Nonlinear Pricing in Yellow Pages

Yao Huang

Pennsylvania State University

Isabelle Perrigne

Pennsylvania State University &

University of California Los Angeles

Quang Vuong

Pennsylvania State University &

University of California Los Angeles

October 2007

Preliminary Draft

* The last two authors are grateful to the National Science Foundation for research support through grant SES-0452154.

Correspondence to: I. Perrigne, Department of Economics, Pennsylvania State University, University Park, PA 16802. e-mail address : perrigne@psu.edu.

Abstract

Nonlinear pricing is a standard practice used by firms in telecommunications, electricity and advertising among others to discriminate among consumers. In this paper, we analyze nonlinear pricing in yellow pages using a structural approach. First, we develop a model that incorporates some features of the industry such as the requirement by law to propose a minimal advertisement size at zero price to all firms. Our model also includes a general cost function. The structure of the model is defined by the distribution of unknown firms' types, the inverse demand function and the firm's cost function. Under the assumption of a multiplicative inverse demand function into a socalled base marginal utility function and the firm's type, we show that the structure is nonparametrically identified while using the first-order conditions of the publisher's optimization problem. The cost function is identified through the marginal cost at the total amount produced only. We then propose a simple nonparametric procedure to estimate the type distribution and the base marginal utility function. We establish the asymptotic properties of our two-step nonparametric procedure. In particular it is shown that the base marginal utility fucntion is estimated uniformly at the parametric rate. Lastly, the method is applied to analyze nonlinear pricing in yellow pages in Central Pennsylvania. The empirical results show a substantial heterogeneity in firms' tastes for advertising with a decreasing marginal utility function. Some counterfactuals assess the cost of asymmetric information and the benefits of nonlinear pricing in presence of asymmetric information over other pricing rules.

Key words: Nonlinear Pricing, Nonparametric Identification, Nonparametric Estimation, Advertising.

Nonlinear Pricing in Yellow Pages

Yao Huang, Isabelle Perrigne & Quang Vuong

1 Introduction

When firms face heteregeneous consumers, offering different prices across purchase sizes is profitable by discriminating consumers according to their preferences. This practice is often referred to as nonlinear pricing or second degree price discrimination. Nonlinear pricing is a standard practice in electricity market where lower rates apply to successive blocks of kilowatt hours. In the celllular phone industry, calling plans usually specify a monthly fee and a minute rate, where a lower (higher) monthly fee is associated with a higher (lower) minute rate. In advertising, a discount is offered to larger advertisement making it less expensive to buy (say) a full page relative to two half pages. See Wilson (1993) for examples of nonlinear pricing. Economists analyze nonlinear pricing in a framework of imperfect information with adverse selection. Seminal papers by Spence (1977), Mussa and Rosen (1978) and Maskin and Riley (1984) provide nonlinear pricing models for a monopoly. The basic idea is to consider the consumer's unobserved taste (type) as a parameter of adverse selection. The principal or firm designs an incentive compatible tariff through which the consumers will reveal their types. Revelation occurs because the firm gives up some rents to consumers. The principal will induce all consumers except those with the highest type to consume less than the efficient (first-best) amount. The resulting optimal price schedule is concave in quantity implying quantity discounts. Thus nonlinear pricing can be considered as the simplest model of adverse selection relative to comlex contract models. Extensions include Oren, Smith and Wilson (1983) who analyze oligopoly competition with homogenous products. A recent literature considers oligopoly competition with differentiated products with Ivaldi and Martimort (1994), Stole (1995), Armstrong and Vickers (2001), Rochet and Stole (2003) and Stole (2007). When competition is introduced, the optimal schedule is less tractable and closed form solutions can be obtained for some simple specifications only.[1]

The economic importance of nonlinear pricing and the substantial theoretical developments have given rise to an increasing empirical literature. Early empirical studies focus on identifying the presence of nonlinear pricing. Lott and Roberts (1991) provide an alternative cost-based explanation for some commonly viewed second degree price discrimination cases. Later empirical studies have attempted to properly account for cost differences while presenting evidence of nonlinear pricing. Examples include Shepard (1991) who shows evidence of second degree price discrimination for full-service versus self-service by gas stations providing both services, Clerides (2002) for hardback and softback books and Verboven (2002) for diesel and gasoline cars. Other studies also document the impact of competition on patterns of nonlinear pricing. Borenstein (1991) shows that decreased competition in the leaded gas market has reduced differences in price margins between unleaded gas and leaded gas. Borenstein and Rose (1994) show that price dispersion in the US airline industry increases as competition increases. Busse and Rysman (2005) study advertising prices in yellow page directories. They show that a larger competition is associated with a larger degree of curvature, i.e. competition raises the discounts to large buyers.

