ASOCIACION ARGENTINA
DE ECONOMIA POLITICA

ANALES | ASOCIACION ARGENTINA DE ECONOMIA POLITICA

LII Reunión Anual

Noviembre de 2017

ISSN 1852-0022
ISBN 978-987-28590-5-3

An empirical analysis of competitive nonlinear pricing

Aryal, Gaurab
Gabrielli, María Florencia

AN EMPIRICAL ANALYSIS OF COMPETITIVE NONLINEAR PRICING*

GAURAB ARYAL† AND MARIA F. GABRIELLI‡

ABSTRACT. We estimate a model of duopoly competition with nonlinear pricing where consumers have private information about their two-dimensional taste for the two differentiated products. We apply the model to a novel data on advertisements placed with two Yellow Pages directories that operate in Central Pennsylvania. Using both supply and demand side optimality conditions, we identify and estimate the distribution of unobserved consumer preferences, marginal costs and common utility parameters. We find substantial unobserved heterogeneity among consumers, and estimate a significant welfare loss due to asymmetric information.
Keywords: Nonlinear Pricing, Competition, Identification, Advertisement, Copula.
JEL classification: C14, D22, D82, L11, L13.

1. INTRODUCTION

Second degree price discrimination is widely used by profit maximizing seller(s) in many markets; see Wilson [1993]. Nonlinear pricing is an example of second degree price discrimination where the tariff is not strictly proportional to the quantity purchased. It is widely used by sellers with market power to increase profits when faced with heterogeneous consumers, for instance in advertisements, and in cable television, electricity to name a few. Under nonlinear pricing, menus of quantities and corresponding charges are offered, and each consumer chooses her best preferred quantity and pays the associated tariff. It affects both the allocative efficiency and total welfare [Mussa and Rosen, 1978; Maskin and Riley, 1984; Rochet and Stole,

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

2002], but the exact extents are ambiguous because they depend on the primitives of the market and the market structure. And because most of these papers focus exclusively on monopoly markets, we know very little about how these extents change with competition. In this paper we estimate a model of competitive nonlinear pricing, and contribute to the empirical literature that uses the *principal-agent framework* to study market outcomes under competition such as Ivaldi and Martimort [1994] and Miravete and Röller [2004].

The effect of competition on welfare depends crucially on the exact nature of the product. On the one hand if consumers choose only one product, competition can improve the consumer welfare because consumers now have better outside options. On the other hand if consumers can buy from multiple sellers consumers might be worse off. Furthermore, as we discuss in more detail below, to model competition we might have to allow consumers to have multidimensional taste parameters, then competition can affect different consumers differently.[1] In particular competition for consumers with "low" tastes might be more severe than competition for consumers with "high" tastes, leading to heterogenous effects on consumers. Ultimately, which forces dominate are all open empirical questions.

In this paper we use the *competing principal-agent framework* to propose a model of duopoly nonlinear pricing, and estimate it using a novel data on advertisements placed in Yellow Page Directories. Yellow pages advertisement provides an interesting setting to study these problems. Our market is simple and is characterized by competitive nonlinear pricing, where consumers buy ads on both directories. To rationalize the data we propose a rich and yet tractable two-dimensional unobserved consumer heterogeneity. Thus we estimate a multidimensional screening with competition, which is known to be a difficult problem. We rely on and develop further the model in Ivaldi and Martimort [1994] and estimate the model using a novel and unique data on Yellow Page advertisements in Central Pennsylvania.

[1] To motivate multidimensional taste, Rochet and Stole [2002] write: ``..because in a common setting of screening contracts -- that of nonlinear pricing -- an empirically reasonable model of consumer preferences arguably requires at least two dimensions of type uncertainty and a more general modeling of the participation decision.''

Advertisement is one of the markets where nonlinear pricing is widely used. Consumers can "self-select" from a menu of options, and often (like in our data) some choose to advertise in multiple outlets, while others do not advertise at all. We posit that there is asymmetric information between the two Yellow Page directories and the consumers (local businesses) about the latter's willingness to pay for advertisement. Our model characterizes the "optimal" nonlinear pricing when the two publishers of the Yellow Page directories commonly know the joint distribution of preference and their publishing costs.

Once the menu is offered, consumers choose the best option(s), and pay the corresponding tariff. We then ask: What must the parameters of the model be for the observed data to have been generated by this model? In other words we explore the problem of identification of our model. Our data is unique because it contains *all* adverts bought by *all* consumers in the market, and thus we can back-out the complete list of all options sold by both publishers. To do that we manually read all advertisements in both Yellow Pages and recorded them as choices, and used information about prices from the National Yellow Pages association to determine the payments. That said, one weakness of our data is that we do not observe any other *observed* consumer covariates. But despite that we use the supply side optimality conditions for identification. Thus we show that we can rely on theory of competitive principal-agent when rich data on consumer covariates is unavailable.

Under the assumption that the data is the outcome of the model, we first determine data features, and any additional conditions, that are sufficient to identify the joint density of consumer preferences, the marginal cost of printing and common utility parameters. Some consumers choose the outside option (the free entry option where your name is recorded in the books for free), so we can nonparmetrically identify only the truncated (marginal) distribution of each type, where the truncation is the highest type who chooses the free option. We also provide conditions to identify the marginal costs and common utility parameters. To combine the two marginals, we estimate a parametric Joe copula.[2]

Using the estimates and counterfactual analysis we find that: (a) there is substantial consumer heterogeneity; (b) competition is stronger for smaller adverts; (c) consumers treat the two directories as (weak) substitutes; and (d) the welfare cost of asymmetric information is $\approx 3.8\%$ of sales revenue.

Related Work. There is a rich literature on price discrimination that is directly or indirectlt related to the empirical analysis of competitive nonlinear pricing that we consider here. In the remainder we highlight the key references here, and discuss other similarities throught the paper.

We refer the reader to Wilson [1993] and Tirole [1998] for a detailed description on the theory of nonlinear pricing and price discrimination. Although the majority of the papers focus on monopoly seller, there some theory papers that allow competition; see, for instance, Oren, Smith, and Wilson [1983]; Epstein and Peters [1999]; Armstrong and Vickers [2001]; Rochet and Stole [2002]; Armstrong [2006]; Martimort and Stole [2002]; and Stole [2007]. But [Ivaldi and Martimort, 1994] and Miravete and Röller [2004] are the only two important exceptions that estimate a model of price discrimination with competition.

There is also a burgeoning literature [Crawford and Shum, 2006; Einav, Finkelstein, and Cullen, 2010; Einav, Finkelstein, Ryan, Schrimpf, and Cullen, 2012]; and Einav, Jenkis, and Levin [2012] that estimates the welfare effect of asymmetric information. But all of these papers only consider a monopoly seller with one dimensional (either discrete or continuous) private information. Thus our paper complements this line of inquiry by considering competition with richer unobserved consumer preference heterogeneity.

Since we estimate consumer preferences or "demand" from their choices, our paper is also related to the literature on demand estimation pioneered by Berry [1994] and Berry, Levinsohn, and Pakes [1995]. Also see Nevo [2000] and Train [2009].

[2] We use Cramér-von Mises test and Vuong [1989] test to select a copula from seven copulas.

