An Empirical Analysis of Competitive Nonlinear Pricing

By

Gaurab Aryal
Australian National University

ANU Working Papers in Economics and Econometrics
610

May 2013 JEL: C14, D22, D82, L11, L13

ISBN: 0 86831 610 5

AN EMPIRICAL ANALYSIS OF COMPETITIVE NONLINEAR PRICING*

GAURAB ARYAL†

ABSTRACT. In this paper I estimate a model of competitive nonlinear pricing with multidimensional adverse selection. I model competition using a Stackelberg duopoly and solve the multidimensional screening problem by aggregating the multidimensional type into a single dimensional type. I study identification and estimation of the utility and cost parameters and the joint density of consumer types. The truncated marginal densities of the aggregated types can be nonparametrically identified but not the joint density. I use the classic Cramér-von Mises and Vuong's test to select one parametric family of copula to estimate the joint density from the unspecified marginals. Using a unique data for advertisements collected from two Yellow Pages Directories in Central Pennsylvania I find that: (a) Joe copula characterizes the joint density of adverse selection; (b) there is a substantial heterogeneity among advertisers; (c) the estimated density rationalizes why there is more competition at the lower end of the ads than at the upper end; (d) consumers treat the ads as substitutes; and (e) a counterfactual exercise suggests that there is a substantial (3.8% of the sales) loss of welfare due to asymmetric information.

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

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

1. INTRODUCTION

The objective of this paper is to propose a method for empirical analysis of competitive markets for differentiated products where consumers have multidimensional private information. In particular, I use a multidimensional screening model with competition, that builds on [Ivaldi and Martimort, 1994], to analyze a market where sellers compete by sequentially offering a nonlinear pricing. Using a unique data on sales of advertisements and the prices charged by two sellers, I study identification and estimation of the utility and cost functions and the joint density of the consumers' types that is unspecified (nonparametric). The estimates rationalize the data and is consistent with the theory, which suggests that the method has a potential application to other markets with competition and possibly multidimensional private information.

Date: May 12, 2013.

* Previously this paper was circulated under the title "Competition and Nonlinear Pricing in Yellow Pages." I am deeply indebted to my friend Yao Huang for her help with an earlier version of the paper. I want to thank Herman Bierens, Vijay Krishna, Mark Roberts, Neil Wallace and participants at Collegio Carlo Alberto, Columbia Uni., OSU, PSU, Queen Mary Uni. of London, UNSW and UQ. Special thanks are due to Isabelle Perrigne and Quang Vuong for their intellectual impetus and suggestions. I am solely responsible for any error. Recent version of the paper is available at the author's webpage: click here.

† The Australian National University.
e-mail:gaurab.aryal@anu.edu.au.

The presence of private or hidden information is now a widely accepted characteristic for most markets and is studied under the rubric of principal-agent problem. The theory for a monopoly seller where consumers have only one dimensional adverse selection parameter is well understood, see [Spence, 1977; Mussa and Rosen, 1978; Maskin and Riley, 1984; Wilson, 1993; Laffont and Martimort, 2001; Bolton and Dewatripont, 2005], and subsequently has been fruitfully used in empirical analysis by [Crawford and Shum, 2006; Einav, Finkelstein, and Cullen, 2010; Perrigne and Vuong, 2011a; Einav, Jenkis, and Levin, 2012; Einav, Finkelstein, Ryan, Schrimpf, and Cullen, 2012], among others. Most, if not all, markets are, however, served by more than one seller, so it is important to allow for imperfect competition. But as soon as we have more than one seller, the problem becomes difficult; the revelation principal fails [Peck, 1997; Epstein and Peters, 1999; Martimort and Stole, 2002] and solutions can be determined under assumptions that might be, often, restrictive for empirical analysis; see [Oren, Smith, and Wilson, 1983; Stole, 1995; Armstrong and Vickers, 2001; Rochet and Stole, 2003; Stole, 2007; Yang and Ye, 2008]. For example, [Yang and Ye, 2008] assume that the vertical and horizontal types are independent and are *uniformly* distributed and there is exclusive dealing, all of which are untenable with the data used in this paper (c.f. Figures 1 & 2).[1] With these models, the most difficult part is to model the optimal nonlinear pricing as a function of both multidimensional adverse selection and competition, but if one only cares about the demand side, one can use the random utility framework. [Leslie, 2004; McManus, 2006; Cohen, 2008] follow this approach, but they can identify the model only under strong non testable (parametric) distributional assumptions. Moreover, ignoring the supply side limits the scope of such models when it comes to quantifying the effect of adverse selection and/or merger on welfare and product line design.[2] Finally, with multiple sellers it becomes imperative, like in this paper, to allow for multidimensional taste parameters, and as [Armstrong, 1996; Rochet and Choné, 1998; Ekeland and Moreno-Bromberg, 2010] have articulated this is a difficult problem even for a single seller. The seminal paper by [Ivaldi and Martimort, 1994], is the closest to this paper, where they consider a duopoly competition, but under a parametric assumption on joint density of consumers' type.

In view of the data, I use Stackelberg-duopoly model and index consumer's type by a two-dimensional parameter without specifying the joint density. Under the assumption that the utility function is quadratic and concave and the cost functions are linear, a new random variable, one for each seller, can be (endogenously) determined that aggregates the two-dimensional types into one and transforms the multidimensional adverse selection problem into a single-dimensional. The

[1] The data rejects the null of independence and as seen in the figures there non-exclusivity because some consumers buy both the sellers.

[2] See [Borenstein, 1991; Lott and Roberts, 1991; Shepard, 1991; Borenstein and Rose, 1994; Clerides, 2002; Verboven, 2002; Busse and Rysman, 2005] for reduced form analysis.

aggregators act as a "sufficient statistic" for the sellers and can be used, in the place of the two dimensional type to determine the equilibrium pairs of price and an allocation functions. Incentive compatibility implies that the allocation rules are monotonic, which can then be inverted to write the unobserved aggregated types as a function of observed demand and hence they become the key source of identification. If the utility of the lowest type from consuming outside options is normalized, the utility and cost parameters can be identified, but since some consumers do not participate while not all who participate buy from both the sellers, only truncated marginal densities of aggregated types can be nonparametrically identified. Then I use a copula to estimate the joint density from the two unspecified marginals, but in order to determine the family of copula I use Cramér-von Mises test and the non nested model selection test of [Vuong, 1989] and select the family that provides the best fit among seven of the most widely used families. Since the asymptotic distribution of these tests are nonstandard, I use the multiplier Bootstrap procedures proposed by [Kojadinovic and Yan, 2011]. The utility and cost parameters are estimated from structural equations and the densities are estimated using Kernel density based on diffusion process, see [Botev, Grotowski, and Kroese, 2010], which have better performance at the boundaries than the others.

I apply the estimation procedure to a unique data on advertisements in two Yellow Pages Directories for the Center County, Pennsylvania, US. The data contains information about the menus of advertisement options and prices offered by two publishers Verizon and Ogden and the advertisements chosen by local business-units in the county.[3] A working assumption of this paper is that advertisement is a final product for businesses. The estimation results suggest that : (a) Joe copula provides the best fit for the joint density; (b) there is a substantial heterogeneity in how the two ads are valued; (c) competition is severe at the lower end of the market, which also has more mass, than the upper end of the market; (d) consumers treat the two ads as substitutes; and (e) counterfactual exercise shows that there is a loss of welfare, approximately 3.8% of sales revenue, due to asymmetric information.

This paper is also related to the literature on demand estimation, in particular the discrete choice models to study demand for differentiated products, pioneered by [Berry, 1994; Berry, Levinsohn, and Pakes, 1995]. But for identification they rely on linearity of the utility function, strong parametric assumption about the distribution of (one dimensional) consumer heterogeneity and exogenous demand shifters. Micro level data are imperative for identification in this paper while they only need aggregate level data. Another difference is that in their framework the consumer heterogeneity is only unknown to the econometricians, which means they have to take the product varieties as exogenous, while I model both the product varieties/qualities and pricing rules as endogenous

[3] Throughout the paper I use the terms business-units, firms and consumers to mean the same. The level of advertisements bought by each business units were manually recorded by reading the two directories. See the section 2 for more.

functions of unobserved consumers' type. This framework is especially useful to simulate mergers with endogenous product varieties with ease; see footnote 12 of [Berry, Levinsohn, and Pakes, 1995]. The paper is also similar to papers that study identification of principal-agent models by [Aryal, Perrigne, and Vuong, 2010; Perrigne and Vuong, 2011b], and the hedonic models by [Ekeland, Heckman, and Nesheim, 2002, 2004; Heckman, Matzkin, and Nesheim, 2010]; both these strands of literature use demand and supply side for identification.[4]

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

2. Nonlinear Pricing in Yellow Pages

The data contains information about advertisements in Yellow Pages directories sold by two publishers, Verizon (henceforth, VZ) and Ogden Directories Inc. (henceforth, OG) in Central Pennsylvania (State College and Bellefonte), US for the year 2006. The data contains information about the advertisement options (different sizes and color combinations) offered by the two publishers and the ads bought by each of the business-units in the market. A business-unit is someone with a phone registered as "for-business". The price data was provided by The Yellow Page Association, an umbrella organization of Yellow Pages publishers and the choice of advertisements were manually read-off from the two directories. It is a norm in the market to put names and addresses of each business-unit in the directory, which ensures that the list is exhaustive. Then, each business-unit was matched with the ads placed in the two directories.

VZ entered the market earlier and is a dominant player who also provides utility services. OG, on the other hand, does not provide any utility service. VZ's directory is slightly bigger, with three columns per page, thicker and the quality of the paper is better than that of OG, whose directory has only two columns per page. VZ distributes more than 215,400 copies while Ogden distributes only 73,000, but they cover the same market. Although these informations are not used in model or estimation, they do suggest to be the source of product differentiation.

Both publishers offer different advertisement options that can be classified into three general categories: (i) listing; (ii) space listing; and (iii) display. VZ and OG both offer variations within each category. For example, VZ offers three fonts sizes with listing, which is the most basic option where the name, address and phone number(s) are listed, and OG offers only two font sizes, but the listing

[4] [D'Haultfoeuille and Février, 2010] provide an important insight by showing that modeling supply side is not necessary if there is some (meaningful) exogenous variation in contracts. Although they consider single dimensional type, such variation can be useful for nonparametric identification of multidimensional adverse selection, a subject that is explored in [Aryal, 2013a].

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

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

with smallest font (known as the standard listing) is free with both.[5] Listing accounts for 30% (resp. 53%) of the total ad sales in VZ (resp. OG). Space listing refers to an option where a space is allocated within the column under an appropriate business heading (such as Doctors, Salons, etcetera), and both VZ and OG offer five different variations and it accounts for 30% (resp. 26%) of the total ad sales in VZ (resp. OG). Finally, display refers to an option where a space (that could cover up to two pages) with colorful pictures is allocated under for the buyer. VZ offers nine different variations and OG offers seven different variations within this category, which is also the most expensive of the three options. One can also choose different colors and sizes. VZ offers five color options – no color, one color, white background, white background plus one color and multiple colors including photos and OG offers four – no color, one color, white background plus one color and multiple colors including photos; see Table A-1.

The unit of measurement is picas, which is approximately 1/6 of an inch. For example, a standard listing in VZ is equal to 12 picas and in OG is equal to 9 picas, and a full page ad in VZ is equal to 3,020 picas and in OG is equal to 1,845 picas. From the table Table A-1 one can see that: (i) for any size, color accounts for most of the differences in prices, e.g. a full-page display ad with no color costs \$18,510 in VZ (resp. \$6,324 in OG), which increase to \$32,395 (resp. \$9,435) with multiple colors; (ii) VZ's price is significantly higher than that of Ogden's across all the comparable advertising options, e.g., a half-page no color display costs \$10,093 in VZ while it costs only \$3,372 in OG; (iii) the price differences between VZ and OG is smaller for the lower-end options, such as listing, than for the upper-end options, e.g., VZ's average price is 130% higher than OG's for the display option and this difference decreases drastically to 18% for space listing and to 17% for standard listing (no color); and (iv) for a given color category, both offer quantity-discount: the price per square pica decreases with the ad size, e.g., the unit price per square picas for a double-page, a full-page, and a half-page display advertisements with no color are \$5.68 (resp. \$3.43) \$6.13 (resp. \$3.68) and \$6.90 (resp. \$3.72), respectively for VZ (resp. OG); see Table 1.

Similarly, from Table A-2 one can see that: (i) the display option, which is the most expensive option, accounts for more than 70% of the revenue for both VZ and OG; (ii) roughly 66% of the

[5] To reiterate what was said earlier, standard listing provides an exhaustive list of consumers in the market and are modeled as It is because of this feature that the data contains exhaustive list of all business-units in the market.

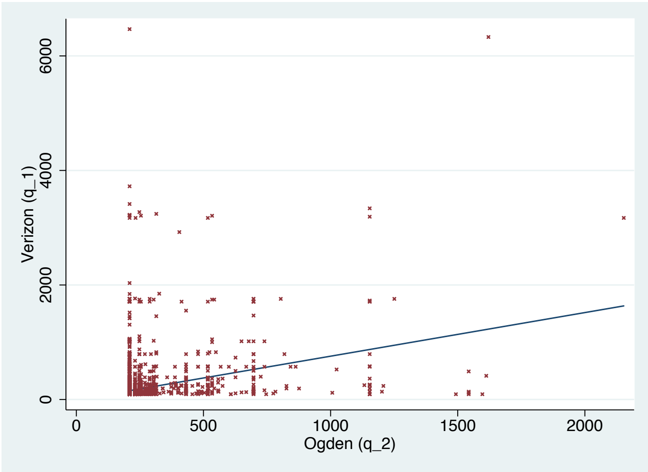

FIGURE 1. X-axis: Ogden ads; Y-axis: Verizon ads.

business-units choose listing and 14% choose display option in VZ, while the numbers for OG are 94% and 3.8%, respectively; (iii) 54% choose only Verizon, 12% choose both and only 2% choose only from Ogden and the rest choose the (default) standard listing. 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 evaluation of advertising among this group. A similar pattern is observed with respect to advertisement sizes.

In the theoretical model I allow consumers to have two-dimensional types. The reason why one dimensional private information is not sufficient and I need at least two-dimensional types lies in the demand pattern as seen in Figure 1. If consumers had only one dimensional type then within the quasi-linear environment, both VZ and OG would rank consumers identically so that a scatter plot of demand for VZ and OG ads would coalesce around an increasing straight line. Since this is inconsistent with the data, where the over all correlation between two ads is quite low at 0.25 increasing up to only 0.32 for the subset who buys from both. However, I reject the hypothesis that the two ads are independent: the Cramer-von Misses statistic for independence was 1.66 with $p \approx 0$. The (normalized) rank plot (or the probability of a sale) of VZ and OG ads is presented in Figure 2.

Quality-Adjusted Quantity. One difficulty with the data is that the offered size and colors are discrete while the theoretical model treats either quantity or quality as continuous variable, even though there are many such combinations. In view of this, I propose a way to combine the size and color into a continuous one dimensional variable and is referred to as the 'quality-adjusted-quantity'. But to verify that such a transformation preserves the ranking from the perspective of the consumers, I use the following features. First, note that if both size and color were important then the sellers

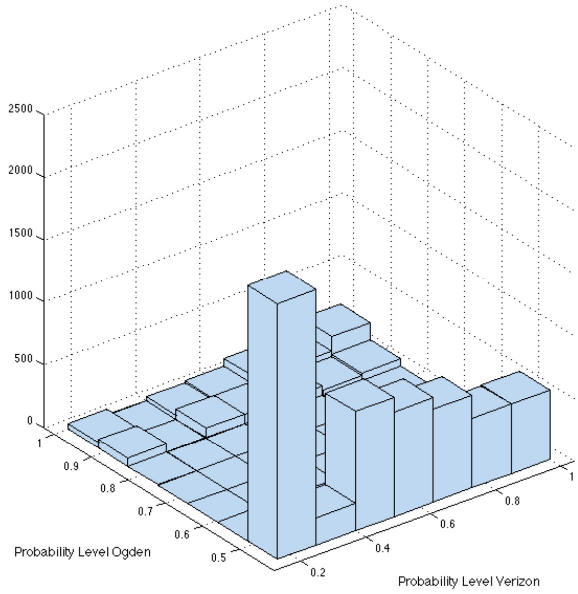

FIGURE 2. Rank plot of advertisement bought from Verizon and Ogden.

should discriminate across both the dimensions, but once I control for the size (picas) the relative price is constant across colors. In particular, discounts are offered for large advertisements while no such discounts are observed for multi color ads and the ratio of the (marginal) prices for two different colors are constant across different sizes. Second, according to [Maskin and Riley, 1984], an optimal bundle of quantity and quality should lie on a unique curve in the quantity-quality space and the optimal quantity allocation should increase with quality along this curve, something that is not observed in the data.

Now, consider the price schedule for multi colored options and fit a continuous function that represents the price schedule. Then I can project each size into this multi-color price to get a new quantity (in picas) in terms of the multi-color size. Then I fit the following quadratic functions using OLS:

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

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

TABLE 2. Summary of Quality Adjusted quantity

where T_{ji} is the price in dollars for publisher i, and q_{ji} is the advertisement size for multi-colored choices measured in square picas purchased by consumer j. The estimates are $\{\hat{\gamma}_1, \hat{\alpha}_1, \hat{\beta}_1/2\} = \{1512, 11.27, -0.00027\}$ and $\{\hat{\gamma}_2, \hat{\alpha}_2, \hat{\beta}_2/2\} = \{103, 6.25, -0.00066\}$, with $R^2 = 0.99$ and all estimates are significant at 1%. Then the quality-adjusted quantities are constructed by plugging other (non multi-colored choices) onto these regression functions. For example, one-page with no color ad in Verizon, which is is $3,020$ sq. picas, is equal to $1,470$ sq. picas in multi color. The summary of the quality-adjusted quantities are given in Table (2). Similar argument is also used by [Perrigne and Vuong, 2011a].

3. THE MODEL

Let the leader (VZ) be indexed as $P1$, who moves first, and let the follower (OG) be indexed as $P2$, who moves after observing $P1's$ choices. Let $u(\mathbf{q}, \theta, A)$ be the gross utility that a consumer of type $theta := (\theta_1, \theta_2)$ gets from choosing $\mathbf{q} := (q_1, q_2)$, and let A be the set of utility parameters that are common for all consumers. Then if a $(\theta_1, \theta_2)-$type consumer chooses (q_1, q_2), then let the net utility be

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

where $T_i(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ is the pricing function chosen by Pi. So $A := \{b_1, b_2, c\}$ and let $b_i > 0, i = 1, 2$ and let $b_1 b_2 - c^2 > 0$ (for concavity).[6] Then, q_1 and q_2 are substitutes, neutral or complements depending on if c is negative, zero or positive, respectively. To wit, note that the (net) marginal utility of q_1 is $MU_1 = \theta_1 - bq_1 + cq_2 - T'(q_1)$, which increases in q_2 (complementarity) if and only if $c > 0$. But if $c > 0$ then q_1 and q_2 should be positively-assortive given the concavity of the gross utility function. However, that is incompatible with the data, see Figures 1 & 2, so I restrict $c \leq 0$. The type (θ_1, θ_2) are independently and identically distributed as $F(\cdot, \cdot)$. The sellers, know their and their opponent's cost and $F(\cdot, \cdot)$ but not the realizations. I make the following assumptions

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

[6] Quadratic utility is general enough to capture the essential features of the data but at the same time it is simple enough to keep the model tractable; [Ekeland, Heckman, and Nesheim, 2002, 2004] also use similar utility. Nonetheless it is more general than $u(q) = \theta.q$, which is widely used in the mechanism design literature.

(3) *The cost function is assumed to be* $C_i(q_i) = K_i + m_i q_i$ *with* $K_i \geq 0$ *and* $m_i > 0$ *for* $i = 1, 2$.

All assumptions are self explanatory except the last, which is partly motivated by the observation in printing industry where two main components of cost include the cost of printing machine and fixed marginal cost of ink, paper and labor and partly by the consideration for identification. A general cost function will complicate the model without any guarantee of better inference because such a function cannot be identified, see [Perrigne and Vuong, 2011a] for non identification with a monopoly seller and single dimensional type.