Recent empirical studies evaluate the economic impact of nonlinear pricing on profits,

[1]See Wilson (1993) and Stole (2007) for more references and a comprehensive treatment of the early and recent literature.

consumer surplus and economic efficiency. Based on a random utility discrete choice model for consumers' preferences, Leslie (2004) analyzes tickets sales for a Broadway play, where consumers choose between tickets for various seat qualities. The author estimates the consumers' random utility function and their taste distribution. He then compares the consumer surplus and the firm's profits under the observed pricing rule and alternative pricing rules. McManus (2007) studies an oligopolistic market of specialty coffee in which the coffee shops differ in terms of locations. The author estimates the consumers' random utility function, which depends on the consumers' unobserved tastes towards different sizes and types of coffee products. Using the estimated consumer utility function and cost data, the magnitude of efficiency distortion (the difference between marginal utility and marginal cost) across different sizes and types of coffee products is assessed. Cohen (2007) applies a similar framework to measure the extent to which price differences across different sizes of paper towels are the result of second degree price discrimination and how profits and consumer welfare change under different pricing rules.

Another trend in the empirical literature on nonlinear pricing endogeneizes the optimal price schedule to recover the demand and cost structure. Ivaldi and Martimort (1994) solve a nonlinear pricing duopoly competition model under a specific structure. In their model, a duopoly produces two differentiated products and the consumers are characterized by a two-dimensional type parameter. The authors then apply this theoretical model to French diary firms for electricity and oil products. Miravete and Roller (2004) apply the same model to analyze the early U.S. cellular telephone industry. The estimated structural parameters are used to evaluate the effect of competition, the policy change in cellular license awarding and welfare under alternative pricing rules. Because their data do not contain individual customers' transactions, the ability to identify demand parameters is limited and their policy conclusions are contingent on the validity of those estimates. Miravete (2002) studies a situation where agents have uncertainty concerning their future consumption. The monopolist can offer either a mandatory nonlinear tariff based on consumers' ex post type or a two-part tariff based on consumers' ex ante type. The author analytically solves for the optimal nonlinear tariff and optimal optional two-part tariff by using a linear demand function and specific distributions for ex post and ex ante types. The author then uses data from a tariff experiment run by South Central Bell to estimate the structural parameters. Crawford and Shum (2007) investigate the magnitude of quality distortion when a cable television monopoly chooses endogenously the quality of the services. The authors observe the services each cable system provides, their prices and the corresponding demand for each service, but not the quality of each service. They use a discrete choice model with the implicit assumption that each cable system chooses the quality and price for each service optimally.

In this paper, we propose a fully structural analysis of nonlinear pricing in yellow page advertising. Yellow page advertising data consitute a well suited case for such an analysis. First, the data contain a large number of different price and quantity options, which can approximate the continuous menu offered by the principal in the theoretical model. Second, individual advertising purchases can be readily observed from phone directories. A characteristic of yellow page advertising is that publishers are required by law to incorporate all businesses by providing basic information such as their name, address and phone number at zero price. We then modify Maskin and Riley (1984) monopoly model to incorporate such an institutional feature. This is equivalent to an optimal exclusion problem, i.e. to find the optimal type below which firms will be offered the standard listing at zero price. The inclusion of such firms has, however, a cost to the publisher which should be taken into account. In contrast to the theoretical and empirical literature which assume a constant marginal cost, we consider a general cost function. The previous empirical literature relies heavily on parametric specifications of the structure. The econometric literature and the recent literature on the structural analysis of auction data have documented that identification of models may depend on particular functional forms.[2] In the spirit of the structural analysis of auction data, we investigate the nonparametric identification of the nonlinear pricing model from observables, which are mainly individual purchase data and tariffs offered by the firm. In a more general perspective of development of a structural analysis of contract data, nonlinear pricing represents one of the simplest case of incomplete information models for contractual relationships. Perrigne and Vuong (2007) establish the nonparametric identification of incentive regulation models but the application of their results on data is pending due to the complexity of estimation. Moreover, nonparametric identification and estimation allow to have policy conclusions robust to functional misspecification.