However, our method relies on having continuous choices while these papers consider discrete choices. Another distinction stems from the way we model consumer heterogeneity. Our primary goal is to understand the effect of asymmetric information between the sellers and the buyers, and as such we allow for a rich (multidimensional) consumer heterogeneity that is unobserved to either the sellers or the researcher. Using the theory of competing principal-agent model we can not only allow consumers to choose from more than one seller but also endogenize the observed product space. The other literature however allows discrete options and a rich market structure but only under the assumption of exogenous product space.[3] Another difference between our method is that we need individual level data from consumers but data from only one market suffices while for the discrete choice method only market share suffices but as a result data from multiple markets is needed.

This is also the reason why the so called "BLP instruments" are inapplicable in our setting. A benefit of using a rich consumer heterogeneity is that it allows us to rationalize why competition appears stiff for smaller advertisements than for larger advertisement. Moreover, if sellers are using nonlinear pricing, the average prices of advertisement and the marginal prices are different, and because consumers respond to marginal prices, using average prices (as in the discrete choice) is incorrect.

Lastly, because we "lean on" the supply side to make up for the fact that we do not observe any consumer characteristics, the discrete choice literature relies heavily on exogenous and rich variation in the consumer characteristics. Thus our method can be viewed as an alternative method to estimate demand when only the individual level choice data is available albeit for only one market and one year.

Another paper that is related to our paper is Nevo, Turner, and Williams [2016]. However, unlike us, their method is based on researchers having access to a very rich set of data on consumer covariates. In that regard our papers can be viewed as

[3] Fan [2013] extends BLP model to allow endogenous product characteristics, and uses geographic variation in competition for identification and is thus inapplicable in our setting.

methods that are on the opposite ends of a spectrum that can be used to estimate demand with rich (unobserved) consumer heterogeneity.

In terms of the application, Rysman [2004] has studied the market for Yellow Page advertisements. While we focus on asymmetric information and (exogenous) consumer heterogeneity Rysman [2004] takes the two-sided market with complete information approach. In another paper Busse and Rysman [2005] estimate the relationship between competition and prices in the Yellow Page industry. Our model delivers the basic feature found in Busse and Rysman [2005] that the competition is severe at the lower quantity and for higher quantity the prices diverge.

Finally, in terms of our identification strategy, because we rely on "inverting first-order condition(s)," our paper is also related to the identification strategy in Guerre, Perrigne, and Vuong [2000].[4] Since then the method has also been applied to price discrimination problems by Luo, Perrigne, and Vuong [2017]; Aryal [2016] and Aryal, Perrigne, and Vuong [2016], among others. Lastly, our paper is also related to the hedonic models of Ekeland, Heckman, and Nesheim [2002, 2004]; Heckman, Matzkin, and Nesheim [2010], and Chernozhukov, Galichon, Hallin, and Henry [2015].

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

The data contains information about print advertisement placed by local businesses with two Yellow Pages directories in central Pennsylvania (State College and Bellefonte), U.S. for the year 2006. The directories are published by Verizon (henceforth, VZ), a utility provider, and Ogden (henceforth, OG). These directories differ in paper quality and their sizes, but they cover the same geographic area. Each publisher offers a large number of adverts to choose from. A consumer is a business

[4] [Leslie, 2004; McManus, 2006; Cohen, 2008] focus only on data where the consumer buys from at most one seller, and use only the demand side for estimation. Their identification strategy relies on exclusion restriction and exogenous variation in consumer covariates, which is infeasible here.

unit (say a doctor) who has a phone registered under a clinic. We have information about the advert(s) bought by all consumers, and the corresponding tariff(s). We observe the advert(s) placed by all consumers and the price(s) for the ad(s).

Data Sources. There are two key components of the data: information about the adverts and the tariffs. The information on the ads bought, their sizes and colors, were manually read and recorded from the two directories by us. It is a norm in the market to publish the names and addresses of all business phone numbers in the directory for free. This standard listing provides us the name of every consumer in the market. We also search and match the names with any additional advert(s) bought. The price data is gathered from the Yellow Page Association –an umbrella organization of Yellow Page publishers in the U.S. As members of this association, VZ and OG provide the association with their prices for all offered options. Once we have the advertisement rate and the advert, we can construct the total tariff(s) paid by each consumer. Throughout the paper we maintain the hypothesis that the tariffs thus constructed are the actual payment made out-of pocket prices.

Differentiated Directories. The standard unit of measurement in this industry is called pica, which is approximately $1/6$ of an inch. The free ad where the directories publish the name and phone number of a business unit is called the standard listing, and is of different size in VZ and OG. In VZ a standard listing is 12 square picas and in OG it is 9 square picas. The differences increases with the size of the advert: for example a full page ad in VZ is $3,020$ sq. pica while it is only $1,860$ sq. pica in OG. And VZ offers a two-page ad (in fact it is approximately 2.1 times the one full page ad) while OG does not offer such an option.

In general, VZ directory is slightly bigger, thicker with higher quality (glossy) papers than OG directory. As a consequence of selling bigger directory, VZ has three columns in a page while OG has only two. In terms of number of circulation, VZ distributes more than $215,400$ copies while Ogden distributes $73,000$, but in terms of geographic coverage both cover similar places. These features suggest that the two directories are differentiated "products."

Verizon	# Sales	% Sales	Revenue ($)	% Revenue
Standard Listing	2,152	31.54%	0	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	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		842,148	100%

TABLE 1. Distribution of Sales and Revenues by Category.

We classify the ads into three general categories: listing, space listing, and display. Both publishers offer different options within each category. For example, VZ offers three font sizes to just list the names, address and phone number(s). OG offers listing with only two font sizes. Listing with the smallest font size is known as the standard listing, and as we mentioned earlier this was offered without a charge by both publishers, as long as the phone was registered as a business-phone.

We present the different options in Table 1. Listing option accounts for 36% and 6.5% of the total ad sales in VZ and OG, respectively. Space listing refers to an option where a space is allocated within the column under an appropriate business heading (such as Doctors, Salons, etc.). Both VZ and OG offer five different options within this category and it accounts for 22% and 3.53% of the total ad sales in VZ and OG, respectively. From this table we can see that : (i) the display option, which is the most expensive option, accounts for 70% or more of the revenue for both VZ and OG; (ii) roughly 67% of consumers choose listing and 10% choose display in VZ, while 94% and 3.3% choose listing and display in OG.

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,

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

TABLE 2. Quantity Discounts: Price per Pica for different sizes.

white background plus one color and multiple colors including photos and OG offers the same options except the 'white background' option.

One of the features in price discrimination is that the sellers offer a menu of options to choose from. To give an idea that indeed there are many options to choose from, we present a subset of the options in Table A-1 (in the Appendix). From this table we can conclude 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 the comparable advertising options, e.g., a half-page display without color costs $\$10,093$ in VZ and only $\$3,372$ in OG; (iii) the price differences between VZ and OG are smaller for the lower-end options, such as listing, than for the upper-end options such as display. 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 nonlinear pricing. We present these numbers in Table 2. As can be seen from the table, it appears that the competition is stronger for the smaller sized ads than for the larger sized ads suggesting that the competition is not uniform across all types of consumers.