The timing of the game is as follows: $P1$ chooses $\{q_1(\cdot), T_1(\cdot)\}$ and then $P2$ chooses $\{q_2(\cdot), T_2(\cdot)\}$ after observing $P1's$ choices. Then each consumer chooses (q_1, q_2) and pays accordingly. As mentioned earlier, pricing functions are roughly quadratic (1), so I restrict $T_1(\cdot)$ to be

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

which is characterized by parameters$\{\gamma_1, \alpha_1, \beta_1 : \gamma_1 > 0, \alpha_1 > 0, \beta_1 < 0\}$. But to find the solution, $T_2(\cdot)$ doesn't have to be restricted to be quadratic, as long as $T_1(\cdot)$ is.[7] In order to determine the participation (IR) and truth-telling/incentive compatibility constraint (IR), I will use the consumer's first order conditions

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

to determine four types of consumers: those who do not participate and choose (q_{10}, q_{20}) denoted as C_0, those who choose only from either $P1$ or $P2$ and are denoted, respectively as C_1 and C_2 and those who buy both $q_1 > q_{10}$ and $q_2 > q_{20}$, denoted as C_b; all in Figure 3. These four subsets are determined endogenously given $T_1(\cdot)$ and $T_2(\cdot)$.

Consider the set C_0, where the types do not participate. Then for all (θ_1, θ_2) the net marginal utilities $MU_i(\cdot, \cdot; \theta_1, \theta_2) \leq 0$ when evaluated at the pair (q_{10}, q_{20}) for $i = 1, 2$. $MU_i(q_{10}, q_{20}; \theta) \leq 0$ for $i = 1, 2, \theta \in C_0$. From (3), these two conditions can be simplified to

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

[7] In fact, in the supplementary note [Aryal, 2013b] I derive a condition that is necessary and sufficient for the optimal $T_2(\cdot)$ to be quadratic, which can provide some testable restrictions on the data.

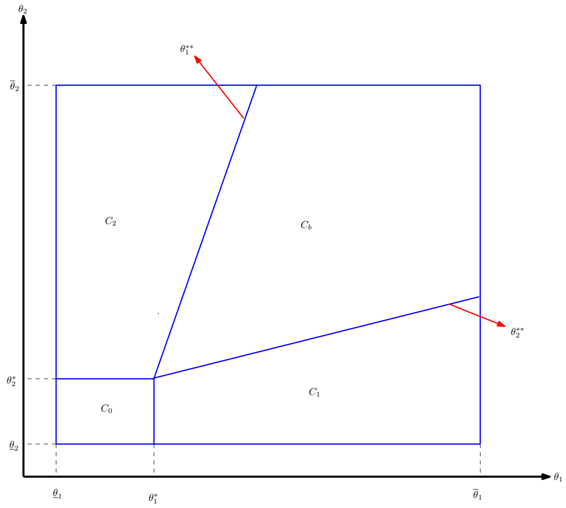

FIGURE 3. Partition of Consumer types.

without restricting $T_2(\cdot)$ to be quadratic.[8] Let (θ_1^*, θ_2^*) the marginal type who choose (q_{10}, q_{20}), i.e.

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

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

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

$$\theta_1 - b_1 q_1 + cq_{20} = \alpha_1 + \beta_1 q_1$$

$$\theta_2 - b_2 q_{20} + cq_{10} \leq T_2'(q_{20}),$$

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

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

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

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

[8] I assume that $T_i(\cdot)$ is right differentiable at q_{i0} for $i = 1, 2$.

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

$$\theta_1 - b_1 q_1 + c q_2 = \alpha_1 + \beta_1 q_1, \tag{9}$$

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

3.1. **The Follower's Problem.** In this subsection I solve for optimal nonlinear pricing (best reply) for the follower $P2$ after observing $T_1(\cdot)$ given in (3).[9] For those consumers who buy $q_2 > q_{20}$, the corresponding q_1 can be determined from (9) as

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

Substituting (11) in (10) gives the necessary condition for q_2 to be optimal for type (θ_1, θ_2) consumer, i.e.

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

Notice that the unobserved types appear only in the LHS of (12) and given $T_1(\cdot)$ they can be treated as exogenous from the point of view of $P2$. This suggests that the LHS can be treated as an aggregated one-dimensional type. Let $z_2 = \theta_2 + \frac{c\theta_1}{b_1 + \beta_1}$ be such a new type, then a type (θ_1, θ_2) consumer who is also now a type z_2 choses an optimal q_2 that solves

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

For $P2$, z_2 is unobserved and is exogenously determined, and hence can be treated as the (unobserved) preference of a firm for q_2. It aggregates (θ_1, θ_2) in the sense that it captures the taste for q_2: it is increasing in θ_2 and decreasing in θ_1 (those who value q_1 more like q_2 less) and decreases with β_1 (cheaper q_1 implies less demand for q_2, ceteris paribus). Therefore, z_2 aggregates (θ_1, θ_2) and acts as sufficient statistic in the sense that a mechanism that depends on z_2 will do as good as a mechanism that depends on (θ_1, θ_2). This also means that there will be pooling at equilibrium, i.e. two different types with same z_2 will be allocated the same good, but pooling is inevitable because the seller has only one dimensional instrument $q_2 \in \mathbb{R}$, while consumers have two dimensional types. The only important question is how should the types θ_1 and θ_2 be pooled, and argument above shows that z_2 is one such way.

[9] In the supplement [Aryal, 2013b], I solve $P2'$ problem without using this aggregation method and show that solution does not change. The advantage of using this aggregation method is its simplicity and it is keeps the identification arguments clean and simple.

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

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

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

$$\max_{T_2(\cdot), q_2(\cdot), z_2^0} \left\{ \mathbb{E}\Pi_2 = \int_{z_2^0}^{\overline{z}_2} \Big(T_2\left(q_2(z_2)\right) - m_2 q_2(z_2) \Big) g_2(z_2) \mathrm{d}z_2 - K_2 - m_2 q_{20} G_2(z_2^0) \right\}, \tag{14}$$

subject to the appropriate IC and IR constraints (see below). The threshold type z_2^0 that corresponds to the types who choose the outside option q_{20}, i.e. the area C_1 and C_0 types in Figure 3 is defined as

$$z_2^0 = \begin{cases} \theta_2^* + \frac{c\theta_1^*}{b_1+\beta_1}, & \theta_1 \le \theta_1^* \\ \theta_2^{**} + \frac{c\theta_1}{b_1+\beta_1}, & \theta_1 \ge \theta_1^*. \end{cases}$$

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

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

the net utility can be written as

$$\begin{aligned} W_2(\theta_1, z_2) &:= \max_{q_1 \ge q_{10}, q_2 \ge q_{20}} \left[u\left(q_1, q_2; \theta_1, z_2 - \frac{c\theta_1}{b1+\beta_1}\right) - T_1(q_1) - T_2(q_2) \right] \\ &= w_2(\theta_1, z_2) + s_2(q_2, z_2). \end{aligned}$$

Then the IC constraint becomes $s_2(q_2(z_2); z_2) \ge s_2(q_2(\tilde{z}_2); z_2)$ for all $z_2, \tilde{z}_2 \in [\underline{z}_2, \overline{z}_2]$. Moreover, $s_2(\cdot)$ is continuous, convex and satisfies the envelope condition

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

and

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

From (15) and (16), $P2's$ can be viewed as choosing $s_2(z_2)$, the rent function, á la [Mirrlees, 1971]. In a seminal paper Rochet [1987] showed:

Lemma 1. *The global IC constraint is satisfied if, and only if, (a) (Envelope Condition):* $s_2(z_2) = \int_{z_2^0}^{z_2}(q_2(t) - q_{20})\mathrm{d}t + s^+, \forall z_2 \in [z_2^0, \bar{z}_2]$, *where* $s_2^+ \equiv \lim_{z_2 \downarrow z_2^0} s_2(z_2)$*; and (b)* $s_2(\cdot)$ *is convex or equivalently* $q_2(z_2)$ *is increasing in* z_2.

From (15) it follows that the global IC is satisfied if and only if $q_2(\cdot)$ is strictly increasing in z_2. The participation (IR) constraint becomes $W_2(\theta_1, z_2) = w_2(\theta_1, z_2) + s_2(z_2) \geq \max\{w_2(\theta_1, z_2), 0\}$,, which is equivalent to $s_2(z_2) \geq 0$.[10] Then, $P2's$ optimization becomes

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

subject to the $q_2'(z_2) > 0$ (IC) and $s_2(z_2) \geq 0$ (IR) for all $z_2 \in [\underline{z}_2, \bar{z}_2]$. To solve the above problem, I consider the relaxed problem where the constraints are verified ex-post. Since $s_2(\cdot)$ is increasing and the optimal allocation rule must be increasing in z_2, ensuring $s_2(z_2^0) = 0$ is sufficient for (IR) constraint to be satisfied for all $z_2 \in (z_2^0, \bar{z}_2]$. It is immediate to see that $s_2^+ = 0$ is optimal. Then the existence and uniqueness of the solution can be guaranteed:

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

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

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

(1) *The optimal allocation function is*

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

[10] An advantage of using dual approach is that it not only makes finding optimal allocation rule easier (using Euler-Lagrange equation) but because implementability requires that the choice be convex, which then guarantees that the quantity allocation rule is continuous in the agent's type - a corollary of Envelope theorem. This monotonicity features prominently in identification.

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

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

(2) $T_2(q)$ *must satisfy (16) such that the corresponding for the price schedule (Ramsey rule) is*

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

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

$$\begin{aligned}
&\frac{c^2}{b_1+\beta_1}(q_2 - q_{20}) + (z_2 - \frac{c\alpha_1 - c^2 q_2}{b_1+\beta_1}) - b_2 q_2 - \frac{1-G_2(z_2)}{g_2(z_2)} - m_2 = 0\\
\Rightarrow q_2(z_2) \quad :=\quad & \frac{z_2 - m_2 - \frac{c^2 q_{20} + c\alpha_1}{b_1+\beta_1} - \frac{1-G_2(z_2)}{g_2(z_2)}}{b_2 - \frac{2c^2}{b_1+\beta_1}}.
\end{aligned}$$

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

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

And since $q_2(z_2^0) = q_{20}$, and as desired it is immediate to see that z_2^0 solves

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

Since $T_2'(q_2) = z_2 - \frac{c\alpha_1}{b_1+\beta_1} - (b_2 - \frac{2c^2}{b_1+\beta_1})q_2 + \frac{c^2}{b+1+\beta_1}q_{20}$, which follows from differentiating (16), the Ramsey equation (18), follows immediately. $\square$

The equation (18) connects the quantity discount to the distribution of the demand (i.e. z_2). For instance, the markup is smaller if either the distribution of z_2 is skewed towards the lower end, i.e. $(1 - G(z_2))$ is smaller or if $\frac{\partial s_2(q_2(z_2))}{\partial q_2}$ is higher.

3.2. **The leader's problem.** In this subsection I characterize the (optimal) nonlinear pricing for $P1$ given the continuation strategy (best response path) of $P2$. The fact that $P1's$ choices are restricted to be on the best response path complicates the pricing function, while the equilibrium allocation rule $q_1(\cdot)$ can be determined as a point-wise maximal. This suggests that competition affects the equilibrium allocation only through the price schedule. Such a phenomenon is documented by Borenstein and Rose [1994]; Busse and Rysman [2005], who find that competition leads to either higher price dispersion or higher discount, without substantial changes in product varieties. Even though $T_1(\cdot)$ is quadratic, determining the functional form of $(\gamma_1, \beta_1, \alpha_1)$ is complicated because $P2's$ choices are a function of these parameters and hence z_1 is not exogenous for $P1$. To that end I follow Wilson [1993] to give conditions that determine $T_1(\cdot)$.

Recall that firms with type $\theta_1 \geq \theta_1^{**}$ choose $q_1 > q_{10}$ and while those with $\theta_1 \leq \theta_1^{**}$ choose q_{10}. Using (11) and $T_2(\cdot)$, the allocation rule q_2 is a function of q_1:

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

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

$$z_1 = b_1 q_1 + c \left[\frac{\alpha_2 - c q_1}{b_2 + \beta_2} \right] q_1 + \alpha_1 + \beta_1 q_1. \tag{20}$$

and the threshold type

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

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

$$s_1(q_1, z_1) := \max_{\{q_1 \geq q_{10}\}} \left(z_1 + \frac{c^2 q_1 - c\alpha_2}{b_2 + \beta_2} \right) (q_1 - q_{10}) - \frac{b_1}{2}(q_1^2 - q_{10}^2) - \gamma_1 - \alpha_1 q_1 - \frac{\beta_1}{2} q_1^2.$$

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

$$T_1(q_1) = \left(z_1 + \frac{c^2 q_1 - c\alpha_2}{b_2 + \beta_2} \right) (q_1 - q_{10}) - \frac{b_1}{2}(q_1^2 - q_{10}^2) - \int_{z_1^0}^{\bar{z}_1} (q_1(t) - q_{10}) \mathrm{d}t - s_1^+. \tag{21}$$

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

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

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

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

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

and the optimal price schedule satisfies

$$T_1'(q_1) = z_1 - \frac{c\alpha_2 + c^2 q_{10}}{b_2 + \beta_2} + \left(\frac{2c^2}{b_2 + \beta_2} - b_2 \right) q_1$$

The proof is straightforward and is omitted. The optimal quantity is determined for a particular price schedule. Note that in the optimization of P1 the boundaries, of the integration in profit function, are also a function of the price schedule. Following Wilson [1993], note that if the pricing rule $T_1(q_1)$ is well behaved then the quantity allocation $z_1 \to q_1(z_1)$ is unique and regular (sufficiently many times differentiable): $q_1 = q_1(z_1) \Leftrightarrow \frac{\partial T_1(q_1)}{\partial q_1} = \Psi(q_1, z_1)$. Then making the change of variable in the profit function gives

$$\mathbb{E}\Pi_1 \quad = \quad \int_{q_{10}}^{\bar{q}_1} (T_1(q_1) - m_1 q_1) g_1(T_1'(q_1)) T_1''(q_1) \mathrm{d}q_1 - m_1 G_1(z_1^0) q_{10} - K_1,$$

where $G_1(z_1^0) = \Pr(\text{selling } q_{10})$. The optimal price is then determined by choosing α_1 and β_1. $P1's$ optimal behavior is characterized by choosing $q_1(\cdot)$ such that $q_1(z_1) = \arg\max_{q_1} s_1(q_1; z_1)$. The first order condition with respect to q_1 is $\frac{\mathrm{d}s_1(\cdot;z_1)}{\mathrm{d}z_1} = 0$. which gives $z_1 = \Psi(q_1)$ where

$$\Psi(q_1) = \left(\frac{2c^2}{b_2 + \beta_2} - (b_1 + \beta_1) \right) q_1 - \frac{c\alpha_2 + c^2 q_{10}}{b_2 + \beta_2} - T_1'(q_1).$$

Hence $\frac{\mathrm{d}z_1}{\mathrm{d}q_1} = \Psi'(q_1)\mathrm{d}q_1 = \left\{ \left(\frac{2c^2}{b_2+\beta_2} - b_1 \right) \right) - T_1''(q_1) \right\}$. Then the expected profit can be rewritten by (using the fact that $q_1(\cdot)$ is increasing) as

$$\begin{aligned} \mathbb{E}\Pi_1 &= \int_{q_{10}}^{\bar{q}_1} (T_1(q_1) - C_1(q_1))\, g_1(\Psi(t))\Psi'(t)\mathrm{d}t - m_1 G_1(z_1^0)q_{10} - K_1 \\ &= \int_{q_{10}}^{\bar{q}_1} (T_1(t) - C_1(t))\, G_1'(\Psi(t))\mathrm{d}t - m_1 G_1(z_1^0)q_{10} - K_1. \end{aligned}$$

Then integrating by parts, the first term in the right hand side becomes $\int_{q_{10}}^{\bar{q}_1}(T_1'(t) - m_1)(1 - G_1(\Psi(t)))\mathrm{d}t - m_1 q_{10} - K_1$.

Now the objective is to choose α_1 and β_1 that maximize the expected profit. The optimal quantity allocation rule determined above will be used to determine the the pricing parameters $\{\alpha_1, \beta_1, \gamma_1\}$. $P1$ chooses α_1 and β_1 to maximized the expected profit ($\mathbb{E}\Pi_1$) while γ_i will be such that the lowest type of firm's participation constraint is binding. Any extra utility resulting from interaction between the two advertisements is extracted by the principal. Since the equations that characterize these parameters are not explicitly used in empirical analysis, their derivation are not presented here but only the supplement [Aryal, 2013b].

4. Identification and Estimation

4.1. **Identification.** In this section I study the identification problem of the model, which concerns the possibility of drawing inferences from the observed data on advertisement-prices to the theoretical structure outlined above. Failure to identify the model structure implies that the data lacks sufficient information to distinguish between alternative structures. The model primitives are the joint distribution of types $F(\cdot,\cdot)$ and the set of utility and cost parameters $X = [m_1, m_2, K_1, K_2, b_1, b_2, c]$. The data provides information on the price functions $\{\alpha_i, \beta_i, \gamma_i : i = 1, 2\}$ and the ads bought by J business-units $\{q_{1j}, q_{2j}\}_{j=1}^{J}$. A structure is a set of hypothesis that implies a unique distribution consistent with the data. Two structures $\{F(\cdot,\cdot), X\} \neq \{\tilde{F}(\cdot,\cdot), \tilde{X}\}$ are said to be observationally equivalent if they imply the same probability distribution of the observed data. And the model is said to be identified if there are no two observationally equivalent models. Given our environment, I assume that the joint distribution $F(\cdot,\cdot)$ is defined on $\Theta := [\underline{\theta}_1, \overline{\theta}_1] \times [\underline{\theta}_2, \overline{\theta}_2]$ such that, all $F \in \mathcal{F}$ is absolutely continuous with continuously differentiable and nowhere vanishing density $f(\cdot,\cdot)$ and X is such that (i) $b_1 b_2 - c^2 > 0$; (ii) $b_i + \beta_i > 0$ for $i = 1, 2$; and (iii) $(b_1 + \beta_1)(b_2 + \beta_2) - 2c^2 > 0$.[11] Since the optimal nonlinear pricing functions are defined in terms of (z_1, z_2), I consider the joint distribution $G(\cdot,\cdot)$ as the model structure and study its identification. This is without loss of generality

[11]These inequalities are sufficient conditions for the utility function to be concave, and the firms' optimization problem to be convex.

because of the one-to-one mapping between the pair (θ_1, θ_2) and (z_1, z_2). Recall that the joint distribution $G(\cdot,\cdot)$ is defined on a compact support $[\underline{z}_1, \overline{z}_1] \times [\underline{z}_2, \overline{z}_2]$ and with joint density $g(\cdot,\cdot) > 0$ and the marginal densities $g_i(\cdot) > 0, i = 1, 2$ that are continuously differentiable.

Since the optimal nonlinear pricing do not depend on the fixed cost K_1 and K_2, they cannot be identified. Incentive compatibility implies that the equilibrium allocation rule $q_i(\cdot) : [z_i^0, \overline{z}] \mapsto [q_{i0}, \overline{q}_i]$ is monotonic, hence can be inverted to provide a (inverse) mapping $z_i(\cdot) \equiv q_i^{-1}(\cdot)$, from sales (data) to type. Then $z_{ij} = z_i(q_{ij})$ is the consumer $j's$ type who chose $q_{ij} > q_{i0}$ from Pi. Let $H(\cdot,\cdot)$ be the conditional joint distribution of (q_1, q_2) given $q_i > q_{i0}$. $H(\cdot,\cdot)$ and the two marginals $H_i(q_i) = \int H(q_i, q_j) dq_j$ for $i = 1, 2$ and $j \in \{1, 2\}, j \neq i$. can be identified from the data. Then

$$\begin{aligned} H_i(q) &= \Pr[q_i \leq q | q_i > q_{i0}] = \Pr[z_i \leq z_i(q) | z_i > z_i(q_{i0})] = \frac{G_i(z_i) - G_i(z_i^0)}{1 - G_i(z_i^0)}, \\ h_i(q) &= \frac{\partial H_i(q)}{\partial q_i} = \frac{g_i(z_i) z_i'(q)}{(1 - G_i(z_i^0))}, \end{aligned}$$

and hence

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

Therefore, the distribution of q_i provides some information about the distribution of the unobserved type $z_i = z_i(q_{ij})$, as shown above, which is the main source of identification.[12]

Identification of Cost Parameters. To identify the marginal cost, I use the fact that there is no distortion on the top. In other words, the highest type gets the quantity that maximizes the social welfare. Consider $P2$: the data identifies $\overline{q}_2 = \max\{q_{2i}; j = 1, \ldots, J\}$, but monotonicity implies that the type that buys $\overline{q}_2$ is $\overline{z}_2$, i.e., $q_2(\overline{z}_2) = \overline{q}_2$. Then the pricing function gives $T_2'(q_2(\overline{z}_2)) = T_2'(\overline{q}_2) = \alpha_2 + \beta_2 \overline{q}_2$. Substituting this in the Ramsey rule (18) gives $\alpha_2 + \beta_2 \overline{q}_2 - m_2 = 0$, which identifies m_2. Because of the sequentiality of the game, the same argument cannot be applied to identify m_1. Differentiating $T_1(\cdot)$ in Equation (21) with respect to q_1 and solving for $z_1'(q_1)$ gives $z_i'(q_i) = T_i''(q_i) + b_i - \frac{2c^2}{b_j + \beta_j}$, which can be used in the optimal allocation rule

$$T_i'(q_i) = m_i + \frac{1 - H_i(q_i)}{h_i(q_i)} z_i'(q_i). \tag{23}$$

for $q_1(\cdot)$ to get

$$T_1'(q_1) = m_1 + \frac{1 - H_1(q_1)}{h_1(q_1)} \left(T_1''(q_1) + b_1 - \frac{2c^2}{b_2 + \beta_2} \right), \quad \forall q_1 \in [q_{10}, \overline{q}_1].$$

[12]The problem is slightly complicated by the utility and the cost parameters.