After developing the model, we establish its nonparametric identification. To do so, we follow the approach proposed by Guerre, Perrigne and Vuong (2000) and exploit the monotonicity of the equilibrium strategy to rewrite the first-order conditions of the publisher's maximization problem in terms of observables. The equilibrium strategy defines the unique mapping between the firm's and its purchase. Because the structure contains multiple elements such as the inverse demand, the type distribution and the publisher's cost function, the identification problem is more involved than in auctions. Under a multiplicative decomposition of the inverse demand function into a socalled base marginal utility function and the firm's type, we show that the model structure is nonparametrically identified. As matter of fact, the base marginal utility function and the firm's type distribution are not identified below a truncation introduced by the optimal exclusion of firms. Moreover, the cost function is only identified at the margin for the total amount produced. Based on this identification result, we propose a natural nonparametric esti-

[2] For instance, the common value auction model can be estimated using particular functional forms for the underlying distributions while this model is not identified in general as shown by Laffont and Vuong (1996) and Li, Perrigne and Vuong (2000).

mator for the base marginal utility function and the type distribution. The asymptotic properties of our estimator are established. In particular, it is shown that the estimator of the marginal utility function is uniformly consistent while its asymptotic distribution converges at the parametric rate.

Next, we analyze a unique data set that we constructed from a phone directory in Pennsylvania and from the Yellow Page Association. The data display a nonlinear pricing pattern as previously documented by Busse and Rysman (2005) for the yellow page advertising industry.[3] The price schedule provides a large number of advertising categories, which allows us to treat the price schedule as nearly continuous. In other empirical studies, only a limited number of different price options is observed. A difficulty arises as the classification of advertisements involves different qualities in addition to advertising space.[4] We construct a quality-adjusted quantity to incorporate different qualities in the price schedule. In addition to the full price schedule, the data contain each firm's purchase totaling more than 7,000 observations. In particular, we observe 149 different advertising categories chosen by firms. As such, this data set is unique because (i) the full price schedule offered to firms is available, (ii) individual data on price and quantity chosen by firms are provided and (iii) the whole population of firms is included.[5] Our empirical results show an important heterogeneity in businesses' taste for advertising or willingness to pay. The estimated base marginal utility function is decreasing as expected.

[3]Rysman (2004) studies the yellow page industry from another angle by considering the phone directory market as a two-sided market where consumers value directories for information and advertisers value directories as a mean to advertise to consumers, giving rise to network effects. The author assumes that publishers use prices for a single quality.

[4]Busse and Rysman (2004) use only advertising spaces within the one color category.

[5]Such nice data features do not exist in other sectors such as telecommunications and electricity. For instance, in telecommunications, marketing survey data contain information on purchased amount in value but not always in quantity (minutes), while the price schedule offered to each consumer is usually unknown. Moreover, part of the market is usually observed.

Counterfactuals assess the cost of asymmetric information in terms of lost profit for the publisher relative to a complete information setting. When asymmetric information is prevalent, we simulate the gain in publisher's revenue and firms' utility levels of nonlinear pricing over other pricing rules. *To be completed.*

The paper is organized as follows. Section 2 introduces the model. Section 3 establishes the nonparametric identification of the model and presents the nonparametric estimation procedure. Section 4 describes the data. Section 5 is devoted to the estimation results and counterfactual experiments. Section 6 concludes with some future lines of research. An appendix collects the proofs.

2 The Model

We extend the Maskin and Riley (1984) nonlinear pricing model in two ways. In our case, the principal is the publishing company and the agents are the firms buying advertising in yellow pages. Firms are characterized by an unknown taste (type) parameter θ for advertising. First, the utility publisher is required to list all businesses' phone numbers in the yellow page section. Thus, the minimum quantity of advertising provided in the price schedule is strictly positive and the price for it is zero. We consider this lowest quantity offered as exogenous and we denote it by q_0, $q_0 > 0$.[6] The problem becomes similar to an optimal exclusion of consumers. The publisher chooses optimally the threshold level θ_0 of the firm's type below which q_0 is provided at zero price. Second, we assume a general cost function instead of a constant marginal cost. A constant marginal cost is a parametric restriction. Doing so will allow us to examine to what extent the cost function can be identified from the observables. These two features alter the publisher's optimization

[6]This quantity is exogenous as it is set by technical constraints, it corresponds to two lines of space over a column.

problem.

Each firm has a utility function defined as

$$\begin{aligned} U(q,\theta) &= U_0 + \int_{q_0}^{q} v(x,\theta)dx - T(q) \quad \forall q > q_0 \\ U(q,\theta) &= U_0 - T(q_0) \quad \forall q \leq q_0, \end{aligned} \tag{1}$$

where q is the quantity of advertising purchased and $T(q)$ is the total payment for q units of advertising. In particular, for $q = q_0$, the firm's utility is a constant independent of θ. The latter assumption is made to avoid countervailing incentives, which arises when the firm's reservation utility depends on its type θ as studied by Lewis and Sappington (1989) and Maggi and Rodriguez-Clare (1995). The term U_0 can be interpeted as the level of utility for the firm to have its name, address and phone number in the yellow pages. For $q > q_0$, the function $v(q,\theta)$ expresses the θ firm's willingness to pay for the qth unit of advertising. The function $v(q,\theta)$ is then the marginal utility of consuming the qth unit of advertising or the inverse demand function for each firm's type. The term θ is distributed as $F(\cdot)$ with continuous density $f(\cdot) > 0$ on its support $[\underline{\theta}, \overline{\theta}]$, $0 \leq \underline{\theta} < \overline{\theta} < \infty$. The publisher does not know each firm's type but knows the distribution $F(\cdot)$. The following assumptions are made on $v(q,\theta)$.