In Figure 1 we present the smoothed average price and marginal prices charged by Verizon and Ogden. The figures show four important features of our data: a) that the average prices of advertisement and the marginal prices are different, which is

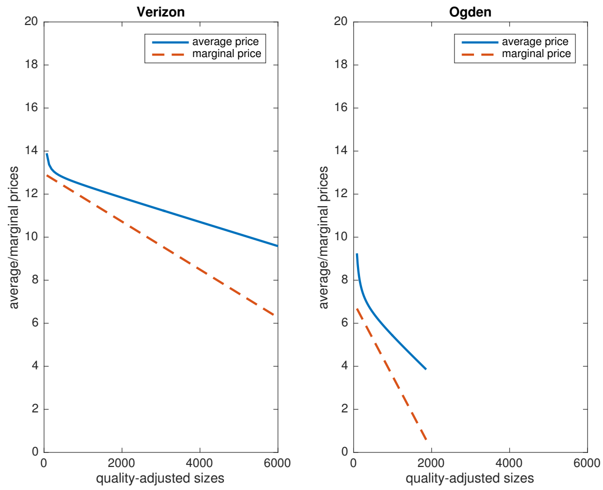


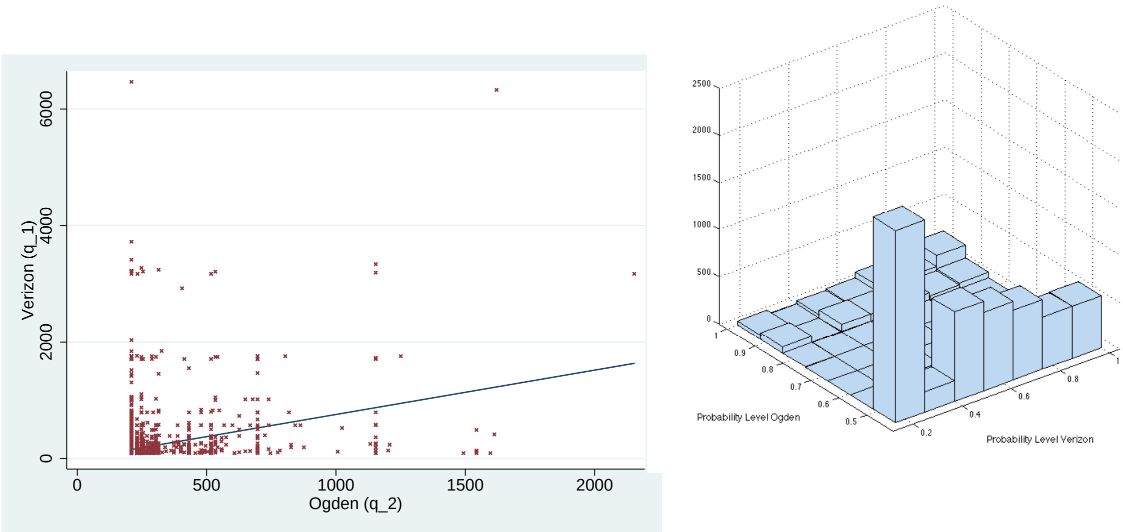
FIGURE 1. Average/Marginal Prices

consistent with sellers using nonlinear pricing;[5] b) the marginal prices are decreasing which shows that the sellers are using quantity discount (concave nonlinear prices);[6] c) VZ offers more options and charges higher prices for each option; and d) OG offers higher quantity discount than VZ.

Moreover, in our data around 54% advertise exclusively with VZ, whereas only 2% advertise exclusively with OG, while 12% advertise with both. 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. Through our empirical exercise we aim to determine the distribution of preference heterogeneity among consumers for VZ and OG that rationalizes these salient features of our data. Before we explain the model we show that the consumption pattern in the data is such that we have to allow consumers

[5]One implication of nonlinear pricing is that we cannot use average prices to estimate the demand, because consumers respond to the marginal prices.

[6]We tried several specifications for the curves, in particular different degrees for the polynomial for the Tariff function. We found out that the best fit was provided by a quadratic price function which implies a linear average price function. This guides our choice of tariff function in our model



(A) Scatter plot of advertisements, in sq. picas (B) Rank plot of advertisements.

FIGURE 2. Advertisement Choices

to have at least two-dimensional preference (one for VZ and one for OG), i,e., assuming that consumers have only one dimensional preference inconsistent with the data.

Two-Dimensional Preferences. In Figure 2 we show scatter plot and a rank plot of ads bought. As can be seen, given the variation in the choices, we need consumers to have at least two-dimensional taste parameter to rationalize the choices. If a consumer could be indexed by a one-dimensional taste parameter that govern the willingness to pay for the ad then under the quasi-linear utility environment, this would mean that both sellers value the consumer in the same way. In other words, if the one-dimensional parameter takes high (or low) value then it means both VZ and OG consider the consumer to be a high- (or low-) valued consumer. That would mean that the observed ads bought would coalesce around one increasing line in the scatter plot and the correlation between VZ and OG would be high. And as can be seen in Figure 2 (A) the best fit line is a poor representation of the heterogeneity in choices. The correlation in ad choices is only 0.25, and it goes up to only 0.32 for those who buy from both. Therefore the data suggests that we want an environment where the same consumer can be high-valued for OG but low-valued for VZ, or vice versa. To achieve that level of rich prediction we need consumers to have at least two-dimensional unobserved preference heterogeneity that are possibly correlated. The correlation between the ad choices are indicative of the correlation between the two types. In fact the Cramér-von Mises statistic for independence between the two sales was 1.66 ($p \approx 0$), rejecting the null of independence. This motivates our modeling assumption that the consumers' taste for VZ and OG ads could be correlated, resulting in a multidimensional screening problem.

Quality Adjusted Quantity. It is clear from the description about the data above that ads vary in both sizes and in quality (colors). The question that an empirical researcher faces is if we should model the products as having multiple-attributes. Although it is conceptually easy to write down the model where a firm sells a product with multiple-attributes, the problem becomes quickly intractable once we factor in multidimensional types and competition. To the best of our knowledge there is no theory paper that provides solution to such a general problem.

Instead, we propose to use a dimension reduction technique by projecting all ads onto the "space" of the most expensive (multi-colored display) ad. This projection is based on the following simple idea. We consider a non-color listing ad of size 20 sq. pica 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 is our one-dimensional aggregated quantity, which, henceforth, we refer to as the "quality-adjusted quantity." To achieve this reduction we estimate a quadratic single-index model on tariffs (defined shortly below) that accounts for proper trade-off between size and color. The modeling decision boils down to the trade off between a general yet complicated model versus a simpler model that captures the main features of the data. Given that our data is rich on choices but not on consumer covariates, we choose this dimension reduction approach.

For this aggregation to work it must not be the case that the publishers use color as a tool for price-discrimination and the aggregation must not alter the ordering of ads choices from consumers' perspective. We find suggestive evidence that color is not used as a tool to discriminate consumers: once we control for the size, the relative price is constant across colors. In particular, discounts are offered for large ads while no such discounts are observed for ads with multiple colors and the ratio of the (marginal) prices for two different colors are constant across different sizes.