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

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

Identification of b_1, b_2, c and the support. The utility function is concave and the parameters b_1 and b_2 induce "love for variety," ceteris paribus. To see this consider the extreme of when the utility is linear, i.e. $b_1 = b_2 = 0$, then the business-unit will care only about ads ($q_1 + q_2$), but not the composition so let $q_2 = 0$. But if $b_1 > 0$, the marginal utility from q_1 falls and q_2 starts to become important leading to $q_2 > 0$. Intuitively, as the utility becomes more concave the choice will become less and less asymmetric. This constraint is therefore most applicable to the highest type $\overline{z}_1-$ who buys the most asymmetric options: the maximum q_1 and minimum q_2. Hence, the value of b_1 must be small enough to rationalize this choice. The $\overline{z}_1-$ type's optimality condition (marginal utility equals marginal price)

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

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

$$\overline{\theta}_2 - b_2 \overline{q}_2 + c q_{10} = \alpha_2 + \beta_2 \overline{q}_2 \Rightarrow b_2 = \frac{c q_{10} + \overline{\theta}_2 - \alpha_2}{\overline{q}_2} - \beta_2$$

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

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

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

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

that can be solved to identify b_1 as

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

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

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

Since some only choose (q_{10}, q_{20}), some normalization is important to determine the support. Although there are many possible normalization that works, I make two that have some intuitive meaning. First, I assume that the type space $\Theta = [\underline{\theta}_1, \bar{\theta}_1] \times [0, \bar{\theta}_2]$ and second, normalize the utility from (q_{10}, q_{20}) of a firm with the lowest type $(\underline{\theta}_1, \underline{\theta}_2)$ at zero:

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

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

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

Identification of Marginal Densities. Since implementability implies that the equilibrium allocation rules $(q_1(\cdot), q_2(\cdot))$ are monotonic in $z's$, these mappings can be inverted to express conditional distribution of types as a function of observed demand distribution. That is using the ads placed with Verizon (resp. Ogden) I can identify the conditional marginal distribution of z_1 (resp. z_2) given that $z_1 \geq z_1^0$ (resp. $z_2 \geq z_2^0$). Recall that the relationship between type z_{ij} and ad q_{ij} is as follows $\tilde{G}_i(z_{ij}) := \frac{G_i(z_i) - G_i(z_i^0)}{1 - G_i(z_i^0)} = \frac{1 - H_i(q_{ij})}{h_i(q_{ij})} m_i$, which can then be used to recover (z_{1j}, z_{2j}) from the consumption bundle (q_{1j}, q_{2j}). It is important to note, however, that the transformation is unique only for some subset. For instance, for the set C_b it must be the case that both types are greater than the threshold, so I can invert (11) and (17) to recover the pseudo-types

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

and hence the conditional joint distribution $G(\cdot,\cdot|z_1 \geq z_1^0, z_2 \geq z_2^0)$. Now consider C_1 (resp. C_2), which includes the types that buy only from Verizon i.e. $z_2 \leq z_2^0$. In this region, I can invert only the allocation corresponding to Verizon (resp. Ogden) i.e. Equation (11) (resp. 17), to recover the corresponding z_{1j} (resp. z_{2j}). On the other hand, I can only recover the proportion of firms with $(z_1, z_2) \leq (z_1^0, z_2^0)$, i.e. $\Pr(z_1 = z_1^0, z_2 = z_2^0) = \Pr(q_1 = q_{10}, q_2 = q_{20})$.[13]

[13] Henceforth, I use the short hand (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.

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

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

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

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

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

$$\begin{aligned} \hat{H}_i(q) &= \frac{1}{N_i^*} \sum_{j=1}^{N_i^*} \mathbb{1}(q_{ij} \le q), \text{for } \ q \in [q_{i0}, \bar{q}_i] \\ \hat{h}_i(q; \xi) &= \frac{1}{N_i^*} \sum_{j=1}^{N_i^*} \frac{1}{\xi} K\left(\frac{q - q_{ij}}{\xi}\right), \end{aligned} \tag{30}$$

where ξ is a bandwidth, $K(\cdot)$ is a kernel. Among other things, however, it is known that: (a) the Kernel 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, plug-in, method is adversely affected by the normal-reference rule [Sheather and Jones, 1991; Jones, Marron, and Sheather, 1996; Devrôye, 1997]. Although various solutions have been proposed to address those shortcomings, I use adaptive kernel density estimator based on linear diffusion processes as proposed by [Botev, Grotowski, and Kroese, 2010]. The main idea is to view the kernel estimator as a transition density of a diffusion process, which leads to a simple kernel estimator with substantially reduced asymptotic bias and mean squared error with no boundary problem and an improved plug-in bandwidth selection method. A key observation is that if suppose the Kernel in Equation (30) is a Gaussian with location q_j and scale $\xi = \sqrt{t}$, i.e. (suppressing the index i for the Verizon or Ogden) $K(q, q_j; \xi) = (2\pi t)^{-1/2} \exp(-(q - q_j)^2/2t)$ then it is the unique solution to the diffusion partial differential equation (Fourier heat equation):

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

with $q \in [q_0, \overline{q}]$ and initial condition $\hat{h}(q;0) = \frac{1}{N}\sum_{i=1}^{N}\delta(q - q_j)$ (the empirical density) and the boundary condition

$$\frac{\partial}{\partial t}\hat{f}(q;t)\Big|_{q=q_0} = \frac{\partial}{\partial t}\hat{f}(q;t)\Big|_{q=\overline{q}} = 0.$$

This means that a solution to a linear diffusion process (with proper boundary condition) is a valid (nonparametric) kernel density estimator that is locally adaptive to boundary conditions. So I follow [Botev, Grotowski, and Kroese, 2010] for estimation of the densities. Since I am not interested in these estimates per se, in view of the space I do not present the estimates in the paper.[14] Once $\{\hat{H}_i(\cdot), \hat{h}_i(\cdot); i = 1, 2\}$ are estimated the parameter set X can be estimated. See Appendix (A) for more on bandwidth selection.

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

$$\hat{X}_N = \arg\min_{X \in \mathcal{X}} s_N(X)' s_N(X),$$

where

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

are the equations that identify the paramaters. Since the parameters are identified there is a unique solution to the above minimization problem. After estimating X, we can recover the pseudo types

[14] I use the computer code provided by [Botev, Grotowski, and Kroese, 2010] to estimate the densities.

z_1 and z_2 and estimate the marginal (conditional) distributions and densities $\{\hat{G}_i^*(\cdot), \hat{g}_i^*(\cdot)\}$ using the same diffusion method as outlined above. The basic consistency result is stated below without the proof, which is a straightforward extension of the consistency results by [Guerre, Perrigne, and Vuong, 2000; Perrigne and Vuong, 2011a].

Lemma 2. *Suppose all the assumptions mentioned so far is 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* $z_i > z_i^0$.

Once again, I do not present these estimates, because ultimately I am interested only in the joint density, which is addressed in next.

4.2.3. *The Joint Distribution.* Ultimately, I want to estimate $F(\cdot,\cdot)$ and not just the conditional marginals $\tilde{G}_1(z_1)$ and $\tilde{G}_2(z_2)$. Recall the data generating process: a firm $j \in \{1, 2, \ldots, N\}$ draws $(\theta_{1j}, \theta_{2j})$ *i.i.d.* from $F(\cdot,\cdot)$. Given $(T_1(\cdot), T_2(\cdot))$ the corresponding (z_{1j}, z_{2j}) is determined and j chooses $q_{1j}(z_{1j})$ and $q_{2j}(z_{2j})$. Thus unless the types are independent, the observed ads are not independent. Also recall that the null of independence was rejected. The question I am interested in is to combine the two (conditional) marginals and then extend it to the whole support. To achieve that goal, I propose to use copula. Although the marginals are censored, they are nonparametric, and hence it is convenient to adopt a parametric form for the dependence function $C_\kappa(\cdot,\cdot)$ (defined later) while keeping marginals unspecified. But there is no guidance as to what should the parameter κ be as there are many families of one parameter copulas, e.g. Gaussian, $t-$ copula, etcetera. Since it is not entirely obvious what family is appropriate, choosing one without due diligence with respect to the data will defy the whole purpose of nonparametric identification of the conditional densities. So I propose to use the classic goodness-of-fit test and Vuong's non-nested model selection test to find the "best" family. In essence the method estimates the dependence between types using data on those who buy from both and uses this dependence to select the family that provides the best global fit. I rely heavily on the current statistic literature on empirical copula and begin by formally introducing copula.

$C : [0,1]^2 \to [0,1]$ is the two-dimensional copula if $C(\cdot,\cdot)$ is the joint distribution of random variable in $[0,1]^2$ with uniform marginals. Since $G_i(\cdot)$ is a uniform random variable, the copula representation of $G(z_1, z_2)$ is $C(z_1, z_2) := C(G_1(z_1), G_2(z_2))$. The seminal Sklar's theorem [Nelson, 1999] guarantees that $C(\cdot,\cdot)$ is unique, but as mentioned earlier cannot be determined (nonparametrically) from the data. I assume that $C(\cdot,\cdot)$ is known up to a parameter, i.e. it belongs to the class

$\mathcal{C}_0 = \{C_\kappa : \kappa \in \Gamma \subset \mathbb{R}\}$, where Γ is an open set and estimate κ. Some of the widely used parametric families are Clayton, Archemedean, Gaussian copulas. If the family of copula was known, i.e. if the null $H_0 : C \in \mathcal{C}_0$, were known to be true, then the parameter κ could be estimated either by maximizing the joint likelihood function or by matching some measure of dependence such as Kendall's τ, or Spearman's ρ. However, I do not know the null, in other words I do not know if the copula is Clayton or Gaussian.

Since the marginal distribution of z_i is unspecified, I can replace it by its empirical counterpart $\hat{G}_i(\cdot) = \frac{1}{J}\sum_{j=1}^{J} \mathbb{1}(z_{ij} \leq \cdot)$. It is easier to work with (imputed) rank $\hat{r}_{ij}$ instead of the variable z_{ij} and view the copula to be based on a collection of pseudo values $(u_1, \ldots, u_J) \in \mathbb{R}^{2J}$ where $\hat{u}_{ij} := r_{ij}(J+1) = \hat{G}_i(z_{ij}) \times J/(J+1)$.[15] The first sensible thing to do is to check if z_1 and z_2 are independent, even though it was verified that q_1 and q_2 are not independent for robustness. If they are independent then the joint distribution is simply product of two marginals. So, I test

$$\begin{aligned} H_0 : \forall (u_1, u_2) \in [0,1]^2, C(u_1, u_2) &= u_1 u_2 \\ H_A : \exists (u_1, u_2) \in [0,1]^2, C(u_1, u_2) &\neq u_1 u_2. \end{aligned}$$

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

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

to test for independence. There are two issues that complicates implementation of the test. First, the asymptotic distribution of T_J under the null is not distribution free [Genest and Rémillard, 2004], and second, the distribution is also affected by the first-step errors from estimating the pseudo z_1 and z_2. So, I compute the critical values using Bootstrap procedures as outlined in [Genest and Rémillard, 2004; Kojadinovic and Holmes, 2009]. The test statistic is estimated to be $\hat{T}_J = 1.66467$ with the $p-$value equal to 0.000499. Therefore I conclude that z_1 and z_2 are not independent.

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

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

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

be an empirical process, and for a given family, let $\hat{\kappa}$ be the value of the parameter that maximizes the pseudo-log likelihood, i.e.

$$\hat{\kappa} = \arg\max_{\kappa \in \Gamma} \left\{ l(\kappa) := \sum_{i=1}^{N} \log \left[C_\kappa(\hat{U}_{i1}, \hat{U}_{i2}) \right] \right\},$$

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

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

can be used as a goodness-of-fit criteria and the test is consistent. To characterize the asymptotic distribution of the test I use the following weak convergence result for (31) from [Fermanian, Radulović, and Wegkamp, 2004]. Let $C_\kappa^{[j]} = \frac{\partial C_\kappa}{\partial u_j}$ and η_κ be a $C_\kappa-$ Brownian bridge.[16]

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

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

Using this result, the $p-$ value can be approximated from the limiting distribution of $\mathbb{T}_N$. Approximating the $p-$ value, however, is computationally costly because the limiting distribution not only depends on the asymptotic behavior of $\mathbb{C}_J$ but also on the estimator $\hat{\kappa}$.[17] Therefore, the approximate $p-$ values can only be obtained from a Bootstrap procedure outlined by [Genest and Rémillard, 2008]. In practice this is a slow method so, I use the multiplier central limit theorem to determine the large sample distribution of the test statistic; see [Kojadinovic and Yan, 2011]. I implement Cramér-von Mises the test for seven widely used families of copula and for each family estimate the test statistic, the corresponding parameter $\hat{\kappa}$ and the $p-$ value based on 10,000 Bootstrap replications. The results are reported in first column of Table 3. It is evident then that only Joe copula provides the best fit.

I also consider Vuong's non-nested model selection test à la [Vuong, 1989]. To implement the test, I consider any two families (from the seven families) and and give $+1$ to the one that is selected by this test and -1 otherwise. I follow the bootstrap procedure of [Clarke, 2007] to compute the $p-$ values For example, between Frank and Gaussian if Vuong's test selects Frank, it gets $+1$ and Gaussian gets -1. I repeat this pair-wise test for all such pairs and add all the scores and present that in Table 3).

---

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

[17] Using pseudo log likelihood is just one of many ways to estimate the parameters in the literature. For robustness, I also estimated the parameters that maximize the Kendall's Tau and Spearman's Rho, and reach the same conclusion.

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

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

The family with the highest score is the one selected. Again I find that this criteria also selects Joe copula as the best model. Therefore, one can conclude that the best family of Copula is Joe and with $\hat{\kappa} = 1.206497$ $(s.e. = 0.0137)$. Then the estimated joint density of (z_1, z_2) becomes[18]

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

Then the estimated joint density of (θ_1, θ_2), Figure 4, evaluated at $z_j(\theta) \equiv z_j(\theta_1, \theta_2)$ becomes

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

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

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

As can be seen $\hat{c} < 0$, which shows that the two ads can be treated as substitutes, although the rate of substitution is weak. The marginal cost of printing for VZ at $\hat{m}_1 = 7.768$ is twice as that of OG at $\hat{m}_2 = 3.145$, which captures the differences in the paper size and quality. The support is estimated to be $[109.39, 896.15] \times [0, 896.15]$. Recall that z_i^0 is the threshold type below which consumers buy q_{i0}. [Armstrong, 1996] showed that in a multidimensional screening, it is always optimal for the seller to price the goods in such a way that some positive fraction of consumers are not served. The threshold type z_i^0 then depends on the density of consumer type, e.g., if $G_i(\cdot)$ has thicker lower tail than upper tail then z_i^0 should be closer to $\underline{z}_i$ as fewer types should be excluded and vice versa. The estimates of the threshold types are $z_1^0 = 978.51$ and $z_2^0 = 298.83$, for VZ and OG, respectively, which suggests that $\hat{g}_2(\cdot)$ has relatively more mass at the lower end than $\hat{g}_1(\cdot)$. This also means that

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

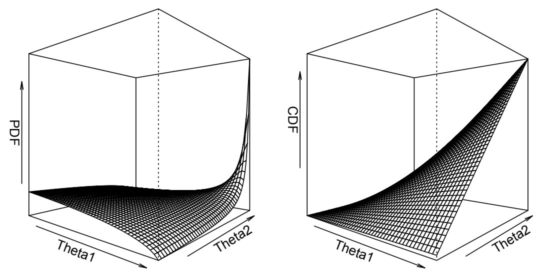

FIGURE 4. Estimated Joint PDF and CDF of recovered $(\hat{\theta}_1, \hat{\theta}_2)$

there is stiffer competition between VZ and OG at the lower end than at the upper end, which is reflected in the differences in prices: the difference in average price per picas widens as I move from lower category to higher, see Table 1. And the fact that VZ's prices are consistently higher across comparable categories than that of Ogden suggests that Verizon enjoys a higher brand effect.

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

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

where $\hat{\theta}_{ij}$ is the firm $j's$ (pseudo) marginal willingness to pay for advertisement with Pi, $\#C_{ij}$ is the number of firms with the same sub-heading as j who advertise with Pi under the same heading as j, likewise $(\#C_{ij})^2$ is the square of that, $avg(\cdot)$ and $std.dev(\cdot)$ are the average and standard deviation of advertisement sizes bought by firms in that industry. $National_{ij}$ is a dummy if the firm has a national brands (or trademark) and $Guide_{ij}$ is a dummy if j opts for guide option. Recall the Guide provides

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

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

additional advertising space by listing specialities and it covers Attorneys, Dentists, Physicians Insurance companies, etcetera. The result for both Verizon and Ogden is presented in Table 4 where the standard errors are reported in parenthesis and the (**) and (*) denote estimates that are significant at 5% and 10% confident level, respectively.

As it can be seen the direct effect of number of similar business units on the willingness to pay is positive but insignificant while the effect decreases significantly as shown by the coefficient of the Square number of firms. This suggests that the effect of competition decreases with the level of competitors, and has inverse-U shape effect which has been suggested by [Porqueras, Julien, and Chengsi, 2012]. An interesting implication of the regression is that whether or not a firm has national presence affects $\hat{\theta}_1$ negatively but $\hat{\theta}_2$ positively. The model does not explain either the demand pattern of firms with national brand or the brand effect of the publisher, because the demand side is captured by reduced form parameters. Explicitly modeling the demand side is important but beyond the scope of this paper.

Cost of Asymmetric Information. Adverse selection leads to second best outcome, and hence a loss of social welfare. It is, therefore, important to quantify the loss and how it is distributed across different consumers; in some markets, significant departure from the first best might also warrant government intervention. I conclude this section by computing the effect of asymmetric information on consumer welfare. To achieve that, I solve a Stackelberg duopoly game where the sellers know the type of each consumer and chooses a price that extracts all the rent and allocates a quantity that equate (residual) marginal utility with marginal cost, while assuming that even in this counterfactual exercise both seller offer q_{10} and q_{20} for free. Once the allocation rule and prices are determined, one can compare the difference in the utility in the data with this perfect information case.

In the second stage, a $(\theta_1, \theta_2)-$ consumer who buys $\tilde{q}_1$ from VZ and paid t_1 gets gross utility

$$D(q_2; \tilde{q}_1; \theta) = u(q_2, \tilde{q}_1; \theta) - t_1$$

if she buys q_2 from OG. For such q_2 the maximum price she is willing to pay (and what will be charged by OG) is

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

Ogden 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 (32), OG's best response is $q_2(\tilde{q}_1) = \frac{\theta_2 + c\tilde{q}_1 - m_2}{b_2}$. Now, in the first period the maximum price 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 (32) 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. Given the quasi-linear utility, I find that VZ gains $2,651,052,914 while OG gains $48,330,062 and the firms will lose $2,699,115,638. The resulting net social welfare gain is in the order of $267,337. One would expect that under full information, the seller will extract all consumer surplus, but because (q_{10}, q_{20}) is free, the consumer's indirect utility under complete information will not be zero but be equal to its valuation for (q_{10}, q_{20}), which is increasing in type. See Table 5, which presents the quantity pair under incomplete information, under full information and the corresponding difference in utility. As predicted by the theory, since the quantity allocation is not distorted for the highest type, the difference in the quantity under the two informational regime decreases with the allocation under incomplete information. The total welfare loss amounts to approximately 3.8% of the sales revenue.