Assumption A1: *The marginal utility function $v(\cdot,\cdot)$ is continuously differentiable on $[q_0, +\infty) \times [\underline{\theta}, \overline{\theta}]$, and $\forall q \geq q_0, \forall \theta \in [\underline{\theta}, \bar{\theta}]$*

(i) $v(q,\theta) > 0$

(ii) $v_1(q,\theta) < 0$

(iii) $v_2(q,\theta) > 0$.[7]

[7]Whenever a function has more than one variable, we denote its derivative with respect to the kth argument by a subscript k.

Assumption A1-(i) says that the marginal utility of consuming advertising is always non-negative. Assumption A1-(ii) says that the marginal utility is decreasing in the quantity purchased. Assumption A1-(iii) says that firms with larger willingnesses to pay θ for advertising enjoy a larger marginal utility across every q. This property is known as the single crossing property.

The publisher chooses optimally the functions $q(\cdot)$ and $T(\cdot)$ and a cutoff type $\theta_0 \in [\underline{\theta}, \overline{\theta})$ to maximize its profit. The function $q(\cdot)$ is defined on $[\underline{\theta}, \overline{\theta}]$ with $q(\theta) = q_0$ for $\theta \in [\underline{\theta}, \theta_0]$ and $q(\theta) \geq q_0$ for $\theta \in (\theta_0, \overline{\theta}]$. The payment $T(\cdot)$ is defined on $[q_0, q(\overline{\theta})]$ with $T(q_0) = 0$ as the publisher has to offer the minimum advertising q_0 at zero price. As usually done in the literature, we restrict $q(\cdot)$ to be continuously differentiable on $(\theta_0, \overline{\theta})$. We also assume for the moment that $q(\cdot)$ is a strictly increasing function on $(\theta_0, \overline{\theta}]$. Later, we will show that with additional assumptions on the game structure the resulting optimal $q(\cdot)$ is strictly increasing.[8]

Without loss of generality, we assume that the publisher faces a population of firms of size one. The publisher's profit can then be written as

$$\Pi = \int_{\theta_0}^{\overline{\theta}} T(q(\theta)) f(\theta) d\theta - C\left[q_0 F(\theta_0) + \int_{\theta_0}^{\overline{\theta}} q(\theta) f(\theta) d\theta\right],$$

where the first term is the revenue collected from all firms buying advertising and the second term expresses the cost for producing the total advertising quantity with a cost function $C(\cdot)$. Because firms characterized by a type below θ_0 do not pay for their advertising quantity q_0, these firms do not show up in the publisher's revenue. On the other hand, this production has a cost that the publisher needs to take into account. This explains the argument of the cost function in two parts: (i) $q_0 F(\theta_0)$ represents the total quantity provided to firms choosing the minimum quantity q_0 and (ii) $\int_{\theta_0}^{\overline{\theta}} q(\theta) f(\theta) d\theta$ is

[8] This result is based on the complete sorting optimum result in Maskin and Riley (1984, Proposition 4).

the total quantity provided to other firms. The cost function is assumed to be strictly increasing.[9]

Assumption A2: *The marginal cost function $C'(\cdot)$ satisfies $C'(q) > 0 \quad \forall q \geq q_0$.*

The publisher's profit needs to be maximized subject to the individual rationality (IR) and the incentive compatibility (IC) constraints of the firms. These two constraints are derived from the firm's optimization problem. For the IR constraints, consider first a firm with type $\theta > \theta_0$. It must prefer to buy $q(\theta)$ rather than q_0, i.e.

$$U(q(\theta), \theta) \geq U_0 \quad \forall \theta \in (\theta_0, \overline{\theta}]. \tag{2}$$

The previous assumption of a firm's utility independent of θ at q_0 is crucial here. Without it, the reservation utility level of the firm would have been a function of θ leading to countervailing incentives. For a firm with type $\theta \leq \theta_0$, it receives q_0, which provides a utility level U_0 thereby satisfying trivially its individual rationality constraint.