To that end, we consider the price schedule and fit a continuous function that represents the price schedule while taking into account the tradeoff between color and size. We consider the following "single index model" and estimate the coefficients using OLS. With those estimates we then solve for the roots of a quadratic polynomial in q and we predict $\hat{q}$, i.e. the quality adjusted quantity for those ads that are not colored (in picas) in terms of the multicolored size. In particular we propose the following specification

$$\begin{aligned} T_{ji} &= \tilde{T}_i(q_{ji} + \delta_1 \times \texttt{color}_{ji}) + \varepsilon_{ji} \\ &= \gamma_i + \tilde{\alpha}_{1i} q_{ji} + \tilde{\alpha}_{2i}(q_{ji} \times \texttt{color}_{ji}) + \frac{\beta_i}{2} q_{ji}^2 + \varepsilon_{ji}, \\ &= \gamma_i + \underbrace{(\tilde{\alpha}_{1i} + \tilde{\alpha}_{2i} \times \texttt{color}_{ji})}_{:=\alpha_i} q_{ji} + \frac{\beta_i}{2} q_{ji}^2 + \varepsilon_{ji}, \\ &= \gamma_i + \alpha_i q_{ji} + \frac{\beta_i}{2} q_{ji}^2 + \varepsilon_{ji} \end{aligned} \tag{1}$$

where T_{ij} is the observed payment made by consumer j to publisher $i \in \{VZ, OG\}$ from buying an ad of size q_{ij} sq. pica that can be colored ($\texttt{color} = 1$) or not ($\texttt{color} = 0$), and ε_{ij} is the pure measurement error in recording the payments. We allow the intercept to be different for the three small sizes offered (beyond the free outside option) and also we allow a differentiated effect for ads of size 24 sq. picas sold by VZ and for ads of size 15 sq. picas sold by OG. Thus, $\gamma_1 = \{\gamma_{11} + \gamma_{12} \times \texttt{dsmall} + \gamma_{13} \times \texttt{d24}\}$ in the model for VZ and $\gamma_2 = \{\gamma_{22} + \gamma_{21} \times \texttt{dsmall} + \gamma_{23} \times \texttt{d15}\}$.

	Min	1st Quartile	Median	Mean	Max
Verizon	5.18	17.38	31.09	98.47	6147
Ogden	7.99	8.74	15.00	123.27	1860

TABLE 3. Summary of Quality Adjusted Quantity

The estimates are $\hat{\gamma}_1 = \{85.48 - 68.60 \times \texttt{dsmall} - 136.36 \times \texttt{d24}\}$, for the intercept, and $\{\hat{\hat{\alpha}}_{11}, \hat{\hat{\alpha}}_{21}, \hat{\beta}_1/2\} = \{9.19, 3.50, -0.0005\}$ for the model for VZ and $\hat{\gamma}_2 = \{303.48 - 223.24 \times \texttt{dsmall} + 100.95 \times \texttt{d15}\}$ for the intercept, and $\{\hat{\hat{\alpha}}_{21}, \hat{\hat{\alpha}}_{22}, \hat{\beta}_2/2\} = \{5.16, 1.59, -0.0017\}$ for OG. The $R^2 = 0.99$ for VZ and $R^2 = 0.96$ for OG. All estimates are significant at 5%.

Then, as mentioned above, the quality-adjusted quantities are constructed by plugging other (choices that are non-multicolored) onto these regression functions. For example, in VZ, a one-page ad with no color measures 3,074 sq. pica; for the same price, if this ad were multicolored the quality adjusted size is 2,616 sq. pica. See Table (3) for the summary. This transformation automatically captures the data feature that Ogden offers only a relatively small menu of listing choices, especially within each color category, and hence it is competing only in a subset of Verizon's nonlinear tariff.

One empirical regularity that has been observed, since Wilson [1993], and more recently by Chu, Leslie, and Sorensen [2011], is that the price schedules tend to be "simple," e.g., two-part tariff or quadratic, and yet approximately optimal. Given (1), in the model we can restrict the optimal prices to be a quadratic function, which will considerably simplify the model.

3. THE MODEL

In this section we present a model of competing nonlinear pricing that builds on Ivaldi and Martimort [1994]. Let $P1$ and $P2$ be two sellers that stand for VZ and OG, respectively. Even though VZ is a more dominant firm, it is not clear whether VZ acts as a Stackelberg leader or not. So for estimation we model and use the optimal allocation rules for both sellers and only the optimal price for P2. Because competition affects the equilibrium only through the price schedules and not through quantities [Busse and Rysman, 2005], we can use the allocation rules for both sellers, and the price function for P2 (P2's reaction function) is the same whether or not P2 moves at the same time as P1 or after P1.

Let $u(\mathbf{q}, \theta, A)$ be the gross utility that a consumer of type $\theta := (\theta_1, \theta_2)$ gets from choosing $\mathbf{q} := (q_1, q_2)$, where A is the set of common utility parameters. Let $T_i(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ be the pricing function chosen by P_i. If a $(\theta_1, \theta_2)-$type consumer chooses (q_1, q_2), let the net utility $U(\cdot; \theta_1, \theta_2)$ be a quasi-linear function given by

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

with $A := \{b_1, b_2, c\}, b_i > 0, i = 1, 2$ and $b_1 b_2 - c^2 > 0$ (for concavity). The (net) marginal utility for the $\theta = (\theta_1, \theta_2)$ consumer from $q_i, i = 1, 2$ is $MU_i = \theta_i - bq_i + cq_{-i} - T'(q_i)$. Therefore type θ_i denotes the intercept of the marginal utility, and higher θ_i means a consumer's marginal utility from q_i is higher and hence is willing to pay more. In view of our data (Figure 2) we model the two products as weak substitutes, and hence impose $c \leq 0$.

Therefore all consumers have the same form of utility but different (random) intercepts for marginal utility. In other words, consumers are homogenous with respect to the degree of substitution between q_1 and q_2, ceteris paribus. The quasi-linearity assumption ensures that there is no income effects, and is a widely used functional form assumption in the literature of mechanism design, discrete choice and hedonic models. We begin with the following assumptions.

Assumption 1. (1) *Let A be such that $b_1 > 0, b_2 > 0, b_1 b_2 - c^2 > 0$ and $c \leq 0$.*

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

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

The first part of the assumption 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 which simplifies the model. The second assumption says 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 adverts. Implicitly we take consumers' values as exogenously given. A richer model would model a "deeper" game among consumers (say, doctors) in their value for ads. Implicitly this means that a consumer's willingness to pay for an ad is independent of the composition of the ads placed by others in the market. In that regard, we follow the modeling assumptions used in the media economics literature; see, for example, Anderson and Waldfogel [2016] and Berry and Waldfogel [2016]. Lastly we assume 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 printing machine distribution cost and the constant marginal cost is associated with the cost of printing such as ink, paper and labor.

3.1. **Nonlinear Pricing.** Now we explain the sellers problem. Publishers do not observe $\theta's$ but have a correct belief that $\theta \overset{i.i.d}{\sim} F(\cdot)$. The objective of the sellers is to choose a pricing mechanism that maximizes expected profit, given the choice of the other seller.

Without loss of generality, from the revelation principle [Myerson, 1981], we consider a direct mechanism. A direct mechnaism 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 price 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. The pair of allocation rule and price function $\{q_i(\cdot), T_i(\cdot)\}$ 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 choice of other seller $\{q_{-i}(\cdot), T_{-i}(\cdot)\}$.

However, the application of the revelation principle when there are more than one seller is not straightforward. At the heart of the problem is the fact that the incentive compatibility constraints for P1 depends on $\{q_2(\cdot), T_2(\cdot)\}$, which in turn must satisfy the incentive compatibility constraints for P2 that depends on $\{q_1(\cdot), T_1(\cdot)\}$ ad infinity; see Epstein and Peters [1999]. Furthermore, guaranteeing that a unique equilibrium exists in our environment without any further restrictions imposed on the class of functions considered, and characterization of such equilibrium is still an open question in the literature on common agency. One way to simplify the problem is by restricting the class of price function. In view of our data (1) we assume that $T_1(\cdot)$ is quadratic, i.e.,

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

such that $T_1(\cdot)$ is right differentiable at q_{10}. Thus, P1's problem is to choose $\gamma_1 > 0, \alpha_1 > 0, \beta_1 < 0$ and the allocation rule $q_1(\cdot)$.