5. Conclusion

In this paper I propose a framework for empirical analysis of competitive (Stackelberg duopoly) nonlinear pricing with multidimensional adverse selection. I study the problem of identification of utility and cost functions and the joint density two dimensional types. Normalization of the outside option for the lowest types is sufficient for identification except the joint density, and that transpires because some do not participate while only few buy from both sellers and only the truncated marginals can be nonparametrically identified. I use copula to determine the joint density from the

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	\$1,900,000
(572, 843)	(575, 851)	1	\$2,200,000
(1709, 697)	(1711, 706)	6	\$15,000,000
(1709, 1154)	(1711, 1160)	3	\$16,000,000
(3171, 2153)	(3173, 2153)	1	\$55,000,000
(6330, 1621)	(6330, 1624)	1	\$210,000,000

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

unspecified marginals, and use the classic goodness-of-fit and the Vuong's test to determine the parametric family of copula to use. The identification argument is constructive and provides a base for the estimation. Then I implement the estimation procedure in a data on advertisements bought from two Yellow Pages directories in Central Pennsylvania. The estimates reasonably rationalize the observed data and suggest that the consumers are sufficiently heterogeneous in terms of the valuation for the two advertising choices. Interestingly, the estimates suggest that the competition between the two publishers is higher at the lower end of the market, where the density puts more mass. This explains why the difference between per unit prices diverge as we move up the size of ads.

Estimation of the joint density of unobserved consumer types, under competition, is a prerequisite to quantify the effect of competition on welfare, or the effect of mergers on product varieties. To that end, the estimated density can be used to simulate a merger by solving the multidimensional screening problem for a multi-product monopolist à la [Rochet and Choné, 1998; Ekeland and Moreno-Bromberg, 2010]. Such a line of enquiry is new and important, but it is beyond the scope of this paper.

Therefore, this paper aims to contribute to the literature in estimation and inference of market data that are characterized with asymmetric information. In particular it contributes to new literature on adverse selection, which for the most part either considers only a monopoly seller (and ignores competition) or uses only the demand side information.

APPENDIX A. BANDWIDTH SELECTION

In this section I present the procedure followed to estimate the Kernel density and is taken from [Botev, Grotowski, and Kroese, 2010]. Given N IID realizations $Y = \{Y_1, \ldots, Y_N\}$ from an unknown continuous density $\tilde{f}(\cdot)$. The Gaussian kernel density estimator is defined as $\hat{\tilde{f}}(y;t) =$

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

Appendix B. Tables

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

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

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

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

REFERENCES

ARMSTRONG, M. (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, 64, 51–75. 2, 26

ARMSTRONG, M., AND J. VICKERS (2001): "Competitive Price Discrimination," *The RAND Journal of Ecoomics*, 32, 579–605. 2

ARYAL, G. (2013a): "Identification of Multidimensional Adverse Selection Models," *Mimeo*. 4

——— (2013b): "Supplement: An Empirical Analysis of Competitive Nonlinear Pricing," *Mimeo.*, Link to the Note. 9, 11, 13, 17

ARYAL, G., I. PERRIGNE, AND Q. H. VUONG (2010): "Identification of Insurance Models with Multidimensional Screening," *Working Paper, ANU*. 4

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890. 3, 4

BERRY, S. T. (1994): "Estimation of Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25(2), 242–262. 3

BOLTON, P., AND M. DEWATRIPONT (2005): *Contract Theory*. MIT Press. 2

BORENSTEIN, S. (1991): "Selling Costs and Switching Costs: Explaining Retail Gasoline Margins," *The RAND Journal of Ecoomics*, 22, 354–369. 2

BORENSTEIN, S., AND N. ROSE (1994): "Competition and Price Dispersion in the U.S. Airline Industry," *Journal of Political Economy*, 102, 653–683. 2, 15

BOTEV, Z. I., J. F. GROTOWSKI, AND D. P. KROESE (2010): "Kernel Density Estimation Via Diffusion," *The Annals of Statistics*, 38(5), 2916–2957. 3, 21, 22, 30

BUSSE, M., AND M. RYSMAN (2005): "Competition and Price Discrimination Competition and Price Discrimination Competition and Price Discrimination in Yellow Pages Advertising," *The RAND Journal of Ecoomics*, 36, 378–390. 2, 15

CLARKE, K. A. (2007): "A Simple Distribution-Free Test for Nonnested Model Selection," *Politica; Analysis*, 15(3), 347–363. 25

CLERIDES, S. K. (2002): "Book Value: Intertemporal Pricing and Quality Discrimination in the U.S. Market for Books," *International Journal of Industrial Organization*, 20, 1385–1408. 2

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

CRAWFORD, G., AND M. SHUM (2006): "Monopoly Quality Degradation in Cable Television," *Journal of Law and Economics*, 49, 181–209. 2

DEVRÔYE, L. (1997): "Universal Smoothing Factor Selection in Density Estimation: Theory and Practice," *Test*, 6, 223–320. 21

D'HAULTFOEUILLE, X., AND P. FÉVRIER (2010): "The Provision of Wage Incentives: A Structural Using Contracts Variation," *Working Paper CREST, INSEE*. 4

EINAV, L., A. FINKELSTEIN, AND M. R. CULLEN (2010): "Estimating Welfare in Insurance Markets using Variation in Prices," *Quarterly Journal of Economics*, 125(3), 877–921. 2

EINAV, L., A. FINKELSTEIN, S. RYAN, P. SCHRIMPF, AND M. CULLEN (2012): "Selection on Moral Hazard in Health Insurance," *American Economic Reveiw*, Forthcoming. 2

EINAV, L., M. JENKIS, AND J. LEVIN (2012): "Contract Pricing in Consumer Credit Markets," *Econometrica*, 80(4), 1387–1432. 2

EKELAND, I., J. HECKMAN, AND L. NESHEIM (2002): "Identofying Hedonic Models," *American Economic Reveiw*, 92(2), 304–309. 4, 8

——— (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112, 60–109. 4, 8

EKELAND, I., AND S. MORENO-BROMBERG (2010): "An Algorithm for Computing Solutions of Variationsal Problems wuth Global Convexity Constraints," *Numerische Mathematik*, 115, 45–69. 2, 30

EPSTEIN, L. G., AND M. PETERS (1999): "A Revelation Principle for Competing Mechanisms," *Journal of Economic Theory*, 88, 119–160. 2

FERMANIAN, J. D., D. RADULOVIĆ, AND M. H. WEGKAMP (2004): "Weak Convergence of Empirical Copula Processes," *Bernoulli*, 95, 119–152. 24, 25

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. 25

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. 25

GENEST, C., AND B. RÉMILLARD (2004): "Tests for Independence and Randomness Based on Empirical Copula Process," *Test*, 13(2), 335–369. 24

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. 25

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. 25

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–574. 23

HECKMAN, J., R. L. MATZKIN, AND L. NESHEIM (2010): "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78(5), 1569–1591. 4

IVALDI, M., AND D. MARTIMORT (1994): "Competition under Nonlinear Pricing," *Annales d'Économie et de Statistique*, 34, 71–114. 1, 2

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. 21

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. 24

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. 3, 25

LAFFONT, J.-J., AND D. MARTIMORT (2001): *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press. 2

LESLIE, P. J. (2004): "Price Discrimination in Broadway Theatre," *The RAND Journal of Economics*, 35, 520–541. 2

LOTT, J. R., AND R. D. ROBERTS (1991): "A Guide to the Pitfalls of Identifying Price Discrimination," *Economic Unquiry*, 29(1), 14–23. 2

MARRON, J. S., AND M. P. WAND (1992): "Exact Mean Integrated Error," *The Annals of Statistics*, 10, 712–736. 21

MARTIMORT, D., AND L. A. STOLE (2002): "The Revelation and Delegation Principles in Common Agency Games," *Econometrica*, 70(4), 1659–1673. 2

MASKIN, E., AND J. RILEY (1984): "Monopoly with Incomplete Information," *The RAND Journal of Ecoomics*, 15, 171–196. 2, 7

MCMANUS, B. (2006): "Nonlinear Pricing in an Oligopoly Market: The Case of Speciality Coffee," *The RAND Journal of Ecoomics*. 2

MIRRLEES, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(2), 175–208. 13

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

NELSON, R. B. (1999): *An Introduction to Copulas*. Springer-Verlag New York, Inc. 23

OREN, S. S., S. A. SMITH, AND R. WILSON (1983): "Competitive Nonlinear Tariffs," *Journal of Economic Theory*, 29(1), 49–71. 2

PECK, J. (1997): "A Note on Competing Mechanisms and the Revelation Principle," *Mimeo*. 2

PERRIGNE, I., AND Q. VUONG (2011a): "Nonlinear Pricing in Yellow Pages," *Mimeo*. 2, 8, 9, 23

PERRIGNE, I., AND Q. H. VUONG (2011b): "Nonparametric Identification of a Contract Model with Adverse Selection and Moral Hazard," *Econometrica*, 79(5), 1499–1539. 4

PORQUERAS, P. G., B. JULIEN, AND W. CHENGSI (2012): "Informative Advertising in Directed Search," *Mimeo, Monash Univeristy*. 28

ROCHET, J.-C. (1987): "A necessary and sufficient condition for rationalizability in a quasi-linear context," *Journal of Mathematical Economics*, 16(2), 191–200. 13

ROCHET, J.-C., AND P. CHONÉ (1998): "Ironing, Sweeping, and Multidimensional Screeing," *Econometrica*, 66(4), 783–826. 2, 13, 30

ROCHET, J.-C., AND L. A. STOLE (2003): "The Economics of Multidimensional Screening," in *Advances in Economics and Econometrics*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, chap. 5. Cambridge University Press. 2

SHEATHER, S. J., AND M. C. JONES (1991): "A reliable data-based bandwidth selection method for kernel density estimation," *Journal of Royal Statistical Society. Serries B (Methodological)*, 53(3), 683–690. 21

SHEPARD, A. (1991): "Price Discrimination and Retail Price Discrimination and Retail Configuration," *Journal of Political Economy*, 99, 30–53. 2

SPENCE, M. (1977): "Nonlinear Pricing and Welfare," *Journal of Public Economics*, 1, 1–18. 2

STOLE, L. A. (1995): "Nonlinear Pricing and Oligopoly," *Journal of Economic Management and Strategy*, 4, 529–562. 2

——— (2007): "Price Discrimination in Competitive Environment," in *The Handbook of Industrial Organization*, ed. by M. Armstrong, and P. R. H., vol. III. Elsivier. 2, 14

TERRELL, G. R., AND D. W. SCOTT (1992): "Variable Kernel Density Estimation," *The Annals of Statistics*, 20, 1236–1265. 21

VERBOVEN, F. (2002): "Quality-based Price Discrimination and Tax Incidence-the Market for Gasoline and Diesel Cars in Europe," *The RAND Journal of Economics*, 33, 275–297. 2

VUONG, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57(2), 307–333. 3, 25

WAND, M., AND M. C. JONES (1995): *Kernel Smoothing*. Chapman and Hall, London. 31

WILSON, R. (1993): *Nonlinear Pricing*. Oxford University Press. 2, 15, 16

YANG, H., AND L. YE (2008): "Nonlinear Pricing, Market Coverage, and Competition," *Theorretical Economics*, 3, 123–153. 2

SUPPLEMENT: AN EMPIRICAL ANALYSIS OF COMPETITIVE NONLINEAR PRICING

GAURAB ARYAL†

This note is a (optional) note to [Aryal, 2013] and collects all the derivations that are referred to in the main paper and is not meant for publication. Section 1 provides all the proofs and derivation from the main paper and Section 2 provides an alternative method to derive the optimal allocation for Principal 2, following [Rochet and Choné, 1998; Basov, 2001] and the remaining section provides the proofs left unsolved in the main paper.

1. PROOFS

In this section we collect the proofs in the main paper. I begin with Theorem 1 in [Aryal, 2013].

Theorem 1. *(Rochet and Choné [1998]) Under our maintained assumption on preferences and cost and assuming $g_1(\cdot)$ has full support, i.e. there exists $\epsilon > 0$ such that $g_1(z_1) \geq \epsilon$ for all $z_1 \in [\underline{z}, \overline{z}]$, there exists a unique solution to the optimization problem.*

Proof. The proof relies on the existence result in [Rochet and Choné, 1998] and therefore for original treatment see the paper. Even though their paper is concerned with a multidimensional screening for a monopoly, it is general enough to be applicable for the case of follower. For P2, contract chosen by P1 can be treated as exogenous parameters which affect the profit. Therefore, working with the sufficient statistic z_2 the existence result from [Rochet and Choné, 1998], Theorem 1 is applicable. The proof entails standard steps: First, we show that Π_2 as a function of s_2 is continuous and concave on Z_2^* (defined below) and Z_2^* is closed and convex. Then we show that Π_2 is coercive (defined below). The uniqueness of the optimal best response follows from the concavity.

Date: May 11, 2013.
† The Australian National University.
e-mail:gaurab.aryal@anu.edu.au webpage: http:\\sites.google.com\site\gaurabaryal.

We define a normed vector space of functions $\mathcal{H}^1(Z_2)$, functions s_2 from $Z_2 = [\underline{z}_2, \overline{z}_2]$ to $\mathbb{R}$, such that s_2 and $\bigtriangledown s_2$ are square integrable, i.e. $\mathcal{H}^1(Z_2) = \{s_2 : s_2 \in L^2(Z_2), \bigtriangledown s_2 \in L^2(Z_2)\}$,[1] with norm defined as

$$|s_2|_{\mathcal{H}^1} = \int_{Z_2} (s_2^2 + || \bigtriangledown s_2||^2) \mathrm{d}z_2.$$

From lemma (3.1) in the text, we know $\bigtriangledown s_2(z_2) = q_2(z_2)$, and suppressing z_2 we can write the expected profit for P2 as

$$\begin{aligned}
\Pi(s_2) &= \int_{Z_2} [T_2(\bigtriangledown s_2) - m_2 \bigtriangledown s_2] \mathbf{1}(s_2 \geq s_2^0) g_2(z_2) \mathrm{d}z_2 - K_2 - m_2 q_{20} G_2(z_2^0) \\
&= \int_{Z_2^*} \left\{ \left(z_2 - \frac{c\alpha_1 - c^2 \bigtriangledown s_2}{b_1 + \beta_1} \right) (\bigtriangledown s_2 - q_{20}) - \frac{b_2}{2} (\bigtriangledown s_2^2 - q_{20}^2) - m_2(\bigtriangledown s_2) - s_2 \right\} \\
&\quad \times g_2(z_2) \mathrm{d}z_2 - K_2 - m_2 q_{20} G_2(z_2^0),
\end{aligned}$$

where $Z_2^* = \{s_2 \in \mathcal{H}^1(Z_2) : s_2 \geq s_2^0\}$ and s_2^0 is the utility if an agent of type z_2 decides to purchase only q_{20}. Our objective is to show that there exists $s_2^* \in Z_2^*$ such that $\Pi_2(s_2^*) \geq \Pi_2(s_2)$ for all $s_2 \in Z_2^*$. Given the contract choice of P1, continuity of $\Pi_2(s_2)$ as a function of $s_2 \in Z_2^*$ follows from our assumption of continuity of $g_2(z_2)$ and the definition of the functional Π_2. The functional can be shown to be concave (see the main text) then the existence follows if we can show that $\Pi_2(s_2)$ is coercive, i.e. Π_2 is coercive if $\Pi_2(s_2) \to -\infty$ when $|s_2| \to \infty$; see Kinderlehrer and Stampacchia (1980) for proof of this sufficient condition. Intuitively, coercive functions are those functions that decrease without limit on any path that extends to infinity.

For all $s_2 \in \mathcal{H}^1(Z_2)$, let $\underline{s}_2$ be the mean value of s_2 in Z_2, i.e. $\underline{s}_2 = \frac{1}{|Z_2|} \int_{Z_2} s_2(z_2) \mathrm{d}z_2$. Then from Poincaré's inequality implies existence of a constant $M < \infty$ such that for $s_2 \in \mathcal{H}^1(Z_2)$, $|s_2 - \underline{s}_2| \leq M| \bigtriangledown s_2|_{L^2}$. Then using $s_2 = s_2 - \underline{s}_2 + \underline{s}_2$ we get

$$|s_2|_{L^2}^2 = |s_2 - \underline{s}_2 + \underline{s}_2|^2 = |s_2 - \underline{s}_2|_{L^2}^2 + \underline{s}_2^2 \leq M^2 | \bigtriangledown s_2|_{L^2}^2 + \underline{s}_2^2.$$

[1] $\mathcal{H}^1(Z_2)$ is a Hilbert space also known as a Sobolev space, where the functions and their first derivatives are square integrable and vanish at the boundary. We are interested in such spaces because the elements in this space are well behaved, thus enabling one to show that partial differential equations that characterize optimal nonlinear prices have solutions.

Then from the definition of the norm in $\mathcal{H}^1(Z_2)$ we note the following relationship:

$$|s_2|_{H^1} \to \infty \Leftrightarrow \text{either} |\bigtriangledown s_2|^2_{L^2} \to \infty, \text{or} \quad \underline{s}_2^2 \to \infty. \tag{1}$$

Then we observe the following relation:

$$\begin{aligned}
\Pi_2(s_2) &\leq \int_{Z_2} \left\{ \left(z_2 + \frac{c^2 \cdot \bigtriangledown s_2}{b_1 + \beta_1} \right) (\bigtriangledown s_2 - q_{20}) - \frac{b_2}{2} (\bigtriangledown s_2^2 - q_{20}^2) - m_2 \cdot \bigtriangledown s_2 - s_2 \right\} g_2(z_2) \mathrm{d}z_2 \\
&\leq \int_{Z_2} \left\{ \left(z_2 + \frac{c^2 \cdot \bigtriangledown s_2}{b_1 + \beta_1} \right) \bigtriangledown s_2 - \frac{b_2}{2} \cdot \bigtriangledown s_2^2 + \frac{b_2}{2} q_{20}^2 - m_2 \cdot \bigtriangledown s_2 - s_2 \right\} g_2(z_2) \mathrm{d}z_2 \\
&= \int_{Z_2} (z_2 \cdot \bigtriangledown s_2 - (s_2 - \underline{s}_2)) \, g_2(z_2) \mathrm{d}z_2 - \underline{s}_2 + \frac{b_2}{2} q_{20}^2 \\
&\qquad - \int_{Z_2} \left\{ \left(\frac{b_2}{2} - \frac{c^2}{b_1 + \beta_1} \right) \bigtriangledown s_2^2 + m_2 \cdot \bigtriangledown s_2 \right\} g_2(z_2) \mathrm{d}z_2 \\
&\leq \int_{Z_2} (z_2 \cdot \bigtriangledown s_2 - (s_2 - \underline{s}_2)) \, g_2(z_2) \mathrm{d}z_2 - \underline{s}_2 + \frac{b_2}{2} q_{20}^2 - \epsilon \left(\frac{b_2}{2} - \frac{c^2}{b_1 + \beta_1} \right) |\bigtriangledown s_2|^2_{L^2}
\end{aligned}$$

Therefore,

$$\Pi_2(s_2) \leq M |\bigtriangledown s_2|_{L^2} - \underline{s}_2 + \frac{b_2}{2} q_{20}^2 - \epsilon \left(\frac{b_2}{2} - \frac{c^2}{b_1 + \beta_1} \right) |\bigtriangledown s_2|^2_{L^2} \tag{2}$$

which with (A.1) implies that if $|s_2|_{H^1} \to \infty$ then $\Pi_2 \to -\infty$. Now suppose s_2^1 and s_2^2 are two optimal best response. Then, using concavity of the functional Π_2 we can show that the optimal s_2 is unique. Once s_2 is unique, it is clear that $q_2(\cdot)$ that implements s_2 is unique and so is z_2^0.