For the IC constraints, we consider four cases. First, a firm with $\theta > \theta_0$ must prefer to buy $q(\theta)$ rather than any other quantity $q(\tilde{\theta})$ for $\tilde{\theta} \in (\theta_0, \overline{\theta}]$, i.e. it must not pretend to be another type in $(\theta_0, \overline{\theta}]$. Formally, let $U(\tilde{\theta}, \theta) \equiv U(q(\tilde{\theta}), \theta) \quad \forall \theta, \tilde{\theta} \in (\theta_0, \overline{\theta}]$. The IC constraint can be written as $U(\theta, \theta) \geq U(\tilde{\theta}, \theta) \quad \forall \theta, \tilde{\theta} \in (\theta_0, \overline{\theta}]$. The local first-order condition for the IC constraint to hold is

$$U_1(\theta, \theta) = 0 \quad \forall \theta \in (\theta_0, \overline{\theta}]. \tag{3}$$

By definition $U(\tilde{\theta}, \theta) = U_0 + \int_{q_0}^{q(\tilde{\theta})} v(x, \theta)dx - T(q(\tilde{\theta}))$. Thus $U_1(\theta, \theta) = [v(q(\theta), \theta) - T'(q(\theta))]$

[9] If we consider a population of size N, the publisher's revenue needs to be multiplied by N, while the cost would be written as $C_N(N(q_0 F(\theta_0) + \int_{\theta_0}^{\overline{\theta}} q(\theta) f(\theta) d\theta))$. When dividing the total profit by N, we remark that the cost term would become $C(x) = (1/N)C(Nx)$. Thus, the function $C(\cdot)$ can be interpreted as an average cost per firm. On the other hand, because $C'(x) = C'(Nx)$, $C'(\cdot)$ can be interpreted as the marginal cost at the total amount of production.

$q'(\theta)$. Since by assumption $q'(\cdot) > 0$ on $(\theta_0, \overline{\theta}]$, (3) is equivalent to

$$v(q(\theta), \theta) = T'(q(\theta)) \quad \forall \theta \in (\theta_0, \overline{\theta}]. \tag{4}$$

Second, a firm with type $\theta > \theta_0$ must prefer to buy $q(\theta)$ rather than $q(\tilde{\theta})$ for $\tilde{\theta} \in [\underline{\theta}, \theta_0]$, i.e it must not pretend to be another type in $[\underline{\theta}, \theta_0]$. But $q(\tilde{\theta}) = q_0$ providing U_0. Thus, the IC constraint is $U(q(\theta), \theta) \geq U_0$, which is the IR constraint (2). Third, a firm with type $\theta \leq \theta_0$ must prefer to receive $q(\theta) = q_0$ rather than $q(\tilde{\theta})$ for $\tilde{\theta} \in [\underline{\theta}, \theta_0]$, i.e. it must not pretend to be another type in $[\underline{\theta}, \theta_0]$. But $q(\tilde{\theta}) = q_0$ providing U_0. Thus, the IC constraint is trivially verified. Fourth, a firm with type $\theta \leq \theta_0$ must prefer to receive q_0 rather than to buy $q(\tilde{\theta})$ for $\tilde{\theta} \in (\theta_0, \overline{\theta}]$, i.e. it must not pretend to be another type in $(\theta_0, \overline{\theta}]$. Thus the IC constraint is $U_0 \geq U(\tilde{\theta}, \theta)$. To show such an inequality, we need to use some equalities established later. Specifically, from (6) and $U_+ = U_0$, we have $T(q(\theta)) = \int_{q_0}^{q(\theta)} v(x, \theta)dx - \int_{\theta_0}^{\theta} \{\int_{q_0}^{q(u)} v_2(x, u)dx\}du \geq \int_{q_0}^{q(\theta)} v(x, \theta)dx$ for $\theta \in (\theta_0, \overline{\theta}]$, where the inequality uses A1-(ii). Equivalently, we can write $T(q(\tilde{\theta})) \geq \int_{q_0}^{q(\tilde{\theta})} v(x, \tilde{\theta})dx \geq \int_{q_0}^{q(\tilde{\theta})} v(x, \theta)dx$ for $\tilde{\theta} \in (\theta_0, \overline{\theta}]$, where the inequality uses A1-(ii). Thus, adding $U_0 - T(q(\tilde{\theta}))$ to both sides gives $U_0 \geq U(\tilde{\theta}, \theta)$. The next lemma shows that the local FOC defined in (4) is sufficient for the IC constraint to hold globally. The proof can be found in the appendix.

