999): "A Revelation Principle for Competing Mechanisms," *Journal of Economic Theory*, 88, 119–160. 5, 15, 18

ERICSON, R., AND A. PAKES (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62, 53–82. 19

FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. Chapman & Hall. 30

FAN, Y. (2013): "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market," *American Economic Reveiw*, 103(5), 1598–1628. 6

FERMANIAN, J. D., D. RADULOVIĆ, AND M. H. WEGKAMP (2004): "Weak Convergence of Empirical Copula Processes," *Bernoulli*, 95, 119–152. 33

FIGALLI, A., Y.-H. KIM, AND R. MCCANN (2011): "When is Multidimensional Screening a Convex PRogram?," *Journal of Economic Theory*, 146, 454–478. 38

FOWLIE, M., M. REGUANT, AND S. P. RYAN (2016): "Market-Based Emissions Regulation and Industry Dynamics," *Journal of Political Economy*, 124(1), 249–302. 19

FU, C. (2014): "Equilibrium Tuition, Application, Admissions, and Enrollment in the College Market," *Journal of Political Economy*, 122(2), 225–281. 7

GAWRISANKARAN, G., AND R. TOWN (1997): "Dynamic Equilibrium in the Hospital Industry," *Journal of Economics and Management Strategy*, 6(1), 816–846. 19

GENEST, C., K. GHOUDI, AND L.-P. RIVEST (1995): "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrica*, 82(3), 543–552. 33

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. 33

GENEST, C., AND B. RÉMILLARD (2004): "Tests for Independence and Randomness Based on Empirical Copula Process," *Test*, 13(2), 335–369. 33

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. 33

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. 33

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–574. 7, 44

HECKMAN, J., R. L. MATZKIN, AND L. NESHEIM (2010): "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78(5), 1569–1591. 7

IVALDI, M., AND D. MARTIMORT (1994): "Competition under Nonlinear Pricing," *Annales d'Économie et de Statistique*, 34, 71–114. 2, 3, 5, 16, 18

JONES, M. C., J. S. MARRON, AND S. J. SHEATHER (1996): "A brief survey of bandwitdh selection for density estimation," *Journal of Americal Statistical Association*, 91, 401–407. 30

JULLIEN, B. (2000): "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, 93('), 1–47. 39

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. 33

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. 33

KOJADINOVIC, I., AND J. YAN (2017): "Modeling Multivaraite Distributions with Continuous Margins Using the Copula **R** Package," *Journal of Statistical Software*, 34(9). 33

LESLIE, P. J. (2004): "Price Discrimination in Broadway Theatre," *The RAND Journal of Economics*, 35, 520–541. 6

LUO, Y., I. PERRIGNE, AND Q. VUONG (2017): "Structural Analysis of Nonlinear Pricing," *Forthcoming Journal of Political Economy*. 6, 7

MARRON, J. S., AND M. P. WAND (1992): "Exact Mean Integrated Error," *The Annals of Statistics*, 10, 712–736. 30

MARTIMORT, D., AND L. STOLE (2009): "Market Participation in Delegated and Intrinsic Common-Agency Games," *RAND Journal of Economics*, 40(1), 78–102. 2, 5, 18

MARTIMORT, D., AND L. A. STOLE (2002): "The Revelation and Delegation Principles in Common Agency Games," *Econometrica*, 70(4), 1659–1673. 5, 15

MASKIN, E., AND J. RILEY (1984): "Monopoly with Incomplete Information," *The RAND Journal of Ecoomics*, 15, 171–196. 2

MCAFEE, P. (2002): "Coarse Matching," *Econometrica*, 70(5), 2025–2034. 19

MCAFEE, P., J. MCMILLAN, AND M. D. WHINSTON (1989): "Multiproduct Monopoly, Commodity Bundling, and Correlation of Values," *Quarterly Journal of Economics*, 104(2), 371–383. 5

MCMANUS, B. (2006): "Nonlinear Pricing in an Oligopoly Market: The Case of Speciality Coffee," *The RAND Journal of Ecoomics*. 6

MIRAVETE, E. J., AND L.-H. RÖLLER (2004): "Estimating Price-Cost Markups Under Nonlinear Pricing Competition," *Journal of the European Economic Association*, 2, 526–535. 2, 5

MUSSA, M., AND S. ROSEN (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317. 2

MYERSON, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73. 17

NELSON, R. B. (1999): *An Introduction to Copulas*. Springer-Verlag New York, Inc. 31, 33

NEVO, A. (2000): "A Practitioner's Guide To Estimation of Random-Coefficients Logit Models of Demand," *Journal of Economics and Management Strategy*, 9(4), 513–548. 6

NEVO, A., J. TURNER, AND J. W. WILLIAMS (2016): "UsageBased Pricing and Demand for Residential Broadband," *Econometrica*, 84(2), 411–443. 6

OREN, S. S., S. A. SMITH, AND R. WILSON (1983): "Competitive Nonlinear Tariffs," *Journal of Economic Theory*, 29(1), 49–71. 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. 24

ROCHET, J.-C., AND P. CHONÉ (1998): "Ironing, Sweeping, and Multidimensional Screeing," *Econometrica*, 66(4), 783–826. 4, 24, 38

ROCHET, J.-C., AND L. A. STOLE (2002): "Nonlinear Pricing with Random Participation Constraints," *Review of Economic Studies*, 69, 277–311. 2, 5

ROGERSON, W. P. (2003): "Simple Menus of Contracts in Cost-Based Procurement and Regulation," *American Economic Review*, 93(3), 919–926. 19

RYSMAN, M. (2004): "Competition Between Networks: A Study of the Market for Yellow Pages," *Review of Economic Studies*, 71, 483–512. 6

STOLE, L. A. (2007): "Price Discrimination in Competitive Environment," in *The Handbook of Industrial Organization*, ed. by M. Armstrong, and P. R. H., vol. III. Elsivier. 5, 42

SWEETING, A. (2013): "Dynamic Product Positioning in Differentiated Product Markets: The Effect of Fees for Musical Performance Rights on the Commercial Radio Industry," *Econometrica*, 81(5), 1763–1803. 19

TERRELL, G. R., AND D. W. SCOTT (1992): "Variable Kernel Density Estimation," *The Annals of Statistics*, 20, 1236–1265. 30

TIROLE, J. (1998): *The Theory of Industrial Organization*. MIT Press. 5

TRAIN, K. E. (2009): *Discrete Choice Methods with Simulations*. Cambridge University Press. 6

VUONG, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57(2), 307–333. 4, 27, 33

WEINTRAUB, G. Y., C. L. BENKARD, AND B. V. ROY (2008): "Markov Perfect Industry Dynamics with Many Firms," *Econometrica*, 76(6), 1375–1411. 19

WILSON, R. (1993): *Nonlinear Pricing*. Oxford University Press. 2, 5, 19

ZHOU, J. (2017): "Competitive Bundling," *Econometrica*, 85, 145–172. 5