□

Lemma 1. *The P2's best response to a quadratic pricing rule of P1 is also quadratic if, and only if, $G_2(z_2) = 1 - [1 - (\varsigma + \xi z_2)]^\rho$ for some parameters ς, ξ and $\rho > 0$.*

Proof. Optimal pricing function must satisfy

$$T(z_2) = \left(z_2 - \frac{c\alpha_1 - c^2 q_2}{b_1 + \beta_1} \right) (q_2 - q_{20}) - \frac{b_2}{2} (q_2^2 - q_{20}^2) - s_2(z_2). \tag{3}$$

and the optimal allocation rule is given by

$$q_2(z_2) = \frac{z_2 - \frac{1-G_2(z_2)}{g_2(z_2)} - m_2 - \frac{c^2 q_{20} + c\alpha_1}{b_1+\beta_1}}{b_2 - \frac{2c^2}{b_1+\beta_1}}, \forall z_2 \in (z_2^0, \bar{z}_2] \tag{4}$$

such that $q_2(z_2) = q_{20}$ otherwise. Using (4), let $\tilde{h}_2(z_2), \forall z_2 \in (z_2^0, \bar{z}_2]$ and $\tau_2(q_2)$ be such that

$$\tilde{h}_2(z_2) := z_2 - \frac{1-G_2(z_2)}{g_2(z_2)} = q_2\left(b_2 - \frac{2c^2}{b_1+\beta_1}\right) + m_2 + \frac{c^2 q_{20} + c\alpha_1}{b_1+\beta_1} := \tau_2(q_2).$$

Let $z_2 \equiv \tilde{h}_2^{-1}[\tau_2(q_2)]$, then using it in (3) gives

$$\begin{aligned}
T_2(q_2) &= \left(\tilde{h}_2^{-1}(\tau_2(q_2)) - \frac{c\alpha_1 - c^2 q_2}{b_1+\beta_1}\right)(q_2 - q_{20}) - \frac{b_2}{2}(q_2^2 - q_{20}^2) \\
&\quad - \int_{q_{20}}^{q_2} (t - q_{20}) \frac{1}{\tilde{h}_2'(\tilde{h}_2^{-1}(\tau_2(t)))} \tau_2'(t) \mathrm{d}t = \left(\tilde{h}_2^{-1}(\tau(q_2)) - \frac{c\alpha_1 - c^2 q_2}{b_1+\beta_1}\right)(q_2 - q_{20}) \\
&\quad - \frac{b_2}{2}(q_2^2 - q_{20}^2) - [\tilde{h}_2^{-1}(\tau_2(t))(t - q_{20})]\big|_{q_{20}}^{q_2} + \int_{q_{20}}^{q_2} \tilde{h}_2^{-1}(\tau_2(t)) \mathrm{d}t \\
&= \int_{q_{20}}^{q_2} \tilde{h}_2^{-1}(\tau_2(t)) \mathrm{d}t - \frac{c\alpha_1 - c^2 q_2}{b_1+\beta_1}(q_2 - q_{20}) - \frac{b_2}{2}(q_2^2 - q_{20}^2),
\end{aligned}$$

which shows that the $T_2(q_2)$ is quadratic if and only if the first term in the right hand side is quadratic, or the integrand $\tilde{h}_2^{-1}(\tau_2(t)))$ is linear in t or equivalently $\frac{1-G_2(z)}{g_2(z)}$ is linear. It is enough to show that $\frac{1-G_2(\cdot)}{g_2(\cdot)}$ is linear iff $G_2(z) = 1 - [1 - (\varsigma + \xi z_2)]^\rho$. If part is obvious. Now, suppose that $\frac{1-G_2(z_2)}{g_2(z_2)} = A - Bz_2$, $B > 0$, then

$$\frac{g_2(z)}{1-G_2(z)} = \frac{1}{A - Bz_2} \Rightarrow -\frac{g(z_2)}{1-G(z_2)} = -\frac{1}{A-Bz_2} \Rightarrow \frac{\mathrm{d}\ln(1-G(z_2))}{\mathrm{d}z_2} = \frac{1}{B}\frac{\mathrm{d}\ln(A - Bz_2)}{\mathrm{d}z_2}.$$

Integrating both sides allows us to write

$$\begin{aligned}
&\int_{\underline{z}_2}^{z_2} \frac{\mathrm{d}\ln(1-G(w))}{\mathrm{d}w}\mathrm{d}w = \int_{\underline{z}_2}^{z_2} \frac{1}{B}\frac{\mathrm{d}\ln(A-Bw)}{\mathrm{d}w}\mathrm{d}w \\
\Rightarrow \quad &\ln(1-G(z_2)) = \frac{1}{B}\ln\left(\frac{A-Bz_2}{A-B\underline{z}_2}\right) \Rightarrow G(z_2) = 1 - \left(\frac{A-Bz_2}{A-B\underline{z}_2}\right)^{\frac{1}{B}}
\end{aligned}$$

It is easy to check that $G(\underline{z}_2) = 0$ and $0 \leq G(\cdot) \leq 1$ and $G(\bar{z}_2) = 1$ whenever $A = B\bar{z}_2$. Therefore,

$$G(z_2) = 1 - \left(\frac{B\bar{z}_2 - Bz_2}{B\bar{z}_2 - B\underline{z}_2}\right)^{\frac{1}{B}} = 1 - \left(1 - \frac{z_2 - \underline{z}_2}{\bar{z}_2 - \underline{z}_2}\right)^{\frac{1}{B}} = 1 - \left(1 - \left(\frac{z_2}{\bar{z}_2 - \underline{z}_2} - \frac{\underline{z}_2}{\bar{z}_2 - \underline{z}_2}\right)\right)^{\frac{1}{B}} := 1 - (1 - \{\varsigma + \xi z\})^\rho.$$

□

Lemma 2. *The coefficients of P1's optimal tariff function α_1, β_1 are unique.*

Proof. It is important to observe that not only $\overline{q}$ but also $\overline{z}_1, \underline{z}_1$ and $\Psi(t)$ are functions of α_1 and β_1. Therefore, we shall use Leibniz's method to compute the first order necessary conditions. First we do some calculations:

$$\begin{aligned}
\Psi(t) &= \left(\frac{2c^2(b_1+\beta_1)}{b_2l_2(b_1+\beta_1)-2c^2(l_2-1)} - (b_1+\beta_1)\right)t - \frac{(c\alpha_2+c^2q_{20})(b_1+\beta_1)}{b_2l_2(b_1+\beta_1)-2c^2(l_2-1)} - \alpha_1 \\
&= \left(\frac{2c^2(b_1+\beta_1)-b_2l_2(b_1+\beta_1)^2+2c^2(b_1+\beta_1)(l_2-1)}{b_2l_2(b_1+\beta_1)-2c^2(l_2-1)}\right)t - \left\{\frac{(c\alpha_2+c^2q_{20})(b_1+\beta_1)}{b_2l_2(b_1+\beta_1)-2c^2(l_2-1)}+\alpha_1\right\} \\
&= At-B.
\end{aligned}$$

The FOC with respect to α_1 is:

$$\begin{aligned}
& \underbrace{(1-G_1(\Psi(\overline{q})))}_{=0}(\alpha_1+\beta_1\overline{q}_1-m_1)\frac{\mathrm{d}\overline{q}}{\mathrm{d}\alpha_1} + \int_{q_{10}}^{\overline{q}_1}\frac{\rho_1(\overline{z}_1-\Psi(t))^{\rho_1-1}}{(\overline{z}_1-\underline{z}_1)^{\rho_1}}(-\Psi_{\alpha_1})(\alpha_1+\beta_1 t-m_1) \\
& + \left(\frac{\overline{z}_1-\Psi(t)}{\overline{z}_1-\underline{z}_1}\right)^{\rho_1}\mathrm{d}t = 0 \\
\Rightarrow \quad & -\Psi_{\alpha_1}\left\{(\alpha_1-m_1)\int_{q_{10}}^{\overline{q}_1}\frac{\rho_1(\overline{z}_1-\Psi(t))^{\rho_1-1}}{(\overline{z}_1-\underline{z}_1)^{\rho_1}}\mathrm{d}t + \beta_1\int_{q_{10}}^{\overline{q}_1}\frac{\rho_1(\overline{z}_1-\Psi(t))^{\rho_1-1}}{(\overline{z}_1-\underline{z}_1)^{\rho_1}}t\mathrm{d}t\right\} \\
& + \int_{q_{10}}^{\overline{q}_1}\left(\frac{\overline{z}_1-\Psi(t)}{\overline{z}_1-\underline{z}_1}\right)^{\rho_1}\mathrm{d}t = 0 \\
\Rightarrow \quad & -\Psi_{\alpha_1}\left\{(\alpha_1-m_1)\int_{q_{10}}^{\overline{q}_1}\frac{\rho_1(\overline{z}_1-At+B)^{\rho_1-1}}{(\overline{z}_1-\underline{z}_1)^{\rho_1}}\mathrm{d}t + \beta_1\int_{q_{10}}^{\overline{q}_1}\frac{\rho_1(\overline{z}_1-At+B)^{\rho_1-1}}{(\overline{z}_1-\underline{z}_1)^{\rho_1}}t\mathrm{d}t\right\} \\
& + \int_{q_{10}}^{\overline{q}_1}\left(\frac{\overline{z}_1-At+B}{\overline{z}_1-\underline{z}_1}\right)^{\rho_1}\mathrm{d}t = 0.
\end{aligned}$$

We can evaluate each term separately as follows:

First Term:

$$\begin{aligned}
\rho_1\int_{q_{10}}^{\overline{q}_1}(\overline{z}_1-At+B)^{\rho_1-1}\mathrm{d}t &= -\frac{1}{A}\int_{q_{10}}^{\overline{q}_1}\{(\overline{z}_1-At+B)^{\rho_1}\}'\mathrm{d}t = -\frac{1}{A}(\overline{z}_1-At+B)\Big|_{q_{10}}^{\overline{q}_1} \\
&= \frac{\nu_2^{\rho_1}-\nu_1^{\rho_1}}{A},
\end{aligned}$$

where $\nu_1 = \bar{z}_1 - A\bar{q}_1 + B$ and $\nu_2 = \bar{z}_1 - Aq_{10} + B$.

Second Term:

$$\begin{aligned}\int_{q_{10}}^{\bar{q}_1} \rho_1(\bar{z}_1 - At + B)^{\rho_1 - 1} t\mathrm{d}t &= -\frac{1}{A}\int_{q_{10}}^{\bar{q}_1}\{(\bar{z}_1 - At + B)^{\rho_1}\}' t\mathrm{d}t \\ &= -\frac{1}{A}\left[\bar{q}_1\nu_1^{\rho_1} - q_{10}\nu_2^{\rho_1} - \int_{q_{10}}^{\bar{q}_1}(\bar{z}_1 - At + B)^{\rho_1}\mathrm{d}t\right].\end{aligned}$$

Let $\bar{z}_1 - At + B = x$, then we can rewrite the last expression in the parenthesis as

$$-\frac{1}{A}\int_{\nu_2}^{\nu_1} x^{\rho_1}\mathrm{d}x = -\frac{\nu_1^{1+\rho_1} - \nu_2^{1+\rho_1}}{A(1+\rho_1)},$$

which when substituted back to the previous expression allows us to write the second term as

$$\int_{q_{10}}^{\bar{q}_1} \rho_1(\bar{z}_1 - At + B)^{\rho_1 - 1} t\mathrm{d}t = \frac{q_{10}\nu_2^{\rho_1} - \bar{q}_1\nu_1^{\rho_1}}{A} - \frac{\nu_1^{1+\rho_1} - \nu_2^{1+\rho_1}}{A^2(1+\rho_1)}.$$

Third term:

$$\int_{q_{10}}^{\bar{q}_1}(\bar{z}_1 - At + B)^{\rho_1}\,\mathrm{d}t = -\frac{\nu_1^{1+\rho_1} - \nu_2^{1+\rho_1}}{A(1+\rho_1)}.$$

Putting together the FOC becomes

$$-\frac{\Psi_{\alpha_1}(\alpha_1 - m_1)(\nu_2^{\rho_1} - \nu_1^{\rho_1})}{A} - \frac{\Psi_{\alpha_1}\beta_1(q_{10}\nu_2^{\rho_1} - \bar{q}_1\nu_1^{\rho_1})}{A} + \frac{\Psi_{\alpha_1}\beta_1(\nu_1^{1+\rho_1} - \nu_2^{1+\rho_1})}{A^2(1+\rho_1)} - \frac{\nu_1^{1+\rho_1} - \nu_2^{1+\rho_1}}{A(1+\rho_1)} = 0$$

$$\Rightarrow \quad -\Psi_{\alpha_1}(\alpha_1 - m_1)(\nu_2^{\rho_1} - \nu_1^{\rho_1}) - \Psi_{\alpha_1}\beta_1(q_{10}\nu_2^{\rho_1} - \bar{q}_1\nu_1^{\rho_1}) - \left(\frac{\Psi_{\alpha_1}\beta_1}{A} - 1\right)\frac{(\nu_2^{(1+\rho_1)} - \nu_1^{(1+\rho_1)})}{(1+\rho_1)} = 0.$$

We can write $\Psi_{\beta_1}(t) = Jt + D$ after some straightforward calculation and substitution, as shown later. Let $\bar{z}_1' = \frac{\bar{z}_1}{\partial\beta_1}$ and similarly we define $\underline{z}_1'$, and let $\triangle z' = \bar{z}_1' - \underline{z}_1'$.

Some preliminary calculations:

$$\begin{aligned}\Psi_{\alpha_1} &= \frac{\partial\Psi(t)}{\partial\alpha_1} = -\frac{c^2(l_2-1)}{\frac{(b_2l_2(b_1+\beta_1)-2c^2(l_2-1))}{(b_1+\beta_1)}(b_1+\beta_1)} - 1 = \frac{-c^2(l_2-1) - b_2l_2(b_1+\beta_1) + 2c^2(l_2-1)}{(b_2l_2(b_1+\beta_1) - 2c^2(l_2-1))} \\ &= \frac{c^2(l_2-1) - b_2l_2(b_1+\beta_1)}{(b_2l_2(b_1+\beta_1) - 2c^2(l_2-1))}.\end{aligned}$$

Since ν_1 is just the first order condition for $\bar{z}_1$ where the type $\bar{z}_1$ purchases the optimal quantity, a necessary condition for optimality is that it must be equal to zero. Therefore, in what follows we shall substitute $\nu_1 = 0$. So, the first order condition with respect to α_1 becomes

$$\Psi_{\alpha_1}(\alpha_1 - m_1)\nu_2^{\rho_1} + \Psi_{\alpha_1}\beta_1 q_{10}\nu_2^{\rho_1} + \left(\frac{\Psi_{\alpha_1}\beta_1}{A} - 1\right)\frac{(\nu_2^{(1+\rho_1)})}{(1+\rho_1)} = 0$$

$$or, \Psi_{\alpha_1}(\alpha_1 - m_1) + \Psi_{\alpha_1}\beta_1 q_{10} + \left(\frac{\Psi_{\alpha_1}\beta_1}{A} - 1\right)\frac{\nu_2}{(1+\rho_1)} = 0.$$

Next, we solve ν_2 to get

$$\nu_2 = \bar{z}_1 - Aq_{10} + M_1 + M_2\alpha_1,$$

where M_1 and M_2 are determined by B, as shown below:

$$\begin{aligned}
B &= \frac{(c\alpha_2 + c^2 q_{20})(b_1+\beta_1)}{b_2 l_2(b_1+\beta_1) - 2c^2(l_2-1)} + \alpha_1 \\
&= \frac{c(\zeta_2 + l_2 m_2)(b_1+\beta_1) + c(c\alpha_1 + c^2 q_{20})(l_2-1) + c^2 q_{20}(b_1+\beta_1)}{b_2 l_2(b_1+\beta_1) - 2c^2(l_2-1)} + \alpha_1 \\
&= \frac{c(\zeta_2 + l_2 m_2)(b_1+\beta_1) + c^3 q_{20}(l_2-1) + c^2 q_{20}(b_1+\beta_1)}{b_2 l_2(b_1+\beta_1) - 2c^2(l_2-1)} + \alpha_1\left\{1 + \frac{c^2(l_2-1)}{b_2 l_2(b_1+\beta_1) - 2c^2(l_2-1)}\right\} \\
&= \frac{c(\zeta_2 + l_2 m_2)(b_1+\beta_1) + c^3 q_{20}(l_2-1) + c^2 q_{20}(b_1+\beta_1)}{b_2 l_2(b_1+\beta_1) - 2c^2(l_2-1)} + \alpha_1\left\{\frac{b_2 l_2(b_1+\beta_1) - c^2(l_2-1)}{b_2 l_2(b_1+\beta_1) - 2c^2(l_2-1)}\right\} \\
&= M_1 + M_2\alpha_1.
\end{aligned}$$

Substituting everything in the FOC we get

$$\Psi_{\alpha_1}(\alpha_1 - m_1) + \Psi_{\alpha_1}\beta_1 q_{10} + \left(\frac{\Psi_{\alpha_1}\beta_1}{A} - 1\right)\frac{\bar{z}_1 - Aq_{10} + M_1 + M_2\alpha_1}{(1+\rho_1)} = 0$$

$$or, \quad \alpha_1\left\{\Psi_{\alpha_1} + \left(\frac{\Psi_{\alpha_1}\beta_1}{A} - 1\right)\frac{M_2}{(1+\rho_1)}\right\} + \Psi_{\alpha_1}(\beta_1 q_{10} - m_1) + \left(\frac{\Psi_{\alpha_1}\beta_1}{A} - 1\right)\frac{\bar{z}_1 - Aq_{10} + M_1}{(1+\rho_1)} = 0,$$

and solving for α_1, we get

$$\alpha_1 = \frac{A\Psi_{\alpha_1}(1+\rho_1)(m_1 - \beta_1 q_{10}) + (A - \Psi_{\alpha_1}\beta_1)(\bar{z}_1 - Aq_{10} + M_1)}{\{\Psi_{\alpha_1}A(1+\rho_1) + M_2(\Psi_{\alpha_1}\beta_1 - A)\}}$$

which is equation (A.6). To determine optimal β_1 we optimize with respect to β_1:

$$\underbrace{(1 - G_1(\Psi(\bar{q}_1)))}_{=0}(\alpha_1 + \beta_1\bar{q}_1 - m_1)\frac{\partial\bar{q}_1}{\partial\beta_1} + \frac{1}{(\bar{z}_1 - \underline{z}_1)^{\rho_1}}\int_{q_{10}}^{\bar{q}}\Big\{\rho_1(\bar{z}_1 - \Psi(t))^{\rho_1-1}(\bar{z}_1' - \Psi_{\beta_1})(\alpha_1 + \beta_1 t - m_1)$$

$$+(\bar{z}_1 - \Psi(t))^{\rho_1}t\Big\}\mathrm{d}t - \int_{q_{10}}^{\bar{q}_1}\frac{\rho_1 \triangle z_1'}{(\bar{z}_1 - \underline{z}_1)^{\rho_1+1}}(\alpha_1 + \beta_1 t - m_1)(\bar{z}_1 - At + B)^{\rho_1}\mathrm{d}t = 0$$

We can express $\Psi(t)_{\beta_1}$ as an explicit function of t:

$$\begin{aligned}\Psi(t)_{\beta_1} &= \frac{\partial\Psi(t)}{\partial\beta_1} = \frac{2c^2l_2 - 2b_2l_2(b_1+\beta_1) - 2c^2l_2^2b_2(b_1+\beta_1) - b_2^2l_2^2(b_1+\beta_1)^2}{b_2l_2(b_1+\beta_1) - 2c^2(l_2-1)}t \\ &\quad + \frac{\{c(\zeta_2+l_2m_2)+c^2q_{20}\}(b_1+\beta_1)b_2l_2 - b_2l_2c(c\alpha_1+c^2q_{20})(l_2-1) - c(\zeta_2+l_2m_2) - c^2q_{20}}{b_2l_2(b_1+\beta_1) - 2c^2(l_2-1)} \\ &= Jt + D.\end{aligned}$$