Lemma 1: *Under Assumption A1, (2) and $q'(\cdot) > 0$ on $(\theta_0, \overline{\theta}]$, the local FOC (3) is sufficient for the IC constraints to hold globally.*

We can now solve the publisher's optimization problem:

$$\max_{q(\cdot), T(\cdot), \theta_0} \Pi \;=\; \int_{\theta_0}^{\bar{\theta}} T(q(\theta))f(\theta)d\theta - C\left[q_0 F(\theta_0) + \int_{\theta_0}^{\overline{\theta}} q(\theta)f(\theta)d\theta\right], \tag{5}$$

subject to the IR constraint (2), the IC constraint (4) and $q(\cdot)$ being strictly increasing. Following Tirole (1988, Chapter 3), we eliminate the function $T(\cdot)$ from the optimization problem. For $\theta \in (\theta_0, \overline{\theta}]$, we define

$$U(\theta) \equiv U(q(\theta), \theta) = U_0 + \int_{q_0}^{q(\theta)} v(x, \theta) dx - T(q(\theta)).$$

Taking the derivative with respect to θ gives

$$U'(\theta) = [v(q(\theta), \theta) - T'(q(\theta)]\, q'(\theta) + \int_{q_0}^{q(\theta)} v_2(x, \theta) dx = \int_{q_0}^{q(\theta)} v_2(x, \theta) dx,$$

where the second equality uses (4). Let $U_+ \equiv \lim_{\theta \downarrow \theta_0} U(\theta)$. Integrating the above equation from θ_0 to θ gives

$$U(\theta) = \int_{\theta_0}^{\theta} \left\{ \int_{q_0}^{q(u)} v_2(x, u) dx \right\} du + U_+.$$

Using the above definition of $U(\theta)$, we obtain

$$T(q(\theta)) = \int_{q_0}^{q(\theta)} v(x, \theta) dx - \int_{\theta_0}^{\theta} \left\{ \int_{0}^{q(u)} v_2(x, u) dx \right\} du + U_0 - U_+, \tag{6}$$

for $\theta \in (\theta_0, \overline{\theta}]$. Thus, the maximization problem (5) can be written as

$$\begin{aligned} \max_{q(\cdot), \theta_0, U(\cdot)} \Pi \quad = \quad & \int_{\theta_0}^{\overline{\theta}} \left[\int_{q_0}^{q(\theta)} v(x, \theta) dx \right] f(\theta) d\theta - \int_{\theta_0}^{\overline{\theta}} \left\{ \int_{\theta_0}^{\theta} \left[\int_{q_0}^{q(u)} v_2(x, u) dx \right] du \right\} f(\theta) d\theta \\ & + (U_0 - U_+)[1 - F(\theta_0)] - C \left[q_0 F(\theta_0) + \int_{\theta_0}^{\overline{\theta}} q(\theta) f(\theta) d\theta \right]. \end{aligned}$$

Integrating by parts, the second term becomes

$$\begin{aligned} & \left[F(\theta) \int_{\theta_0}^{\theta} \left[\int_{q_0}^{q(u)} v_2(x, u) dx \right] du \right]_{\theta_0}^{\overline{\theta}} - \int_{\theta_0}^{\overline{\theta}} \left\{ \left[\int_{q_0}^{q(\theta)} v_2(x, \theta) dx \right] F(\theta) \right\} d\theta \\ & = \int_{\theta_0}^{\overline{\theta}} \left[\int_{q_0}^{q(\theta)} v_2(x, \theta) dx \right] d\theta - \int_{\theta_0}^{\overline{\theta}} \left\{ \left[\int_{q_0}^{q(\theta)} v_2(x, \theta) dx \right] F(\theta) \right\} d\theta. \end{aligned}$$

After rearranging terms and noting that $U(\cdot)$ appears through U_+ only, the firm's problem becomes

$$\begin{aligned} \max_{q(\cdot), \theta_0, U_+} \Pi = & \int_{\theta_0}^{\overline{\theta}} \left\{ \left[\int_{q_0}^{q(\theta)} v(x, \theta) dx \right] f(\theta) - [1 - F(\theta)] \left[\int_{q_0}^{q(\theta)} v_2(x, \theta) dx \right] \right\} d\theta \\ & + (U_0 - U_+) [1 - F(\theta_0)] - C \left[q_0 F(\theta_0) + \int_{\theta_0}^{\overline{\theta}} q(\theta) f(\theta) d\theta \right]. \end{aligned} \tag{7}$$

Maximization of Π with respect to U_+ gives trivially $U_+ = U_0$. The optimal control problem is nonstandard because θ_0 appears at the boundary of the integral. The optimization problem can be solved as a free terminal time and free-end point control problem as in Kirk (1970, pp 188 and 192). The next proposition establishes the necessary conditions for the solution $[q(\cdot), T(\cdot), \theta_0]$. We make the following assumption.