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

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

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

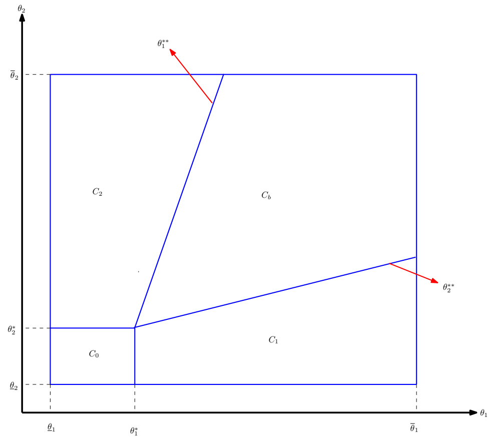


FIGURE 3. Consumer types: C_0 types choose (q_{10}, q_{20}); C_i types choose $(q_i > q_{i0})$ but q_{-io} and C_b types choose $(q_1 > q_{10}, q_2 > q_{20})$.

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

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

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

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

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

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

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

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

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

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} and is:

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

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

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

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

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

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

$q_2 > q_{20}$, the corresponding q_1 can be determined from (6) as

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

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

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

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

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

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

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

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

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

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

subject to the appropriate IC and IR constraints (see below). The threshold type z_2^0 that corresponds to the types who choose the outside option q_{20}, i.e., the area

C_1 and C_0 types in Figure 3 are such that $z_2^0 = \theta_2^* + \frac{c\theta_1^*}{b_1+\beta_1}$ if $\theta_1 \leq \theta_1^*$ and $z_2^0 = \theta_2^{**} + \frac{c\theta_1}{b_1+\beta_1}$ if $\theta_1 \geq \theta_1^*$. Since the utility from q_2 depends on q_1, which depends on q_2, (see (6) and (7)), determining the incentive compatibility constraints for P2 needs some additional work.

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

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

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

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

And let $s_2(z_2)$ be such that

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

Then 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 a sum of the indirect utility from (q_1, q_{20}) and the additional utility from choosing $q_2(z_2) > q_{20}$. Since P2 can only affect $s_2(\cdot)$ it is the only relevant "utility function,"that P2 cares about. Thus the IC constraints can be expressed as follows:

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

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

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

and

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

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

Similarly, the participation or IR constraint becomes

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

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

Then, P2 solves:

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

subject to the following IC and IR constraints, respectively

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

To solve the problem, we follow the literature and ignore the second order IC constraint and verify ex-post that the solution satisfies the constraint. Since $s_2(\cdot)$ is increasing, $s_2(z_2^0) = 0$ implies $s_2(z_2) > 0$ (IR) for all $z_2 \in (z_2^0, \bar{z}_2]$. It is immediate to see that $s_2^+ = 0$ is optimal. Since competition is endogenously accounted for in z_2 this is a one-dimensional screening problem, for which a unique equilibrium exists.[7] The next result provides the characterization of the allocation and pricing rule.

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

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

(1) *The optimal allocation function is*

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

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

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

(2) $T_2(q)$ *must satisfy (13).*

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

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

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

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

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

which determines the optimal allocation rule to be:

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

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

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

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

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

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

$$q_1(z_1)=\frac{z_1-\frac{1-G_1(z_1)}{g_1(z_1)}-m_1-\frac{c\alpha_2+c^2q_{10}}{b_2+\beta_2}}{b_1-\frac{2c^2}{b_2+\beta_2}},\quad \forall z_1\in(z_1^0,\overline{z}_1], \tag{15}$$

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

4. IDENTIFICATION

In this section we study the identification of the model parameters which are the joint distribution of types $F(\cdot,\cdot)$ with density $f(\cdot,\cdot)$ that is strictly positive everywhere in the support $\Theta:=[\underline{\theta}_1,\overline{\theta}_1]\times[\underline{\theta}_2,\overline{\theta}_2]$, the set of utility parameters $[b_1,b_2,c]$ and the set of cost parameters $[m_1,m_2,K_1,K_2]$. Our observables include the parameters from the price functions $\{\alpha_i,\beta_i,\gamma_i : i=1,2\}$, and adverts $\{q_{1j},q_{2j}\}$ placed by consumer $j=1,\ldots,J$ with the two Yellow Pages directories.

Since the outcome of the firms' optimization does not depend on the fixed cost, we cannot identify K_1 and K_2. Let ψ denote all the parameters except $F(\cdot)$, such that (i) $b_1b_2-c^2>0$; (ii) $b_i+\beta_i>0$ for $i=1,2$; and (iii) $(b_1+\beta_1)(b_2+\beta_2)-2c^2>0$. As we mentioned earlier, these conditions ensure the concavity of the utility function, and thus convexity of the optimization problem. Moreover, given the nonlinearity of our model and the fact that some consumers choose (q_{10},q_{20}) and we do not have information about the lowest type of consumers we make the following location normalizations. In particular we assume that $\underline{\theta} = 0$ and normalize the utility that the lowest types $(\underline{\theta}_1, \underline{\theta}_2)$, get from the free ads (q_{10}, q_{20}) to be zero.

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

The model provides a mapping between consumer type θ and observed choices (q_1, q_2), given the parameter ψ, i.e., $\mathbf{q}(\theta;\psi) = \begin{pmatrix} q_1(\theta_1,\theta_2;\psi) \\ q_2(\theta_1,\theta_2;\psi) \end{pmatrix} : \Theta \longrightarrow \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \in \mathbb{R}^2_+$. Since there is a one-to-one mapping between θ and z we can express the choice functions in terms of z as $\mathbf{q}(z;\psi) = \begin{pmatrix} q_1(z_1;\psi) \\ q_2(z_2;\psi) \end{pmatrix} : \Theta \longrightarrow \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \in \mathbb{R}^2_+$, where for $z_i \in (z_i^0, \overline{z}_i]$,

$$q_i(z_i;\psi) = \frac{z_i - \frac{1-G_i(z_i)}{g_i(z_i)} - m_i - \frac{c\alpha_{-i}+c^2 q_{i0}}{b_{-i}+\beta_{-i}}}{b_i - \frac{2c^2}{b_{-i}+\beta_{-i}}}, \quad i = 1,2. \tag{16}$$

There are two key properties of these allocation functions that are vital for the identification. First, the optimal allocation rule (or the choice rule) for each seller is unique. Second, the allocation rule $q_i(\cdot;\psi)$ is strictly increasing in z_i. Thus the mapping between the ad-choices and the unobserved z is unique and invertible. So, our identification strategy will be to invert each q_i to identify the the distribution of z and since $(z_1, z_2) \mapsto (\theta_1, \theta_2)$ is one-to-one, this identifies $f(\cdot,\cdot)$.