Taking $\frac{1}{(\overline{z}_1-\underline{z}_1)^{\rho_1}}$ common from the first order condition, we get

$$\begin{aligned} &\int_{q_{10}}^{\overline{q}} \Big\{\rho_1(\overline{z}_1 - At + B)^{\rho_1-1}(\overline{z}_1' - Jt - D)(\alpha_1+\beta_1 t - m_1) + (\overline{z}_1 - At + B)^{\rho_1} t\Big\}\mathrm{d}t \\ &- \int_{q_{10}}^{\overline{q}_1} \frac{\rho_1 \triangle z_1'}{(\overline{z}_1 - \underline{z}_1)}(\alpha_1+\beta_1 t - m_1)(\overline{z}_1 - At + B)^{\rho_1}\mathrm{d}t = 0 \\ \textit{or,}\quad &\int_{q_{10}}^{\overline{q}} \rho_1(\overline{z}_1 - At+B)^{\rho_1-1}(\overline{z}_1' - D)(\alpha_1 - m_1) - \rho_1(\overline{z}_1 - At + B)^{\rho_1-1}J(\alpha_1-m_1)t \\ &+\rho_1(\overline{z}_1 - At+B)^{\rho_1-1}(\overline{z}_1'-D)\beta_1 t - \rho_1(\overline{z}_1 - At+B)^{\rho_1-1}J\beta_1 t^2 + (\overline{z}_1 - At + B)^{\rho_1} t\mathrm{d}t \\ &- \int_{q_{10}}^{\overline{q}_1} \frac{\rho_1 \triangle z_1'}{(\overline{z}_1 - \underline{z}_1)}(\alpha_1+\beta_1 t - m_1)(\overline{z}_1 - At + B)^{\rho_1}\mathrm{d}t = 0 \\ \textit{or,}\quad &(\overline{z}_1' - D)(\alpha_1 - m_1)\int_{q_{10}}^{\overline{q}} \rho_1(\overline{z}_1 - At+B)^{\rho_1-1}\mathrm{d}t - J(\alpha_1 - m_1)\int_{q_{10}}^{\overline{q}_1}\rho_1(\overline{z}_1 - At+B)^{\rho_1-1}t\mathrm{d}t \\ &+(\overline{z}_1'-D)\beta_1\int_{q_{10}}^{\overline{q}_1}\rho_1(\overline{z}_1 - At+B)^{\rho_1-1}\mathrm{d}t - J\beta_1\int_{q_{10}}^{\overline{q}_1}\rho_1(\overline{z}_1 - At+B)^{\rho_1-1}t^2\mathrm{d}t \\ &+\int_{q_{10}}^{\overline{q}_1}(\overline{z}_1 - At+B)^{\rho_1}t\mathrm{d}t - \int_{q_{10}}^{\overline{q}_1} \frac{\rho_1 \triangle z_1'}{(\overline{z}_1 - \underline{z}_1)}(\alpha_1+\beta_1 t - m_1)(\overline{z}_1 - At + B)^{\rho_1}\mathrm{d}t = 0 \end{aligned}$$

$$\begin{aligned} \textit{or,}\quad &(\overline{z}_1' - D)(\alpha_1 - m_1)\int_{q_{10}}^{\overline{q}} \rho_1(\overline{z}_1 - At+B)^{\rho_1-1}\mathrm{d}t - (J(\alpha_1 - m_1) - (\overline{z}_1' - D)\beta_1) \\ &\times\int_{q_{10}}^{\overline{q}_1}\rho_1(\overline{z}_1 - At+B)^{\rho_1-1}t\mathrm{d}t - J\beta_1\int_{q_{10}}^{\overline{q}_1}\rho_1(\overline{z}_1 - At+B)^{\rho_1-1}t^2\mathrm{d}t \\ &+\int_{q_{10}}^{\overline{q}_1}(\overline{z}_1 - At+B)^{\rho_1}t\mathrm{d}t - \int_{q_{10}}^{\overline{q}_1} \frac{\rho_1 \triangle z_1'}{(\overline{z}_1 - \underline{z}_1)}(\alpha_1+\beta_1 t - m_1)(\overline{z}_1 - At + B)^{\rho_1}\mathrm{d}t = 0 \end{aligned}$$

Now, ignoring the coefficients, we solve the integration. First Term:

$$\int_{q_{10}}^{\overline{q}} \rho_1(\overline{z}_1 - At + B)^{\rho_1-1}\mathrm{d}t = -\frac{1}{A}\int_{q_{10}}^{\overline{q}}\{(\overline{z}_1 - At+B)^{\rho_1}\}'\mathrm{d}t = \frac{\nu_2^{\rho_1} - \nu_1^{\rho_1}}{A}.$$

Second Term:

$$\int_{q_{10}}^{\bar{q}_1} \rho_1(\bar{z}_1 - At + B)^{\rho_1 - 1} t \mathrm{d}t = -\frac{1}{A}\int_{q_{10}}^{\bar{q}_1} t\{(\bar{z}_1 - At + B)^{\rho_1}\}' \mathrm{d}t$$
$$= -\frac{1}{A}\left[\bar{q}_1 \nu_1^{\rho_1} - q_{10}\nu_2^{\rho_1} - \int_{q_{10}}^{\bar{q}_1} (\bar{z}_1 - At + B)^{\rho_1} \mathrm{d}t\right] = \frac{q_{10}\nu_2^{\rho_1} - \bar{q}_1 \nu_1^{\rho_1}}{A} + \frac{1}{A}\int_{q_{10}}^{\bar{q}_1} (\bar{z}_1 - At + B)^{\rho_1} \mathrm{d}t$$
$$= \frac{q_{10}\nu_2^{\rho_1} - \bar{q}_1 \nu_1^{\rho_1}}{A} + \frac{\nu_2^{1+\rho_1} - \nu_1^{1+\rho_1}}{A^2(1+\rho_1)}.$$

Third Term

$$\int_{q_{10}}^{\bar{q}_1} \rho_1(\bar{z}_1 - At + B)^{\rho_1 - 1} t^2 \mathrm{d}t = -\frac{1}{A}\int_{q_{10}}^{\bar{q}_1} t^2\{(\bar{z}_1 - At + B)^{\rho_1}\}' \mathrm{d}t$$
$$= -\frac{1}{A}\left[\bar{q}_1^2 \nu_1^{\rho_1} - q_{10}^2\nu_2^{\rho_1} - 2\int_{q_{10}}^{\bar{q}_1} t(\bar{z}_1 - At + B)^{\rho_1} \mathrm{d}t\right] = \frac{q_{10}^2\nu_2^{\rho_1} - \bar{q}_1^2 \nu_1^{\rho_1}}{A} + \frac{2}{A}\int_{q_{10}}^{\bar{q}_1} t(\bar{z}_1 - At + B)^{\rho_1} \mathrm{d}t.$$

Let $(\bar{z}_1 - At + B)^{\rho_1} = x$, then $\mathrm{d}t = -\frac{x^{\frac{1}{\rho_1}-1}}{A\rho_1}\mathrm{d}x$, which allows us to solve the integration in the last expression:

$$\int_{q_{10}}^{\bar{q}_1} t(\bar{z}_1 - At + B)^{\rho_1} \mathrm{d}t = \int_{\nu_2^{\rho_1}}^{\nu_1^{\rho_1}} x\frac{(\bar{z}_1 + B) - x^{\frac{1}{\rho_1}}}{A}\left(-\frac{x^{\frac{1}{\rho_1}-1}}{A\rho_1}\right)\mathrm{d}x$$
$$= -\frac{(\bar{z}_1 + B)}{A^2\rho_1}\int_{\nu_2^{\rho_1}}^{\nu_1^{\rho_1}} x^{\frac{1}{\rho_1}}\mathrm{d}x + \frac{1}{A^2\rho_1}\int_{\nu_2^{\rho_1}}^{\nu_1^{\rho_1}} x^{\frac{2}{\rho_1}}\mathrm{d}x = -\frac{(\bar{z}_1 + B)\left(\{\nu_1^{\rho_1}\}^{1+\frac{1}{\rho_1}} - \{\nu_2^{\rho_1}\}^{1+\frac{1}{\rho_1}}\right)}{A^2\rho_1(1+\frac{1}{\rho_1})}$$
$$+\frac{\{\nu_1^{\rho_1}\}^{1+\frac{2}{\rho_1}} - \{\nu_2^{\rho_1}\}^{1+\frac{2}{\rho_1}}}{A^2\rho_1(\frac{2}{\rho_1}+1)} = -\frac{(\bar{z}_1 + B)\left(\nu_1^{1+\rho_1} - \nu_2^{1+\rho_1}\right)}{A^2(1+\rho_1)} + \frac{\nu_1^{2+\rho_1} - \nu_2^{2+\rho_1}}{A^2(2+\rho_1)}$$

Therefore, the third term becomes

$$\int_{q_{10}}^{\bar{q}_1} \rho_1(\bar{z}_1 - At + B)^{\rho_1 - 1} t^2 \mathrm{d}t = \frac{q_{10}^2\nu_2^{\rho_1} - \bar{q}_1^2 \nu_1^{\rho_1}}{A} - \frac{2(\bar{z}_1 + B)\left(\nu_1^{1+\rho_1} - \nu_2^{1+\rho_1}\right)}{A^3(1+\rho_1)} + \frac{2(\nu_1^{2+\rho_1} - \nu_2^{2+\rho_1})}{A^3(2+\rho_1)}$$

Fourth term: (following the derivation for the third term)

$$\int_{q_{10}}^{\bar{q}_1} (\bar{z}_1 - At + B)^{\rho_1} t \mathrm{d}t = \frac{(\bar{z}_1 + B)(\nu_2^{1+\rho_1} - \nu_1^{1+\rho_1})}{A^2(1+\rho_1)} + \frac{\nu_1^{2+\rho_1} - \nu_2^{2+\rho_1}}{A^2(2+\rho_1)}.$$

The Fifth Term:

$$\int_{q_{10}}^{\overline{q}_1}\frac{\rho_1\triangle z_1'}{(\overline{z}_1-\underline{z}_1)}(\alpha_1+\beta_1 t-m_1)(\overline{z}_1-At+B)^{\rho_1}\mathrm{d}t=\int_{q_{10}}^{\overline{q}_1}\Big\{\frac{\rho_1\triangle z_1'(\alpha_1-m_1)(\overline{z}_1-At+B)^{\rho_1}}{(\overline{z}_1-\underline{z}_1)}$$
$$+\frac{\rho_1\triangle z_1'\beta_1(\overline{z}_1-At+B)^{\rho_1}t}{(\overline{z}_1-\underline{z}_1)}\Big\}\mathrm{d}t=\frac{\triangle z_1'(\alpha_1-m_1)}{(\overline{z}_1-\underline{z}_1)}\int_{q_{10}}^{\overline{q}_1}\rho_1(\overline{z}_1-At+B)^{\rho_1}\mathrm{d}t$$
$$+\frac{\rho_1\triangle z_1'\beta_1}{(\overline{z}_1-\underline{z}_1)}\int_{q_{10}}^{\overline{q}_1}(\overline{z}_1-At+B)^{\rho_1}t\mathrm{d}t$$
$$=\frac{\triangle z_1'\rho_1(\alpha_1-m_1)(\nu_2^{1+\rho_1}-\nu_1^{1+\rho_1})}{A(\overline{z}_1-\underline{z}_1)(1+\rho_1)}-\frac{\triangle z_1'\rho_1\beta_1(\overline{z}_1+B)(\nu_1^{1+\rho_1}-\nu_2^{1+\rho_1})}{A^2(\overline{z}_1-\underline{z}_1)(1+\rho_1)}+\frac{\triangle z_1'\rho_1\beta_1(\nu_1^{2+\rho_1}-\nu_2^{2+\rho_1})}{A^2(\overline{z}_1-\underline{z}_1)(2+\rho_1)}$$

Now, putting all the terms together we get the following first order condition for β_1:

$$\frac{(\overline{z}_1'-D)(\alpha_1-m_1)(\nu_2^{\rho_1}-\nu_1^{\rho_1})}{A}-\{J(\alpha_1-m_1)-(\overline{z}_1'-D)\beta_1\}\left\{\frac{q_{10}\nu_2^{\rho_1}-\overline{q}_1\nu_1^{\rho_1}}{A}+\frac{\nu_2^{1+\rho_1}-\nu_1^{1+\rho_1}}{A^2(1+\rho_1)}\right\}$$
$$-J\beta_1\left\{\frac{q_{10}^2\nu_2^{\rho_1}-\overline{q}_1^2\nu_1^{\rho_1}}{A}-\frac{2(\overline{z}_1+B)\left(\nu_1^{1+\rho_1}-\nu_2^{1+\rho_1}\right)}{A^3(1+\rho_1)}+\frac{2(\nu_1^{2+\rho_1}-\nu_2^{2+\rho_1})}{A^3(2+\rho_1)}\right\}$$
$$+\left\{\frac{(\overline{z}_1+B)(\nu_2^{1+\rho_1}-\nu_1^{1+\rho_1})}{A^2(1+\rho_1)}+\frac{\nu_1^{2+\rho_1}-\nu_2^{2+\rho_1}}{A^2(2+\rho_1)}\right\}-\left\{\frac{\triangle z_1'\rho_1(\alpha_1-m_1)(\nu_2^{1+\rho_1}-\nu_1^{1+\rho_1})}{A(\overline{z}_1-\underline{z}_1)(1+\rho_1)}\right.$$
$$\left.-\frac{\triangle z_1'\rho_1\beta_1(\overline{z}_1+B)(\nu_1^{1+\rho_1}-\nu_2^{1+\rho_1})}{A^2(\overline{z}_1-\underline{z}_1)(1+\rho_1)}+\frac{\triangle z_1'\rho_1\beta_1(\nu_1^{2+\rho_1}-\nu_2^{2+\rho_1})}{A^2(\overline{z}_1-\underline{z}_1)(2+\rho_1)}\right\}=0.$$

Taking $\frac{1}{A}$ common and re arranging the terms gives us

$$(\overline{z}_1'-D)(\alpha_1-m_1)(\nu_2^{\rho_1}-\nu_1^{\rho_1})-\{J(\alpha_1-m_1)-(\overline{z}_1'-D)\beta_1\}(q_{10}\nu_2^{\rho_1}-\overline{q}_1\nu_1^{\rho_1})$$
$$+\left\{\frac{\overline{z}_1+B}{A}-\frac{\{J(\alpha_1-m_1)-(\overline{z}_1'-D)\beta_1\}}{A}-\frac{2J\beta_1(\overline{z}_1+B)}{A^2}+\frac{\triangle z_1'\rho_1\{\beta_1(\overline{z}_1+B)-A(\alpha_1-m_1)\}}{A(\overline{z}_1-\underline{z}_1)}\right\}$$
$$\times\left(\frac{\nu_2^{1+\rho_1}-\nu_1^{1+\rho_1}}{(1+\rho_1)}\right)-J\beta_1(q_{10}^2\nu_2^{\rho_1}-\overline{q}_1^2\nu_1^{\rho_1})-\left(\frac{\triangle z_1'\rho_1\beta_1}{\overline{z}_1-\underline{z}_1}+2J\beta_1-1\right)\left(\frac{\nu_1^{2+\rho_1}-\nu_2^{2+\rho_1}}{A^2(2+\rho_1)}\right)=0.$$

Then when we substitute $\nu_1=0$, and taking $\nu_2^{\rho_1}(\neq 0)$ common, we get

$$(\overline{z}_1'-D)(\alpha_1-m_1)-q_{10}\{J(\alpha_1-m_1)-(\overline{z}_1'-D)\beta_1\}$$
$$+\left\{\frac{\overline{z}_1+B}{A}-\frac{\{J(\alpha_1-m_1)-(\overline{z}_1'-D)\beta_1\}}{A}-\frac{2J\beta_1(\overline{z}_1+B)}{A^2}+\frac{\triangle z_1'\rho_1\{\beta_1(\overline{z}_1+B)-A(\alpha_1-m_1)\}}{A(\overline{z}_1-\underline{z}_1)}\right\}$$
$$\times\frac{\nu_2}{(1+\rho_1)}-J\beta_1 q_{10}^2+\left(\frac{\triangle z_1'\rho_1\beta_1}{\overline{z}_1-\underline{z}_1}+2J\beta_1-1\right)\frac{\nu_2^2}{A^2(2+\rho_1)}=0.,$$

□

Lemma 3. *In equilibrium the fixed price of purchasing advertisements from each publishers is γ_1 and γ_2, which are respectively given by*

$$\begin{aligned}\gamma_1 \quad = \quad & \frac{(\theta_1^* - \alpha_1)^2(b_2+\beta_2)}{b_1+\beta_1} + \frac{c(\theta_1^*-\alpha_1)(\overline{\theta}_2-\alpha_2)(3-(b_1+\beta_1))}{(b_1+\beta_1)(b_2+\beta_2)-c^2} + \frac{(\overline{\theta}_2-\alpha_2)^2(b_1+\beta_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2} - \frac{(\theta_1^*-\alpha_1)^2}{2(b_1+\beta_1)} \\ & - \frac{c^2(\theta_1^*-\alpha_1)^2}{(b_1+\beta_1)(b_2+\beta_2)-c^2} \times \left(1 - \frac{1}{b_1+\beta_1}\right) + \frac{1}{2}\left(\frac{(\overline{\theta}_2-\alpha_2)(b_1+\beta_1)+c(\theta_1^*-\alpha_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2}\right)^2 (1-(b_2+\beta_2)) \\ & -\theta_1^* q_{10} + \frac{b_1}{2}q_{10}^2 + \frac{(\overline{\theta}_2-\alpha_2+cq_{10})^2}{2(b_2+\beta_2)}\end{aligned}$$

$$\begin{aligned}\gamma_2 \quad = \quad & \left\{(\overline{\theta}_1-\alpha_1)\frac{3(\overline{\theta}_1-\alpha_1)(b_2+\beta_2)(b_1+\beta_1)+c(b_1+\beta_1)[2(\theta_2^*-\alpha_2)+(b_2+\beta_2)q_{20}]-c^2(\overline{\theta}_1-\alpha_1)-c^3q_{20}}{2(b_1+\beta_1)(b_2+\beta_2)-2c^2}\right\} \\ & \left[\frac{c(\theta_2^*-\alpha_2)}{(b_1+\beta_1)(b_2+\beta_2)} + \frac{c^2(\overline{\theta}_1-\alpha_1)(b_2+\beta_2)+c^2(\theta_2^*-\alpha_2)}{(b_1+\beta_1)(b_2+\beta_2)\{(b_1+\beta_1)(b_2+\beta_2)-c^2\}} - \frac{cq_{20}}{b_1+\beta_1}\right] \\ & + (\theta_2^*-\alpha_2)\left[\frac{\theta_2^*-\alpha_2}{b_2+\beta_2} + \frac{c(\overline{\theta}_1-\alpha_1)(b_2+\beta_2)+c^2(\theta_2^*-\alpha_2)}{(b_2+\beta_2)\{(b_1+\beta_1)(b_2+\beta_2)-c^2\}}\right] - \theta_2^*q_{20} - \frac{b_2+\beta_2}{2}\left[\frac{\theta_2^*-\alpha_2}{b_2+\beta_2}\right. \\ & \left. + \frac{c(\overline{\theta}_1-\alpha_1)(b_2+\beta_2)+c^2(\theta_2^*-\alpha_2)}{(b_2+\beta_2)\{(b_1+\beta_1)(b_2+\beta_2)-c^2\}}\right]^2 + \frac{b_2}{2}q_{20}^2 + c\left[\frac{\theta_2^*-\alpha_2}{b_2+\beta_2} + \frac{c(\overline{\theta}_1-\alpha_1)(b_2+\beta_2)+c^2(\theta_2^*-\alpha_2)}{(b_2+\beta_2)\{(b_1+\beta_1)(b_2+\beta_2)-c^2\}}\right] \\ & \left[\frac{(\overline{\theta}_1-\alpha_1)(b_2+\beta_2)+c(\theta_2^*-\alpha_2)}{(b_1+\beta_1)(b_2+\beta_2)-c^2}\right] - \frac{c(\overline{\theta}_1-\alpha_1+cq_{20})q_{20}}{b_1+\beta_1} + (\zeta_2+m_2)q_{20} \\ & + \frac{1}{2}\left(b_2(2l_2-1) - \frac{2c^2(l_2-1)}{b_1+\beta_1}\right)q_{20}^2\end{aligned}$$

Proof. The lowest type for $P1$ amongst all those who buy at least q_{10}, is $(\theta_1^*, \overline{\theta}_2)$. Recall the definition of θ_1^* from previous section. Let $w_2(q_2(\theta_1^*, \overline{\theta}_2); \theta_1^*, \overline{\theta}_2)$ be the utility that the type $(\theta_1^*, \overline{\theta}_2)$ gets by consuming optimal q_2 and only q_{10}. Notice that given the type in other dimension, the agent will always have interior solution for q_2. Then if we recall $W_1(\theta_1^*, \overline{\theta}_2; x)$ to be the indirect utility of consuming optimal pair of (q_1, q_2) when P1 charges $\gamma_1 = x$, then the difference between the two $W_1(\cdot, \cdot; x) - w_2(\cdot; \cdot, \cdot)$ is the extra utility from participating in contract with P1. Since this type is the least favorable one from the perspective of P1, optimal γ_1 should be such that it takes away all the extra utility for such type that accrues from participating with P1. Therefore we get

$$\gamma_2 = \arg\min_x \{W_1(\theta_1^*, \overline{\theta}_2; x) - w_2(q_2(\theta_1^*, \overline{\theta}_2); \theta_1^*, \overline{\theta}_2)\}.$$