Assumption A3: *For every $\theta \in [\underline{\theta}, \overline{\theta}]$, $v(q,\theta) - [(1 - F(\theta))v_2(q,\theta)/f(\theta)]$ is strictly monotone or identically equal to zero in q.*

Proposition 1: *Under A1, A2, A3 and $q'(\cdot) > 0$ on $(\theta_0, \overline{\theta}]$, the functions $q(\cdot)$ and $T(\cdot)$, and the cutoff type θ_0 that solve the publisher's optimization problem (5) satisfy*

$$\begin{aligned} v(q,\theta) &= C'(Q) + \frac{1 - F(\theta)}{f(\theta)} v_2(q,\theta) \quad \forall \theta \in (\theta_0, \overline{\theta}] && (8) \\ \lim_{\theta \downarrow \theta_0} q(\theta) &= q_0 && (9) \\ T'(q) &= v(q,\theta) \quad \forall \theta \in (\theta_0, \overline{\theta}] && (10) \\ \lim_{q \downarrow q_0} T(q) &= 0, && (11) \end{aligned}$$

where $Q \equiv q_0 F(\theta_0) + \int_{\theta_0}^{\overline{\theta}} q(u) f(u) du$ in (8) and $q = q(\theta)$ in (8) and (10).

Conditions (8) and (9) characterize the optimal $q(\cdot)$ and the optimal cutoff type θ_0. Note that (8) becomes $v(q,\theta) = c + [(1 - F(\theta))v_2(q,\theta)/f(\theta)]$ when the marginal cost is assumed to be constant and denoted by c. The marginal utility for each type then equals the marginal cost plus a nonnegative distortion term due to incomplete information. Hence, all firms buy less than the efficient (first best) quantity of advertising except for the firm with the largest type $\overline{\theta}$ for which there is no distortion. When the cost function is nonlinear, the publisher considers only the marginal cost of the last unit of the total quantity produced Q. For the highest type, its marginal utility equals $C'(Q)$. Once the optimal $q(\cdot)$ is known, the differential equation (10) and the boundary condition (11) characterize the optimal price schedule $T(\cdot)$. Equation (10) says that the marginal price for each type is equal to the marginal utility for that type. Equations (9) and (11) imply the continuity of $q(\cdot)$ and $T(\cdot)$ at θ_0 and q_0, respectively. These are obtained using A3.

The next lemma shows that the optimal $q(\cdot)$ is strictly increasing. Some assumptions need first to be made.

Assumption A4: *The marginal utility function $v(\cdot,\cdot)$ is twice continuously differentiable on $[q_0,+\infty)\times[\underline{\theta},\overline{\theta}]$ and $f(\cdot)$ is continuously differentiable on $[\underline{\theta},\overline{\theta}]$. Moreover, $\forall\theta\in[\underline{\theta},\overline{\theta}]$ and $\forall q\in[q_0,+\infty)$*

(i) $\partial\left[-qv_1(q,\theta)/v(q,\theta)\right]/\partial\theta\le 0$,

(ii) $[1/v_2(q,\theta)]\partial[v_2(q,\theta)/\rho(\theta)]/\partial\theta<1$ *where* $\rho(\theta)=f(\theta)/[1-F(\theta)]$,

(iii) $v_{22}(q,\theta)\le 0$,

(iv) $1-d[1/\rho(\theta)]/d\theta>0$ *so that* $\theta-[(1-F(\theta))/f(\theta)]$ *is increasing in* θ.

These assumptions are quite standard in the theoretical literature in nonlinear pricing. Assumption A4-(i) says that the demand elasticity is nonincreasing in types. Maskin and Riley (1984) show that a large classes of preferences satisfy this assumption.[10] Assumption A4-(ii) is difficult to interpret. Lemma 2 shows, however, that A4-(ii) is implied by the two other assumptions A4-(iii) and A4-(iv). Assumption A4-(iii) says that the increase in demand price is diminishing as θ increases. Assumption A4-(iv) states that the hazard rate of the distribution of types does not decline too rapidly as θ increases. A large class of distribution functions satisfy this property. These assumptions on the structure are generally sufficient assumptions for the second-order conditions of the optimization problem. As such, they might be weakened. In a different problem, Perrigne and Vuong (2007) derive the sufficient and necessary conditions for the second-order conditions to hold in terms of observables. Such an exercise which is related to test the model validity

[10] In particular, the family of linear demand functions of the form of $v(q,\theta)=\theta-\beta(\theta)q$ and the family of constant elasticity demand functions of the form of $v(q,\theta)=\alpha(\theta)q^{-1/\eta}$, where $\eta>0$, satisfy this assumption under proper assumptions on $\beta(\theta)$ and $\alpha(\theta)$.

is left for future research.