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 $q_i > q_{i0}$. 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 $z_i(\cdot) \equiv q_i^{-1}(\cdot)$, with $\overline{q}_i := \max_j q_{ij}$ is the largest ad size. Then $z_{ij} = z_i(q_{ij})$ is the type that choses $q_{ij}(> q_{i0})$, so, if $z_i^0 := z_i(q_{i0})$,

$$\begin{aligned} H_i(q) &:= \int_{q_{j0}}^{\bar{q}_j} H(q,\xi)d\xi = \Pr[q_i \leq q | q_i \geq q_{i0}] \\ &= \Pr[z_i \leq z_i(q) | z_i > z_i(q_{i0})] = \frac{G_i(z) - G_i(z_i^0)}{1 - G_i(z_i^0)}; \quad (17) \\ h_i(q) = H_i'(q) &= \frac{g_i(z)}{1 - G_i(z_i^0)} z_i'(q). \quad (18) \end{aligned}$$

From (17) and (18) we get $\frac{1-G_i(z_i)}{g_i(z_i)} = \frac{1-H_i(q)}{h_i(q)} z_i'(q)$, which together with (16) give

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

which can be inverted to give

$$z_i = q_i \left(b_i - \frac{2c^2}{b_{-i} + \beta_{-i}} \right) + m_i + \frac{c\alpha_{-i} + c^2 q_{i0}}{b_{-i} + \beta_{-i}} + \frac{(1 - H_i(q_i))}{h_i(q_i)} z_i'(q_i), \quad (19)$$

which identifies $z_i, i = 1, 2$ conditional on identifying $z_i'(\cdot)$, the marginal cost m_i and the utility parameters (b_1, b_2, c). Identification of $z_i'(\cdot)$ follows from total differentiating (13) twice (and ignoring the index because it holds for both the sellers) as

$$z_i'(q_i) = T_i''(q_i) + b_i - \frac{2c^2}{(b_{-i} + \beta_{-i})}, i = 1, 2.$$

Therefore, for those in C_b, (15) and (14) can be inverted to identify $G(\cdot, \cdot | z_1 \geq z_1^0, z_2 \geq z_2^0)$ from

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

with $i = 1, 2; j = 1, \ldots, 6328.$[8]

If every consumer bought ads with both directories then for each consumer j we would observe (q_{1j}, q_{2j}) which identifies (z_{1j}, z_{2j}) and therefore $(\theta_{1j}, \theta_{2j})$. But the

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

problem is only few consumers buy adverts from both, most patronize only one, 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. Under the assumption that the dependence between those who buy from both is the same as those who choose outside options, the joint density can be estimated. Even though the idea is simple, the empirical steps are involved because economic theory is silent about the nature of dependence between θ_1 and θ_2. In the empirical section (Section 5.2) we use insights from the theory of empirical copulas and non-nested model selection (goodness-of-fit and Vuong) tests to choose a parametric Copula that provides the best fit to the data.

Cost Parameters. One of the results in the price-discrimination literature is that the seller never distorts the output meant for the highest type, even though the output is distorted for everyone else. In other words, the equilibrium allocation is such that the output meant for the highest type, i.e., the largest output, equates marginal utility to marginal cost. This feature of the nonlinear pricing is referred to as "no-distortion-on-top," and helps to identify the marginal costs. The idea is simple. Under no-distortion-on-top, 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., $T_i'(\overline{q}_i) = \alpha_i + \beta_i \overline{q}_i, i = 1, 2$. Since we can identify $\overline{q}_i := \max_j\{q_{ij}\}, i = 1, 2$, we identify $\{m_1, m_2\}$.

Utility Parameters. Since the utility function is concave, and the parameters b_1 and b_2 and c govern the "love-for-variety", i.e., the mix of q_1 and q_2, we exploit the observed extreme choices (those who choose large q_1 but small q_2 and vice versa) to identify these parameters. In other words, since higher b_1 and b_2 forces consumers to advertise with only one Yellow page, if we observe some consumers choosing equal q_1 and q_2, say, then it must mean b_1 and b_2 are smaller. To get further intuition, suppose the utility was linear, i.e., $b_1 = b_2 = 0$, then consumers will care only about total advert sizes $(q_1 + q_2)$, but not the composition, so 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 $q_2 > 0$.

Heuristically speaking, as the utility function becomes more concave the choice will become less and less asymmetric. This constraint is most binding for the highest type $\overline{z}_1$ with the skewed (the largest q_1 and smallest q_2) ad choices. Hence, the value of b_1 must be small enough to rationalize this choice, and similarly for $\overline{z}_2$. 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.,

$$\overline{\theta}_i - b_i\overline{q}_i + cq_{j0} = \alpha_i + \beta_i\overline{q}_i \Rightarrow \overline{\theta}_i = \alpha_i + (b_i + \beta_i)\overline{q}_i - cq_{j0}, \quad j \neq i$$

thereby identifying $\overline{\theta}_1$ conditional on identifying b_1 and c and $\overline{\theta}_2$ conditional on identifying b_2 and c. Evaluating $q_2(z_2)$ in (14) at $(\overline{z}_2)$ gives

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

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

$$\overline{\theta}_2 = \overline{q}_2b_2 + m_2 + \frac{c^2q_{20} + c\alpha_1 - c\underline{\theta}_1 - 2c^2\overline{q}_2}{b_1+\beta_1}$$

$$\alpha_2 + (b_2+\beta_2)\overline{q}_2 - cq_{10} = \overline{q}_2b_2 + m_2 + \frac{-2c^2(\overline{q}_2 - q_{20}) + c(\alpha_1 - \frac{b_1}{2}q_{10} + \frac{b_2}{2}\frac{q_{20}^2}{q_{10}})}{b_1+\beta_1}$$

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

which identifies c as the negative root of the quadratic equation. Next, we consider the identification of b_1 and b_2. For any $q_i < \overline{q}_i$, we rewrite the optimal allocation rule

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

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 then identify $(\overline{\theta}_2, \overline{\theta}_1, \underline{\theta}_1)$.

5. Estimation

We observe (q_1, q_2) for every consumer j where $j = 1, 2, ..., 6823$. Using the optimal allocation functions, define the econometric model as

$$q_{ij} = \begin{cases} [z_{ij} - \frac{1-G_i(z_{ij})}{g_i(z_{ij})} - m_i - \frac{c\alpha_{-i} + c^2 q_{i0}}{b_{-i}+\beta_{-i}}] / [b_i - \frac{2c^2}{b_{-i}+\beta_{-i}}], & z_{ij} \in (z_i^0, \overline{z}_i] \\ q_{i0}, & z_{ij} \in [\underline{z}_i, z_i^0] \end{cases} \tag{21}$$

with $i = 1, 2; j = 1, \ldots, 6328$. In this model, the latent consumer type z_{ij} is the source of randomness, in other words, z_{ij} plays the role of the "error" in the usual regression models. However, the model also depends on $\frac{1-G_i(\cdot)}{g_i(\cdot)}$ which is unknown, but from the identification arguments we know it can be replaced with $\frac{1-H_i(q_{ij})}{h_i(q_{ij})} z_i'(q_{ij})$. Since $\frac{1-H_i(q_{ij})}{h_i(q_{ij})}$ can be estimated nonparametrically (described below), we can replace it with its estimate $\frac{1-\hat{H}_i(q_{ij})}{\hat{h}_i(q_{ij})}$. Besides these two equations, one for each seller, there are also the following vector of conditions that identify ψ:

$$s(\psi) = \begin{pmatrix} m_1 - \alpha_1 + \beta_1 \overline{q}_1 \\ m_2 - \alpha_2 - \beta_2 \overline{q}_2 \\ 2(b_1 + \beta_1) - \frac{\alpha_1 + \beta_1 q_1 - m_1}{\frac{1 - H_1(q_1)}{h_1(q_1)}} - \frac{\alpha_1 + \beta_1 \tilde{q}_1 - m_1}{\frac{1 - H_1(\tilde{q}_1)}{h_1(\tilde{q}_1)}} - \beta_1 \\ \overline{q}_2 c q_{10} - \alpha_2 - (b_2 + \beta_2)\overline{q}_2 + \overline{\theta}_2 \\ \overline{\theta}_1 - \overline{q}_1(b_1 - \beta_1) - \alpha_1 + c q_{20} \\ (b_2 + \beta_2)\overline{q}_2 - c q_{10} - \overline{\theta}_1 + \alpha_2 \\ (\overline{\theta}_2 - \overline{q}_2 b_2 - m_2)(b_1 + \beta_1) - c^2 q_{20} - c\alpha_1 + c\underline{\theta}_1 + 2c^2 \overline{q}_2 \\ (\underline{\theta}_1 + c q_{20}) 2 q_{10} - b_1 q_{10}^2 - b_2 q_{20}^2 \end{pmatrix} = 0 \qquad (22)$$

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 (21) and (22), and estimate the parameters using the nonlinear least squares method.