When only q_{10} is consumed then the first order condition gives us that the optimal $q_2(\theta^*;q_{10}) = \frac{\overline{\theta}_2-\alpha_2+cq_{10}}{b_2+\beta_2}$, where θ^* is the type pair we are concerned with. Therefore,

$$\begin{aligned}
w_2(q_2(\theta^*;q_{10});\theta^*) &= \theta_1^* q_{10} + (\overline{\theta}_2 - \alpha_2 + cq_{10})\left(\frac{\overline{\theta}_2 - \alpha_2 + cq_{10}}{b_2+\beta_2}\right) - \frac{(b_1+\beta_2)}{2}\left(\frac{\overline{\theta}_2 - \alpha_2 + cq_{10}}{b_2+\beta_2}\right)^2 \\
&\quad -\frac{b_1}{2}q_{10}^2 - \gamma_2 \\
&= \theta_1^* q_{10} - \frac{b_1}{2}q_{10}^2 - \gamma_2 + \frac{(\overline{\theta}_2 - \alpha_2 + cq_{10})^2}{2(b_2+\beta_2)}
\end{aligned}$$

However, without any such restrictions, optimal consumption pairs are

$$q_2(\theta^*) = \frac{\overline{\theta}_2 - \alpha_2 + cq_1(\theta^*)}{b_2+\beta_2}; \qquad q_1(\theta^*) = \frac{\theta_1^* - \alpha_1 + cq_2(\theta^*)}{b_1+\beta_1},$$

which when solved simultaneously, gives us

$$\begin{aligned}
q_2(\theta^*) &= \frac{(\overline{\theta}_2 - \alpha_2)(b_1+\beta_1) + c(\theta_1^* - \alpha_1)}{(b_1+\beta_1)(b_2+\beta_2) - c^2} \\
q_1(\theta^*) &= \frac{\theta_1^* - \alpha_1}{b_1+\beta_1} + \frac{c}{b_1+\beta_1}\left[\frac{(\overline{\theta}_2 - \alpha_2)(b_1+\beta_1) + c(\theta_1^* - \alpha_1)}{(b_1+\beta_1)(b_2+\beta_2) - c^2}\right].
\end{aligned}$$

Observe that we intend to minimize the difference between W_1 and w_2 by choosing x, which appears linearly only in W_1 we can choose the γ_1 such that ignoring x in W_1 $\gamma_1 = W_1 - w_2$. To get W_1 we substitute all the expression back into the indirect utility function to get:

$$
\begin{aligned}
\gamma_1 &= (\theta_1^* - \alpha_1)q_1(\theta^*) + (\overline{\theta}_2 - \alpha_2)q_2(\theta^*) - \frac{b_1+\beta_1}{q}{}_1(\theta^*)^2 - \frac{b_2+\beta_2}{2}q_2(\theta^*)^2 \\
&\quad -\gamma_2 + cq_1(\theta^*)q_2(\theta^*) - \theta_1^* q_{10} + \frac{b_1}{2}q_{10}^2 + \gamma_2 - \frac{(\overline{\theta}_2 - \alpha_2 + cq_{10})^2}{2(b_2+\beta_2)} \\
&= \frac{(\theta_1^* - \alpha_1)^2}{b_1+\beta_1} + \frac{c(\theta_1^* - \alpha_1)(\overline{\theta}_2 - \alpha_2)}{(b_1+\beta_1)(b_2+\beta_2)-c^2} + \frac{c^2(\theta_1^* - \alpha_1)^2}{(b_1+\beta_1)\{(b_1+\beta_1)(b_2+\beta_2)-c^2\}} \\
&\quad + \frac{(\overline{\theta}_2 - \alpha_2)^2(b_1+\beta_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2} + \frac{c(\overline{\theta}_2 - \alpha_2)(\theta_1^* - \alpha_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2} - \frac{b_1+\beta_1}{2}\Bigg[\left(\frac{\theta_1^* - \alpha_1}{b_1+\beta_1}\right)^2 \\
&\quad + \frac{2c(\theta_1^* - \alpha_1)}{b_1+\beta_1}\frac{(\overline{\theta}_2 - \alpha_2)(b_1+\beta_1) + c(\theta_1^* - \alpha_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2} + \frac{c^2}{(b_1+\beta_1)^2}\left(\frac{(\overline{\theta}_2 - \alpha_2)(b_1+\beta_1) + c(\theta_1^* - \alpha_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2}\right)^2 \\
&\quad - \frac{b_2+\beta_2}{2}\left(\frac{(\overline{\theta}_2 - \alpha_2)(b_1+\beta_1) + c(\theta_1^* - \alpha_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2}\right)^2 + \frac{c(\theta_1^* - \alpha_1)}{b_1+\beta_1}\frac{(\overline{\theta}_2 - \alpha_2)(b_1+\beta_1) + c(\theta_1^* - \alpha_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2} \\
&\quad \frac{c^2}{b_1+\beta_1}\left(\frac{(\overline{\theta}_2 - \alpha_2)(b_1+\beta_1) + c(\theta_1^* - \alpha_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2}\right)^2 - \theta_1^* q_{10} + \frac{b_1}{2}q_{10}^2 + \frac{(\overline{\theta}_2 - \alpha_2 + cq_{10})^2}{2(b_2+\beta_2)}
\end{aligned}
$$

$$
\begin{aligned}
\gamma_1 &= \frac{(\theta_1^* - \alpha_1)^2(b_2+\beta_2)}{b_1+\beta_1} + \frac{3c(\theta_1^* - \alpha_1)(\overline{\theta}_2 - \alpha_2)}{(b_1+\beta_1)(b_2+\beta_2)-c^2} + \frac{(\overline{\theta}_2 - \alpha_2)^2(b_1+\beta_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2} - \frac{(\theta_1^* - \alpha_1)^2}{2(b_1+\beta_1)} \\
&\quad - \frac{c(\theta_1^* - \alpha_1)(\overline{\theta}_2 - \alpha_2)(b_1+\beta_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2} - \frac{c^2(\theta_1^* - \alpha_1)^2}{(b_1+\beta_1)(b_2+\beta_2)-c^2}\left(1 - \frac{1}{b_1+\beta_1}\right) \\
&\quad + \frac{1}{2}\left(\frac{(\overline{\theta}_2 - \alpha_2)(b_1+\beta_1) + c(\theta_1^* - \alpha_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2}\right)^2(1-(b_2+\beta_2)) - \theta_1^* q_{10} + \frac{b_1}{2}q_{10}^2 + \frac{(\overline{\theta}_2 - \alpha_2 + cq_{10})^2}{2(b_2+\beta_2)} \\
&= \frac{(\theta_1^* - \alpha_1)^2(b_2+\beta_2)}{b_1+\beta_1} + \frac{c(\theta_1^* - \alpha_1)(\overline{\theta}_2 - \alpha_2)(3-(b_1+\beta_1))}{(b_1+\beta_1)(b_2+\beta_2)-c^2} + \frac{(\overline{\theta}_2 - \alpha_2)^2(b_1+\beta_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2} \\
&\quad - \frac{(\theta_1^* - \alpha_1)^2}{2(b_1+\beta_1)} - \frac{c^2(\theta_1^* - \alpha_1)^2}{(b_1+\beta_1)(b_2+\beta_2)-c^2}\left(1 - \frac{1}{b_1+\beta_1}\right) \\
&\quad + \frac{1}{2}\left(\frac{(\overline{\theta}_2 - \alpha_2)(b_1+\beta_1) + c(\theta_1^* - \alpha_1)}{(b_1+\beta_1)(b_2+\beta_2)-c^2}\right)^2(1-(b_2+\beta_2)) - \theta_1^* q_{10} + \frac{b_1}{2}q_{10}^2 + \frac{(\overline{\theta}_2 - \alpha_2 + cq_{10})^2}{2(b_2+\beta_2)}
\end{aligned}
$$

Now, we shall find the optimal γ_2. From the characterization of the threshold θ_2^*, we know that an agents with type pair $(\overline{\theta}_1, \theta_2^*)$ are least prepared to buy more than q_{20} from Principal 2. Hence, γ_2, can be at most the infra gain in utility from consuming the optimal pair from both the principals and consuming only q_{20} from P2. If we let

$w_1(q_1(\overline{\theta}_1,\theta_2^*);\overline{\theta}_1,\theta_2^*)$ to be the utility that this agent gets when he consumes optimal q_1 -which could be greater or equal to q_{10}- and only q_{20}, and $W_2(q_1,q_2;\overline{\theta}_1,\theta_2^*)$ to be the utility when he consumes optimal amount of both q_1 and q_2. Then the optimal fixed cost of purchasing $q_2 > q_{20}$, γ_2 must take away this rent, hence, hence $\underline{s}(\cdot)(\gamma_2) - w_2(\cdot) = 0$, so

$$\gamma_2 = W_2(\cdot) - w_1(\cdot).$$

Since

$$w_1(q_1(\overline{\theta}_1,\theta_2^*);\overline{\theta}_1,\theta_2^*) = \max_{\hat{q}_1} u(\hat{q}_1,q_{20};\overline{\theta}_1,\theta_2^*) - \gamma_1 - \alpha_1\hat{q}_1 - \frac{\beta_1}{2}\hat{q}_1^2.$$

the optimal q_1 is given by the first order necessary condition, and is

$$q_1 = \frac{\overline{\theta}_1 - \alpha_1 + cq_{20}}{b_1+\beta_1}.$$

Hence,

$$\begin{aligned} w_1(q_1(\overline{\theta}_1,\theta_2^{**});\overline{\theta}_1,\theta_2^*) &= (\overline{\theta}_1-\alpha_1)\frac{\overline{\theta}_1 - \alpha_1 + cq_{20}}{b_1+\beta_1} + \theta_2^* q_{20} - \frac{b_1+\beta_1}{2}\left(\frac{\overline{\theta}_1 - \alpha_1 + cq_{20}}{b_1+\beta_1}\right)^2 \\ &\quad -\frac{b_2}{2}q_{20}^2 + cq_{20}\frac{\overline{\theta}_1 - \alpha_1 + cq_{20}}{b_1+\beta_1} - \gamma_1 - \frac{\beta_1}{2}\hat{q}_1^2 \\ &= \frac{(\overline{\theta}_1-\alpha_1)^2}{b_1+\beta_1} + \left(\frac{2c(\overline{\theta}_1-\alpha_1)}{b_1+\beta_1} + \theta_2^*\right)q_{20} - \frac{\left(\overline{\theta}_1 - \alpha_1 + cq_{20}\right)^2}{2(b_1+\beta_1)} \\ &\quad + \left(\frac{c^2}{b_1+\beta_1} - \frac{b_2}{2}\right)q_{20}^2 - \gamma_1. \end{aligned}$$

Now, to find $\underline{s}_2(\cdot)$ we start from the demand for q_1 and q_2 derived from the usual first order necessary condition of optimization, where

$$W_2(q_1,q_2;\overline{\theta}_1,\theta_2^*) = \max_{\hat{q}_1,\hat{q}_2}\left[u(\hat{q}_1,\hat{q}_2;\overline{\theta}_1,\theta_2^*) - \gamma_1 - \alpha_1\hat{q}_1 - \frac{\beta_1}{2}\hat{q}_1^2 - \alpha_2\hat{q}_2 - \frac{\beta_2}{2}\hat{q}_2^2\right].$$

which gives[2]

$$
\begin{aligned}
q_1 &= \frac{\overline{\theta}_1 - \alpha_1 + cq_2}{b_1 + \beta_1}; \qquad q_2 = \frac{\theta_2^* - \alpha_2 + cq_1}{b_2 + \beta_2} \\
q_1(q_{20}) &= \frac{\overline{\theta}_1 - \alpha_1 + cq_{20}}{b_1 + \beta_1}
\end{aligned}
$$

If we simultaneously solve these two equations we get the following final demand as a function of pricing rule, and using $\theta^{**} = (\overline{\theta}_2, \theta_2^*)$:

$$
\begin{aligned}
q_2(\theta^{**}) &= \frac{\theta_2^* - \alpha_2}{b_2 + \beta_2} + \frac{c(\overline{\theta}_1 - \alpha_1)(b_2 + \beta_2) + c^2(\theta_2^* - \alpha_2)}{(b_2 + \beta_2)\{(b_1 + \beta_1)(b_2 + \beta_2) - c^2\}} \\
q_1(\theta^{**}) &= \frac{(\overline{\theta}_1 - \alpha_1)(b_2 + \beta_2) + c(\theta_2^* - \alpha_2)}{(b_1 + \beta_1)(b_2 + \beta_2) - c^2}
\end{aligned}
$$

$$
\begin{aligned}
\hat{\hat{\gamma}}_2 &= (\overline{\theta}_1 - \alpha_1)q_1(\theta^{**}) + (\theta_2^* - \alpha_2)q_2 - \frac{b_1 + \beta_1}{2} q_1(\theta^{**})^2 - \frac{b_2 + \beta_2}{2} q_2(\theta^{**})^2 + cq_1(\theta^{**})q_2(\theta^{**}) \\
&\quad - \gamma_1 - (\overline{\theta}_1 - \alpha_1)q_1(q_{20}) - \theta_2^* q_{20} + \frac{b_1 + \beta_1}{2} q_1(q_{20})^2 + \frac{b_2}{2} q_{20}^2 - cq_1(q_{20})q_{20} + \gamma_1 \\
&= (\overline{\theta}_1 - \alpha_1)(q_1(\theta^{**}) - q_1(q_{20})) + (\theta_2^* - \alpha_2)q_2(\theta^{**}) - \theta_2^* q_{20} - \frac{b_1 + \beta_1}{2}(q_1(\theta^{**})^2 - q_1(q_{20})^2) \\
&\quad - \frac{b_2 + \beta_2}{2} q_2(\theta^{**})^2 + \frac{b_2}{2} q_{20}^2 + c(q_1(\theta^{**})q_2(\theta^{**}) - q_1(q_{20})q_{20}) \\
&= \left\{ (\overline{\theta}_1 - \alpha_1) - \frac{b_1 + \beta_1}{2}(q_1(\theta^{**}) + q_1(q_{20})) \right\} (q_1(\theta^{**}) - q_1(q_{20})) + (\theta_2^* - \alpha_2)\, q_2(\theta^{**}) - \theta_2^* q_{20} \\
&\quad - \frac{b_2 + \beta_2}{2} q_2(\theta^{**})^2 + \frac{b_2}{2} q_{20}^2 + c\{q_1(\theta^{**})q_2(\theta^{**}) - q_1(q_{20})q_{20}\}
\end{aligned}
$$

And since $\gamma_2 = \hat{\hat{\gamma}}_2 - \hat{\gamma}_2$, we get

$$
\begin{aligned}
\gamma_2 &= \left\{ (\overline{\theta}_1 - \alpha_1) - \frac{b_1 + \beta_1}{2}(q_1(\theta^{**}) + q_1(q_{20})) \right\} (q_1(\theta^{**}) - q_1(q_{20})) + (\theta_2^* - \alpha_2)\, q_2(\theta^{**}) - \theta_2^* q_{20} \\
&\quad - \frac{b_2 + \beta_2}{2} q_2(\theta^{**})^2 + \frac{b_2}{2} q_{20}^2 + c\{q_1(\theta^{**})q_2(\theta^{**}) - q_1(q_{20})q_{20}\} + (\zeta_2 + m_2)q_{20} \\
&\quad + \frac{1}{2}\left(b_2(2l_2 - 1) - \frac{2c^2(l_2 - 1)}{b_1 + \beta_1} \right) q_{20}^2
\end{aligned}
$$

[2] Note that while writing $\underline{s}_2(\cdot)$ we have ignored γ_2, as it is automatically taken care of in the definition of optimal γ_2.

Some calculation yields the following:

$$q_1(\theta^{**}) - q_1(q_{20}) = \frac{c(q_2(\theta^{**}) - q_{20})}{b_1+\beta_1} =$$
$$\left[\frac{(\theta_2^* - \alpha_2)}{(b_1+\beta_1)(b_2+\beta_2)} + \frac{c^2(\overline{\theta}_1 - \alpha_1)(b_2+\beta_2) + c^2(\theta_2^* - \alpha_2)}{(b_1+\beta_1)(b_2+\beta_2)\{(b_1+\beta_1)(b_2+\beta_2) - c^2\}} - \frac{cq_{20}}{b_1+\beta_1}\right]$$

$$\frac{b_1+\beta_1}{2}(q_1(\theta^{**}) + q_1(q_{20})) =$$
$$\frac{3(\overline{\theta}_1 - \alpha_1)(b_2+\beta_2)(b_1+\beta_1) + c(b_1+\beta_1)[2(\theta_2^* - \alpha_2) + (b_2+\beta_2)q_{20}] - c^2(\overline{\theta}_1 - \alpha_1) - c^3 q_{20}}{2(b_1+\beta_1)(b_2+\beta_2) - 2c^2}$$

$$q_1(\theta^{**})q_2(\theta^{**}) - q_1(q_{20})q_{20} = \left[\frac{\theta_2^* - \alpha_2}{b_2+\beta_2} + \frac{c(\overline{\theta}_1 - \alpha_1)(b_2+\beta_2) + c^2(\theta_2^* - \alpha_2)}{(b_2+\beta_2)\{(b_1+\beta_1)(b_2+\beta_2) - c^2\}}\right]$$
$$\times \left[\frac{(\overline{\theta}_1 - \alpha_1)(b_2+\beta_2) + c(\theta_2^* - \alpha_2)}{(b_1+\beta_1)(b_2+\beta_2) - c^2}\right] - \frac{(\overline{\theta}_1 - \alpha_1 + cq_{20})q_{20}}{b_1+\beta_1}$$

$$\begin{aligned}
\gamma_2 \;=\; & \Bigg\{(\overline{\theta}_1 - \alpha_1) \\
& \frac{3(\overline{\theta}_1 - \alpha_1)(b_2+\beta_2)(b_1+\beta_1) + c(b_1+\beta_1)[2(\theta_2^* - \alpha_2) + (b_2+\beta_2)q_{20}] - c^2(\overline{\theta}_1 - \alpha_1) - c^3 q_{20}}{2(b_1+\beta_1)(b_2+\beta_2) - 2c^2}\Bigg\} \\
& \times \left[\frac{(\theta_2^* - \alpha_2)}{(b_1+\beta_1)(b_2+\beta_2)} + \frac{c^2(\overline{\theta}_1 - \alpha_1)(b_2+\beta_2) + c^2(\theta_2^* - \alpha_2)}{(b_1+\beta_1)(b_2+\beta_2)\{(b_1+\beta_1)(b_2+\beta_2) - c^2\}} - \frac{cq_{20}}{b_1+\beta_1}\right] \\
& + (\theta_2^* - \alpha_2)\left[\frac{\theta_2^* - \alpha_2}{b_2+\beta_2} + \frac{c(\overline{\theta}_1 - \alpha_1)(b_2+\beta_2) + c^2(\theta_2^* - \alpha_2)}{(b_2+\beta_2)\{(b_1+\beta_1)(b_2+\beta_2) - c^2\}}\right] - \theta_2^* q_{20} \\
& - \frac{b_2+\beta_2}{2}\left[\frac{\theta_2^* - \alpha_2}{b_2+\beta_2} + \frac{c(\overline{\theta}_1 - \alpha_1)(b_2+\beta_2) + c^2(\theta_2^* - \alpha_2)}{(b_2+\beta_2)\{(b_1+\beta_1)(b_2+\beta_2) - c^2\}}\right]^2 + \frac{b_2}{2}q_{20}^2 \\
& + c\left[\frac{\theta_2^* - \alpha_2}{b_2+\beta_2} + \frac{c(\overline{\theta}_1 - \alpha_1)(b_2+\beta_2) + c^2(\theta_2^* - \alpha_2)}{(b_2+\beta_2)\{(b_1+\beta_1)(b_2+\beta_2) - c^2\}}\right]\left[\frac{(\overline{\theta}_1 - \alpha_1)(b_2+\beta_2) + c(\theta_2^* - \alpha_2)}{(b_1+\beta_1)(b_2+\beta_2) - c^2}\right] \\
& - \frac{c(\overline{\theta}_1 - \alpha_1 + cq_{20})q_{20}}{b_1+\beta_1} + (\zeta_2 + m_2)q_{20} + \frac{1}{2}\left(b_2(2l_2 - 1) - \frac{2c^2(l_2 - 1)}{b_1+\beta_1}\right)q_{20}^2.
\end{aligned}$$

□

2. Optimal Allocation Rule

In this section we characterize the optimal nonlinear pricing that does not use the "aggregation method," but uses the multidimensional screening method as in [Rochet and Choné, 1998] and generalized by [Basov, 2001]. In our case, each publisher can sell one type of advertisement ($q_i \in \mathbb{R}_+$) but agents have two types ($\theta \in \mathbb{R}^2$) $\sim F(\cdot,\cdot)$ so perfect screening is not possible, so we use the generalization in [Basov, 2001].[3] The preferences of an agent of type (θ) is: $u(q_1, q_2; \theta) = \theta_1 q_1 + \theta_2 q_2 - \frac{b_1}{2} q_1^2 - \frac{b_2}{2} q_2^2 + c q_1 q_2$. The first publihser offers quadratic tariff function and is given by

$$T_1(q_1) = \begin{cases} \gamma_1 + \alpha_1 q_1 + \frac{\beta_1}{2} q_1^2 & \text{if} q > q_{10} \\ 0 & \text{if} q \leq q_{10} \end{cases} \tag{5}$$

Let the indirect utility of agent with type (θ) be $W(\theta)$ and is defined as

$$W(\theta) = u(q_1(\theta), q_2(\theta); \theta) - T_1(q_1(\theta)) - T_2(q_2(\theta)),$$

where $q_i(\theta)$ is the optimal quantity of good i purchased by agent of type θ. Suppose, an agent decides to buy only from P1, some optimal quantity while consuming q_{20} for free from P2, then let's denote his indirect utility to be $w_1(\theta)$, which is defined as

$$w_1(\theta) = \max_{\tilde{q}_1 \geq q_{10}} \left[\theta_1 \tilde{q}_1 + \theta_2 q_{20} - \frac{b_1}{2} \tilde{q}_1^2 - \frac{b_2}{2} q_{20}^2 + c \tilde{q}_1 q_{20} - \gamma_1 - \alpha_1 \tilde{q}_1 - \frac{1}{2} \beta_1 \tilde{q}_1^2 \right].$$

After determining the optimal $q_1(\theta)$ via the FOC, and substituting back, $w_1(\theta)$ can be written as

$$w_1(\theta) = -\gamma_1 + \frac{(\theta_1 - \alpha_1)^2}{2(b_1 + \beta_1)} + \left(\frac{c(\theta_1 - \alpha_1)}{b_1 + \beta_1} + \theta_2 \right) q_{20} + \left(\frac{c^2}{2(b_1 + \beta_1)} - \frac{b_2}{2} \right) q_{20}^2. \tag{6}$$

If an agent of type θ buys $q_2 > q_{20}$, from P2 and optimal $q_1(\theta)$ from P1, then his/her (indirect)utility is $W(\theta)$. Whereas, if he/she buys only q_{20} from P2, then it is $w_1(\theta)$. This means the difference in the utility from deciding to buy $q_2 > q_{20}$ from P2 is just the difference between $W(\theta)$ and $w_1(\theta)$. With some calculation it can be shown

[3] Whenever possible, we shall use θ to denote a vector of (θ_1, θ_2).

that $W(\theta) = w_1(\theta) + s(q_2, \theta)$, where $s(q_2, \theta)$ is the residual utility that an agent gets by consuming optimal amount of q_2 and is

$$s(q_2(\theta), \theta) = \max_{\tilde{q}_2 \geq q_{20}} \left[\left(\theta_2 + \frac{c(\theta_1 - \alpha_1 + cq_2)}{b_1 + \beta_1} \right) (q_2 - q_{20}) - \frac{b_2}{2}(q_2^2 - q_{20}^2) - T_2(q_2) \right].$$

Therefore, for P2, an agent with type θ can be thought of as having $s(q_2, \theta)$ as the utility function. From earlier assumptions we know that $s(q_2(\theta), \theta)$ satisfies envelope conditions:

$$\begin{aligned} s_2(q_2(\theta), \theta) &= \frac{\partial s(q_2(\theta), \theta)}{\partial \theta_2} = v_2(q_2(\theta)) = (q_2(\theta) - q_{20}) \\ s_1(q_2(\theta), \theta) &= \frac{\partial s(q_2(\theta), \theta)}{\partial \theta_1} = v_1(q_2(\theta)) = \frac{c(q_2(\theta) - q_{20})}{b_1 + \beta_1}. \end{aligned}$$

We denote $q_2 - q_{10} = \tilde{q}_2$ and write the tariff as a function of the information rent (indirect residual utility) as

$$T_2(\theta) = \theta_2 \tilde{q}_2 + \frac{c(\theta_1 - \alpha_1)\tilde{q}_2}{b_1 + \beta_1} + \frac{c^2(\tilde{q}_2 + q_{20})\tilde{q}_2}{b_1 + \beta_1} - \frac{b_2}{2}(\tilde{q}_2^2 + 2q_{20}(\tilde{q}_2 + q_{20})) - s(\theta).$$

Then the expected profit of P2, with $\Theta^* \subset \Theta$ being those who buy more than q_{20}, can be written as

$$\begin{aligned} \mathbb{E}(\Pi_2) \quad &= \quad \int_{\Theta^*} \Big[\left(\theta_2 + \frac{c(\theta_1 - \alpha_1)}{b_1 + \beta_1} \right) \tilde{q}_2 + \left(\frac{c^2 \tilde{q}_2}{b_1 + \beta_1} - b_2 q_{20} \right) (\tilde{q}_2 + q_{20}) - \frac{b_2}{2} \tilde{q}_2^2 \\ &\qquad -m_2 \tilde{q}_2 - m_2 q_{20} - s(\theta) \Big] f(\theta) \mathrm{d}\theta - m_2 q_{20} F(\theta^*) - K_2. \end{aligned}$$

The optimal quantity and pricing rule will maximize the expected profit conditional on the fact that the agents will choose their quantity appropriately and that every body wants to participate. Before we move on, let's look into the issue of participation in detail. First, from the perspective of P2, $s(q_2, \theta)$ is the additional utility that an agent θ gets from consuming q_2, while his total utility is his indirect utility $W(\theta)$. Also, recall that if the agent chooses not to consume more than q_{20} from P2, then he gets $w_1(\theta)$ if he consumes some amount from P1 or he gets $u(q_{10}, q_{20}, \theta)$ if he consumes the free quantity. Therefore, an agent will participate in the contract with P2 if and only if the gain from participating is at least as much as not participating at all. In other words, the participation constraint is: for all $\theta \in \boldsymbol{\Theta}$,

$$\begin{aligned} & W(\theta) \geq \max\{w_1(\theta), u(q_{10}, q_{20}, \theta)\} \\ \Rightarrow \quad & w_1(\theta) + s(q(\theta), \theta) \geq \max\{w_1(\theta), u(q_{10}, q_{20}, \theta)\} \\ \Rightarrow \quad & w_1(\theta) + s(q(\theta), \theta) \geq w_1(\theta) \\ \therefore \quad & s(q(\theta), \theta) \geq 0. \end{aligned}$$

The third inequality follows from the fact that $w_1(\theta)$ is the utility that results when the agent finds it optimal to purchase $q_1 > q_{10}$ while he purchases only q_{20}. Therefore, the utility $q_1(\theta)$ has to be at least as much as $u(q_{10}, q_{20}, \theta)$. Then P2's objective function is to choose optimal rent that is given to an agent of type θ, with the constraints that the rent so induced is implementable and satisfies participation constraints. Following the literature on implementability, the sufficient condition for implementability is that the rent function has to be convex when evaluated at optimal consumption and the participation constraint must be satisfied. So the optimization problem of P2 is $\max_s \mathbb{E}(\Pi_1)$ such that $s(\cdot)$ is convex; and $s(\theta) \geq 0$.

Let, $p_1 = v_1(\theta) = \frac{c}{b_1+\beta_1}\tilde{q}_2$ and $p_2 = v_2(\theta) = \tilde{q}_2$. Since the dimension of type is more than the dimension of goods, for P2, not all points in the utils space will be feasible. The set of feasible points forms a smooth subset (1 dimensional manifold) in $\mathbb{R}^2_+$. This subset can be characterized as :

$$A = \{p \in \mathbb{R}^2_+ : a(p_1, p_2) = 0\},$$

for some function $a : \mathbb{R}^2 \to \mathbb{R}$. Now, in terms of the newly introduced utils, P2's problem is

$$\begin{aligned} \mathbb{E}(\Pi_2) \quad = \quad & \iint_{\Theta^*} \Big[\sum_{i=1}^{2} \theta_i p_i + \left(\frac{c^2 q_{20} - c\alpha_1}{b_1 + \beta_1} - b_2 q_{20} - m_2 \right) p_2 + \left(\frac{c^2}{b_1 + \beta_1} - \frac{b_2}{2} \right) p_2^2 - s(\theta) \Big] \mathrm{d}F(\theta) \\ & -(K_2 + m_2 q_{20}), \qquad (7) \\ & \text{s.t} \bigtriangledown s(\theta) = z, s(\cdot) - \text{convex}, a(p_1, p_2) = 0. \qquad (8) \end{aligned}$$

For existence and uniqueness results and the interpretation of the constraints see [Basov, 2001]. First we drop the convexity assumption, and derive optimal contract for the "relaxed" problem. The Hamiltonian can for the problem becomes

$$\begin{aligned}\mathbb{H}(p) &= \left[\sum_{i=1}^{2}\theta_i p_i + \left(\frac{c^2 q_{20} - c\alpha_1}{b_1+\beta_1} - b_2 q_{20} - m_2\right) h_2 + \left(\frac{c^2}{b_1+\beta_1} - \frac{b_2}{2}\right) p_2^2 - s(\theta)\right] f(\theta)\\ &\quad -(K_2 + m_2 q_{20}) + \sum_{i=1}^{2}\lambda_i(\theta)p_i + \mu(\theta)a(p_1,p_2),\end{aligned}$$

where λ is the costate vector for the envelope condition $\bigtriangledown s(\theta) = p$ while μ is the Lagrange multiplier on the constraint $a(p_1, p_2) = 0$. Let,

$$\left(\frac{c^2 q_{20} - c\alpha_1}{b_1+\beta_1} - b_2 q_{20} - m_2\right) = x, \qquad \left(\frac{b_2}{2} - \frac{c^2}{b_1+\beta_1}\right) = e.$$

then we can use $a(z) = \frac{cp_2}{b_1+\beta_1} - p_1 = 0$ and re write $\mathbb{H}(z)$ as

$$\mathbb{H} = \left(\sum_{i=1}^{2}\theta_i p_i + xp_2 - ep_2^2 - s\right) f(\theta) + \sum_{i=1}^{2}\lambda_i(\theta)p_i + \mu(\theta)\left(\frac{cp_2}{b_1+\beta_1} - p_1\right) - (K_2 + m_2 q_{20}).$$

Let ν be the unit vector, normal to the boundary of participation and pointing outwards. Then from [Basov, 2001] we have

Theorem 2. *Suppose that the rent function $s(\theta)$ solves the relaxed problem. Then the solution characterized by the following conditions: there exists a continuously differentiable vector function $\lambda : \mathbb{R}^2 \to \mathbb{R}^2$, and a continuously differentiable function $\mu : \mathbb{R}^2 \to \mathbb{R}$, such that*

$$div\lambda + \frac{\partial \mathbb{H}}{\partial s} \le 0; \qquad a.e. on \quad \mathbf{\Theta}^* \tag{9}$$

$$<\lambda, \nu> \quad \ge 0; \qquad a.e. on \quad \partial\mathbf{\Theta}^*. \tag{10}$$

Inequalities (5) and (6) becomes equalities at interior of participation region $\mathbf{\Theta}^*$, i.e whenever $s(\theta) > 0$. For a given vector λ, z is determined by

$$p_i \in \arg\max H. \tag{11}$$

In the Hamiltonian, we should interpret the term $e\tilde{q}_2^2 - x\tilde{q}_2$ as the pseudo-cost of producing q. And under the condition that $b_2 > \frac{2c^2}{b_1+\beta_1}$ it can be shown that this cost is strictly convex in $\tilde{q}_2$. As this condition depends on the optimal pricing chosen by P1, it is taken as parameter by P2, while choosing its own optimal contract. For our purpose we shall assume that this inequality is true, and shall verify it ex post.

This new pseudo-cost function makes our case directly comparable to the case in Basov. From these three conditions we get

$$\frac{\mathrm{d}\lambda_1}{\mathrm{d}\theta_1} + \frac{\mathrm{d}\lambda_2}{\mathrm{d}\theta_2} - f(\theta) \leq 0; \qquad \text{a.e. on } \mathbf{\Theta}, \tag{12}$$

$$\lambda_1 \nu_1 + \lambda_2 \nu_2 \geq 0; \qquad \text{a.e. on } \partial\mathbf{\Theta}, \tag{13}$$

$$\frac{\partial \mathbb{H}}{\partial p_1} = 0 \Rightarrow \lambda_1 = \mu(\theta) - \theta_1 f(\theta) \tag{14}$$

$$\frac{\partial \mathbb{H}}{\partial p_2} = 0 \Rightarrow \lambda_2 = -(\theta_2 + x - 2ep_2)f(\theta) - \frac{\mu(\theta)c}{b_1 + \beta_1}$$

$$\Rightarrow \lambda_2 = -(\theta_2 + x - 2es_2(\theta))f(\theta) - \frac{\mu(\theta)c}{b_1 + \beta_1}, \tag{15}$$

where the last equality follows from the envelope condition, i.e. $\frac{\partial s(\theta_1,\theta_2)}{\partial \theta_2} = p_2$. Differentiating (10) and (11) w.r.t θ_1 and θ_2, respectively, we get

$$\begin{aligned} \frac{\mathrm{d}\lambda_1}{\mathrm{d}\theta_1} &= \mu_1(\theta) - f(\theta) - \theta_1 f_1(\theta), \\ \frac{\mathrm{d}\lambda_2}{\mathrm{d}\theta_2} &= -(1 - 2es_{22}(\theta))f(\theta) - (\theta_2 + x - 2es_2(\theta))f_2(\theta) - \frac{\mu_2(\theta)c}{b_1 + \beta_1}. \end{aligned}$$

Here, for functions $f(\theta)$ and $\mu(\theta)$ we use subscript to denote the partial derivative with respect to that argument. Substituting each of the above expressions in (8) gives us

$$\mu_1(\theta) - f(\theta) - \theta_1 f_1(\theta) - (1 - 2es_{22}(\theta))f(\theta) - (\theta_2 + x - 2es_2(\theta))f_2(\theta) - \frac{\mu_2(\theta)c}{b_1 + \beta_1} - f(\theta) \leq 0$$

$$\Rightarrow (2es_{22}(\theta) - 3)f(\theta) + (2es_2(\theta) - \theta_2 - x)f_2(\theta) + \mu_1(\theta)(\theta) - \theta_1 f_1(\theta) - \frac{\mu_2(\theta)c}{b_1 + \beta_1} = 0 \tag{16}$$

For the first part we shall focus only on substitute goods, i.e. $c < 0$. The inequality (9), binds at the boundary where the agent values the good 2 the most, i.e at $(\underline{\theta}_1, \overline{\theta}_2)$, which, in the contract literature is known as no distortion on top. Hence,

$$\mu(\underline{\theta}_1, \theta_2) - \underline{\theta}_1 f_1(\underline{\theta}_1, \theta_2) = 0; \tag{17}$$

$$-(\overline{\theta}_2 + x - 2es_2(\theta_1, \overline{\theta}_2))f(\theta_1, \overline{\theta}_2) - \frac{\mu(\theta_1, \overline{\theta}_2)c}{b_1 + \beta_1} = 0. \tag{18}$$

From (13) we notice that the multiplier is a a function of θ_2 only through the density function. Hence we conjecture that the multiplier is

$$\mu(\theta) = \theta_1 f(\theta).$$

Hence, $\mu_1(\theta) = \theta_1 f_1(\theta) + f(\theta)$, which satisfies (13) and $\mu_2(\theta) = \theta_1 f_2(\theta)$. Now for notational ease we make a change of variable, $y(\theta) = s_1(\theta)$, from (13) we get

$$\begin{aligned}
&(2ey_1(\theta) - 3)f(\theta) + (2ey(\theta) - \theta_2 - x)f_2(\theta) + \theta_1 f_1(\theta) + f(\theta) - \theta_1 f_1(\theta) - \frac{\theta_1 f_2(\theta)c}{b_1 + \beta_1} = 0\\
&2ey_1(\theta)f(\theta) + \left(2ey(\theta) - \frac{\theta_1 c}{b_1 + \beta_1} - \theta_2 - x\right) f_2(\theta) - 2f(\theta) = 0;\\
&\quad\therefore \frac{\partial}{\partial\theta_2}\left[2eyf(\theta) - \left(\frac{c\theta_1}{b_1 + \beta_1} + \theta_2 + x\right) f(\theta)\right] = f(\theta). \qquad (19)
\end{aligned}$$

Therefore, the optimal contract is characterized by the PDE (15) with the boundary condition:

$$-\left(\overline{\theta}_2 + x - \frac{c\theta_1}{b_1 + \beta_1}\right) f(\theta_1, \overline{\theta}_2) + 2ey(\theta_1, \overline{\theta}_2) f(\theta_1, \overline{\theta}_2) = 0. \qquad (20)$$

Integrating both sides of (16), with respect to θ_2, we get

$$\begin{aligned}
&\int_{\theta_2}^{\overline{\theta}_2} \frac{\partial}{\partial t}\left[2ey(\theta_1, t)f(\theta_1, t) - (\frac{c\theta_1}{b_1 + \beta_1} + t + x)f(\theta_1, t)\right] \mathrm{d}t = k_0 + \int_{\theta_2}^{\overline{\theta}_2} f(\theta_1, t)\mathrm{d}t\\
\textit{or,}\quad &\left\{2ey(\theta_1, \overline{\theta}_2)f(\theta_1, \overline{\theta}_2) - \left(\frac{c\theta_1}{b_1 + \beta_1} + \overline{\theta}_2 + x\right) f(\theta_1, \overline{\theta}_2)\right\}\\
&-\left\{2ey(\theta_1, \theta_2)f(\theta_1, \theta_2) - \left(\frac{c\theta_1}{b_1 + \beta_1} + \theta_2 + x\right) f(\theta_1, \theta_2)\right\} = k_0 + \int_{\theta_2}^{\overline{\theta}_2} f(\theta_1, t)\mathrm{d}t\\
\textit{or,}\quad &-\left\{2ey(\theta_1, \theta_2)f(\theta_1, \theta_2) - \left(\frac{c\theta_1}{b_1 + \beta_1} + \theta_2 + x\right) f(\theta_1, \theta_2)\right\} = k_0 + \int_{\theta_2}^{\overline{\theta}_2} f(\theta_1, t)\mathrm{d}t\\
\textit{or,}\quad &y(\theta_1, \theta_2) = \frac{\left(\frac{c\theta_1}{b_1+\beta_1} + \theta_2 + x\right) f(\theta_1, \theta_2) - k_0 - \int_{\theta_2}^{\overline{\theta}_2} f(\theta_1, t)\mathrm{d}t}{2ef(\theta_1, \theta_2)}\\
\textit{or,}\quad &\tilde{q}_2(\theta_1, \theta_2) = p_2 = \frac{\partial s(\theta_1, \theta_2)}{\partial\theta_2} = y(\theta_1, \theta_2) = \frac{\left(\frac{c\theta_1}{b_1+\beta_1} + \theta_2 + x\right) f(\theta_1, \theta_2) - k_0 - \int_{\theta_2}^{\overline{\theta}_2} f(\theta_1, t)\mathrm{d}t}{2ef(\theta_1, \theta_2)},
\end{aligned}$$

where the third equality follows from the boundary condition (16), and evaluating the above at $\overline{\theta}_2$ gives $k_0 = 0$.. Therefore the optimal contract is

$$q(\theta_1, \theta_2) = \frac{\frac{c\theta_1}{b_1+\beta_1} + \theta_2 - \frac{c^2 q_{20} + c\alpha_1}{b_1+\beta_1} - m_2 - \frac{\int_{\theta_2}^{\overline{\theta}_2} f(\theta_1, t)\mathrm{d}t}{f(\theta_1, \theta_2)}}{\left(b_2 - \frac{2c^2}{b_1+\beta_1}\right)}, \tag{21}$$

which is the same as we found using the aggregation method. The allocation rule for the first principal can be determined analogously and hence not pursued. Note that, unlike in [Rochet and Choné, 1998], optimal allocation rule never generates perfect screening because of the difference in dimension of instrument and agent's type. Therefore, the agent's type is divided into only two subsets, one where they are screened out and offered only $q_{i0}, i = 1, 2$ and the other is bunching of "second kind" where agents with type $h_i = \theta_i + \frac{c\theta_j}{b_j+\beta_j}, j \neq i, i, j \in \{1, 2\}$, get the same good $q_i(h_i)$.

References

ARYAL, G. (2013): "An Empirical Analysis of Competitive Nonlinear Pricing," *Working Paper, ANU.* 1

BASOV, S. (2001): "Hamiltonian approach to multi-dimensional screening," *Journal of Mathematical Economics*, 36, 77–94. 1, 17, 19, 20

ROCHET, J.-C., AND P. CHONÉ (1998): "Ironing, Sweeping, and Multidimensional Screeing," *Econometrica*, 66(4), 783–826. 1, 17, 23