5.1. **Estimating Marginal Densities.** Let N_1^* and N_2^* denote the number of firms purchasing advertising space strictly larger than q_{10} and q_{20}, respectively and q_{ij} denotes the quantity purchased by each of those firms from $i = 1, 2$.

It is known that: (a) the kernel density estimation suffers from lack of local adaptability, i.e. it is sensitive to outliers and spurious bumps [Marron and Wand, 1992; Terrell and Scott, 1992]; (b) it surfers from boundary bias; and (c) the most widely used data-driven bandwidth selection method, the plug-in method, is adversely affected by the normal-reference rule [Jones, Marron, and Sheather, 1996; Devrôye, 1997].

So, to estimate $H_i(\cdot)$ and $h_i(\cdot)$ we use a local polynomial method. In particular we follow the simple estimator proposed by [Cattaneo, Jansson, and Ma, 2017]. As the authors highlight, this is an intuitive and easy-to-implement nonparametric density estimator based on local polynomial techniques, which does not require pre-binning or any other transformation of the data while still being fully boundary adaptive and automatic. Also, the same methodology can be used to obtain an estimation of the distribution function.

Local Polynomial Estimators (LPE) can be expressed in terms of a weighted least squares problem and therefore are easy to implement and the intuition behind them is straightforward (for details see [Fan and Gijbels, 1996]).

In the context of auctions, [Aryal, Gabrielli, and Vuong, 2015] have used LPEs to estimate the distribution and density functions of bids.

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

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

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

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

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

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

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

The starting point of our estimation is the Sklar's Theorem [Nelson, 1999] that guarantees that there is a unqiue Copula representation of the joint distribution of (z_1, z_2), i.e.,

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

where $C : [0,1]^2 \to [0,1]$ is an (unknown) Copula. Suppose we assume that the Copula is known up to up κ, i.e., $C(\cdot,\cdot) \in \mathcal{C}_0 = \{C_\kappa : \kappa \in \Gamma\}$, where Γ is the parameter set that contains κ. But as is often the case we do not know the family of Copula. For instance it could be Gaussian or $t-$Copula or Frank, and the choice affects the dependence between (z_1, z_2).

If we know the copula family, then the parameter κ could be estimated either by maximizing the joint likelihood function or by matching dependence measured by Kendall's τ, or Spearman's ρ. So we take a two-step procedure to estimate $G(z_1, z_2)$. First, we consider many widely used families of Copulas and use non-nested model selection criteria of Vuong [1989] and Cramér-von Mises goodness-of-fit tests to select one family that provides the best fit to the data; see Fermanian, Radulović, and Wegkamp [2004] and Genest, Rémillard, and Beaudoin [2009]. For instance we compared Gaussian copula with Frank copula, and gave +1 to the family that was selected and -1 to the other one. If the test could not select either then we gave both families +1. The family with the largest score is chosen. Second, we estimate the parameter of that family using pseudo maximum log-likelihood; see Genest, Ghoudi, and Rivest [1995] and Genest, Quessy, and Rémillard [2006].[9] The testing and estimation procedures can be implemented in R using copula package, so do not provide any detail of the steps here and refer interested reader to Kojadinovic and Yan [2017].

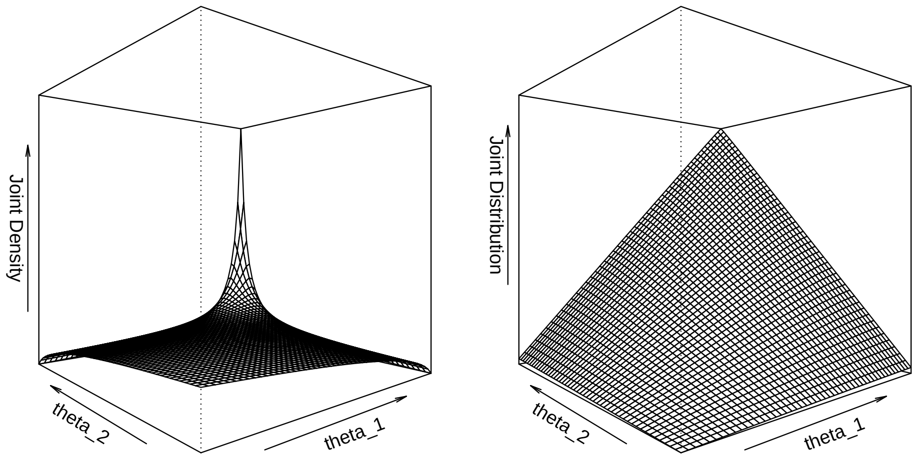
We implement Vuong test and Cramér-von Mises test for top ten families of copula and based on 10,000 simulations for each pair-wise comparison find that a Joe Copula provides the best fit. Joe copula is an example of Archimedean copula [see Nelson, 1999] that are widely used in applications, with a single parameter that controls the degree of dependence. The estimated parameter is $\hat{\kappa} = 1.214$ with bootstrapped standard error of (0.013). Once we have the estimate of (z_1, z_2) we "invert" it to determine the joint density of (θ_1, θ_2) evaluated at $z_j(\theta) \equiv z_j(\theta_1, \theta_2)$ to be

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

In Figure 4 we present the joint distribution and its corresponding joint desnity by simulating draws from them. As expected the estimate density reflects the fact that there is a large probability on high θ_1 and low θ_2.

[9] As with most goodness-of-fit tests the asymptotic distribution of the test statistic is not distribution free [Genest and Rémillard, 2004]. Furthermore we do not observe (z_1, z_2) so the estimation error in $(\hat{z}_1, \hat{z}_2)$ will further affect the asymptotic distribution. Thus we compute the critical values for Vuong-test the using the bootstrap procedure proposed in Clarke [2007] and for goodness-of-fit test we follow Genest and Rémillard [2004]; Genest and Rémillard [2008]; Kojadinovic and Holmes [2009] and Kojadinovic and Yan [2011].

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



5.3. **Estimation Results.** Table 4 presents the results of our estimation procedure as well as the lower and upper ends of the (bootstrap) confidence interval for each estimator.

TABLE 4. Estimates and Confidence Interval

	Lower	Estimate	Upper
$\underline{\theta}_1$	0.16	1.28	1.83
$\overline{\theta}_1$	263.8	567.1	757.7
$\overline{\theta}_2$	193.8	319	378
m_1	9.1	9.7	10.8
m_2	3.3	3.5	4.4
b_1	0.07	0.09	0.12
b_2	0.1	0.17	0.2
c	-0.0004	-0.0003	-0.000002

The estimated gross utility function becomes

$$\hat{u}(q_1, q_2, \theta_1, \theta_2) = \theta_1 q_1 - \frac{0,09}{2} q_1^2 + \theta_2 q_2 - \frac{0.17}{2} q_2^2 - 0.0003 \times q_1 \times q_2.$$

As can be seen $\hat{c} < 0$, which shows that the two ads can be treated as substitutes, although the rate of substitution is weak. The marginal cost of printing for VZ at $\hat{m}_1 = 9.7$ is more than twice the one for at $\hat{m}_2 = 3.5$, which captures the differences in paper size and quality. The support is estimated to be $[1.28, 567.1] \times [0, 319]$. Recall that z_i^0 is the threshold type below which consumers buy q_{i0}.

Armstrong [1996] showed that in a multidimensional screening, it is always optimal for the seller to price the goods in such a way that some positive fraction of consumers are not served. The threshold type z_i^0 then depends on the density of consumer type, e.g., if $G_i(\cdot)$ has thicker lower tail than upper tail then z_i^0 should be closer to $\underline{z}_i$ as fewer types should be excluded and vice versa.

This means that competition between VZ and OG at the lower end is more severe than at the upper end. This is reflected in the differences in prices: the difference in average price per pica widens as we move from lower category to higher, see Table 2. And the fact that VZ's prices are consistently higher across comparable categories than of OG's suggests that VZ enjoys a higher brand effect.

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

To quantify the loss, we consider the case where VZ is the leader in the market and moves before OG and they know each consumer's type.[10] In the second stage (θ_1, θ_2)- consumer who buys $\tilde{q}_1$ from VZ and pays t_1 and buys q_2 from OG gets gross utility $D(q_2; \tilde{q}_1; \theta) = u(q_2, \tilde{q}_1; \theta) - t_1$. For such q_2 the maximum price she is willing to pay is

$$t_2(q_2) = \theta_2(q_2 - q_{20}) - \frac{b_2}{2}(q_2^2 - q_{20}^2) + c\tilde{q}_1(q_2 - q_{20}). \tag{23}$$

OG will make a take-it-or-leave-it offer of q_2 at t_2 that maximizes the profit $t_2(q_2) - m_2 q_2$. From Equation (23), OG's best response is $q_2(\tilde{q}_1) = \frac{\theta_2 + c\tilde{q}_1 - m_2}{b_2}$. Now, in the first period the maximum price VZ can charge for any q_1 is

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

where $t_2(q_{10})$ can be determined by evaluating (23) at q_{10}. Then, $q_1 = (\theta_1 - m_1)/b_1$ maximizes the profit $t_1(q_1) - m_1 q_1$ and the corresponding q_2 (as a function of q_1) is $q_2 = [b_1(\theta_2 - m_2) + c(\theta_1 - m_1)]/[b_1 b_2]$.

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

---

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

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

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

One would expect that under full information, the seller will extract all consumer surplus, but because (q_{10}, q_{20}) is free, the consumer's indirect utility under complete information will not be zero but will be equal to its valuation for (q_{10}, q_{20}), which increases with (θ_1, θ_2).

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

6. CONCLUSION

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

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

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

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

Appendix A. Tables

				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).

References

ANDERSON, S. P., AND J. WALDFOGEL (2016): "Preference Externalities in Media Markets," in *Handbook of Media Economics*, ed. by S. P. Anderson, J. Waldfogel, and D. Stromberg, vol. 1A, chap. 1. Elsevier. 16

ARMSTRONG, M. (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, 64, 51–75. 35

——— (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. 4

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

ARYAL, G. (2016): "Identifying a Model of Screening with Multidimensional Consumer Heterogeneity," *Working Paper*. 6

ARYAL, G., M. F. GABRIELLI, AND Q. VUONG (2015): "Semiparametric Estimation of First-Price Auction Models," *http://arxiv.org/abs/1407.7140*. 31

ARYAL, G., I. PERRIGNE, AND Q. H. VUONG (2016): "Identification of Insurance Models with Multidimensional Screening," *Working Paper*. 6

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

BERRY, S. T. (1994): "Estimation of Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25(2), 242–262. 4

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. 16

BUSSE, M., AND M. RYSMAN (2005): "Competition and Price Discrimination in Yellow Pages Advertising," *The RAND Journal of Ecoomics*, 36, 378–390. 6, 15

CATTANEO, M. D., M. JANSSON, AND X. MA (2017): "Simple Local Polynomial Density Estimators," *Mimeo*. 30

CHERNOZHUKOV, V., A. GALICHON, M. HALLIN, AND M. HENRY (2015): "Monge-Kantororvich Depth, Quantiles, Ranks, and Signs," *Working paper*. 6

CHU, S., P. LESLIE, AND A. SORENSEN (2011): "Bundle-Size Pricing as an Approximation to Mixed Bundling," *American Economic Review*, 101(1), 263–303. 14

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

CRAWFORD, G., AND M. SHUM (2006): "Monopoly Quality Degradation in Cable Television," *Journal of Law and Economics*, 49, 181–209. 4

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. 4

EINAV, L., A. FINKELSTEIN, S. RYAN, P. SCHRIMPF, AND M. CULLEN (2012): "Selection on Moral Hazard in Health Insurance," *American Economic Reveiw*, Forthcoming. 4

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

EKELAND, I., J. HECKMAN, AND L. NESHEIM (2002): "Identifying Hedonic Models," *American Economic Reveiw*, 92(2), 304–309. 6

——— (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112, 60–109. 6

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

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

FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. Chapman & Hall. 31

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

FERMANIAN, J. D., D. RADULOVIĆ, AND M. H. WEGKAMP (2004): "Weak Convergence of Empirical Copula Processes," *Bernoulli*, 95, 119–152. 32

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. 32

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. 32

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–574. 6, 31

HECKMAN, J., R. L. MATZKIN, AND L. NESHEIM (2010): "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78(5), 1569–1591. 6

IVALDI, M., AND D. MARTIMORT (1994): "Competition under Nonlinear Pricing," *Annales d'Économie et de Statistique*, 34, 71–114. 2, 4, 14

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

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

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

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

MASKIN, E., AND J. RILEY (1984): "Monopoly with Incomplete Information," *The RAND Journal of Ecoomics*, 15, 171–196. 1

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, 4

MUSSA, M., AND S. ROSEN (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317. 1

MYERSON, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73. 16

NELSON, R. B. (1999): *An Introduction to Copulas*. Springer-Verlag New York, Inc. 32, 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. 4

NEVO, A., J. TURNER, AND J. W. WILLIAMS (2016): "UsageBased Pricing and Demand for Residential Broadband," *Econometrica*, 84(2), 411–443. 5

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

ROCHET, J.-C. (1987): "A necessary and sufficient condition for rationalizability in a quasi-linear context," *Journal of Mathematical Economics*, 16(2), 191–200. 22

ROCHET, J.-C., AND P. CHONÉ (1998): "Ironing, Sweeping, and Multidimensional Screeing," *Econometrica*, 66(4), 783–826. 22, 38

ROCHET, J.-C., AND L. A. STOLE (2002): "Nonlinear Pricing with Random Participation Constraints," *Review of Economic Studies*, 69, 277–311. 1, 2, 4

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. 4, 23

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. 4

TRAIN, K. E. (2009): *Discrete Choice Methods with Simulations*. Cambridge University Press. 4

VUONG, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57(2), 307–333. 4, 32

WILSON, R. (1993): *Nonlinear Pricing*. Oxford University Press. 1, 4, 14