Lemma 2: *Under A1, A2, and A4-(i,ii) or under assumptions A1, A2 and A4-(i,iii,iv),* $q(\cdot)$ *is strictly increasing and continuously differentiable on* $[\theta_0, \overline{\theta}]$ *with* $q'(\cdot) > 0$ *on* $[\theta_0, \overline{\theta}]$. *Moreover,* $T(\cdot)$ *is strictly increasing and twice continuously differentiable on* $[q_0, \overline{q}]$ *with* $T'(\cdot) > C'(Q)$ *on* $(q_0, \overline{q}]$ *and* $T'(q_0) = C'(Q)$.

Regarding the verification of the second-order conditions, we invite the reader to consult Maskin and Riley (1984). Tirole (1988) indicates that $T''(\cdot) < 0$, i.e. the price schedule is strictly concave in q.

3 Nonparametric Identification and Estimation

3.1 Nonparametric Identification

It is first useful to define the game structure and the observables. The data provide information on the price-advertising schedule, the minimum quantity at zero price, the proportion of firms choosing this minimum quantity, the firms' advertising purchases and the total amount of advertising produced. Using our previous notations, the observables are $[T(\cdot), q_0, F(\theta_0), G^*(\cdot), N]$, where N denotes the number of firms.[11] The function $G^*(\cdot)$ denotes the truncated distribution of firms' purchases, i.e. $G^*(q^*) = \Pr(q \leq q^*)/\Pr(q > q_0)$. The structural primitives of the model are $[v(\cdot,\cdot), F(\cdot), C(\cdot)]$, which are the marginal utility function, the firms' type distribution and the cost function. We adopt a structural approach. Specifically, we assume that the observables are the outcomes of the optimal price schedule and purchasing choices determined by the equilibrium necessary conditions (8), (9), (10) and (11) while θ is the unobserved random variable in the model.

[11] In this subsection, the observables are assumed to be known. The estimation of $T(\cdot)$, $\overline{q}$, $F(\theta_0)$ and $G^*(\cdot)$ is considered in the next subsection.

Identification investigates whether the primitives can be uniquely recovered from the observables. Following results obtained in Perrigne and Vuong (2007), we consider $v(q, \theta)$ being multiplicatively separable in θ. We show later that the general function $v(q, \theta)$ is not identified.

Assumption B1: *The consumer's marginal utility function is of the form*

$$v(q, \theta) = \theta v_0(q), \tag{12}$$

where $v_0(\cdot)$ satisfies $v_0(q) > 0$, and $v_0'(q) < 0$ for $q \geq q_0$ and for all $\theta \in [\underline{\theta}, \overline{\theta}] \subset (0, +\infty)$.[12]

Hereafter, we interpret $v_0(q)$ as the base marginal utility function. It can be easily seen that the assumptions on the marginal utility function A1 and A4-(i,iii) are satisfied under such a specification. The necessary conditions (8) and (10) then become

$$\begin{aligned} \theta v_0(q) &= C'(Q) + \frac{1 - F(\theta)}{f(\theta)} v_0(q) \quad \forall q \in (q_0, \overline{q}] & (13) \\ T'(q) &= \theta v_0(q) \quad \forall q \in (q_0, \overline{q}], & (14) \end{aligned}$$

where $\theta = q^{-1}(q)$ and $\overline{q} = q(\overline{\theta})$ since $q(\cdot)$ is strictly increasing following Lemma 2. The unique monotone mapping between the unobserved type θ and the observed advertising quantity q is the key of our identification result. Hereafter, we let $\mathcal{S}$ be the set of structures $[v_0(\cdot), F(\cdot), C'(\cdot)]$ such that $[v(\cdot, \cdot), F(\cdot), C(\cdot)]$ satisfy B1, A2, A3 and A4-(iv).

Our first identification result concerns the cost function, which is not identified except for the marginal cost at the total amount produced. See footnote 9, which explains that $C'(Q)$ can be interpereted as the marginal cost at the total amount produced MQ.

[12]We could consider a multiplicative function under the form $\psi(\theta)v_0(q)$. The function $\psi(\cdot)$ needs to satisfy $\psi(\cdot) > 0$, $\psi'(\cdot) > 0$ and $\psi''(\cdot) \leq 0$. If $\psi(\cdot)$ is known, our results extend trivially. On the other hand, if $\psi(\cdot)$ is unknown, the model is likely to remain nonidentified even under a suitable normalization. Note also that under this assumption of multiplicative separability, it can be shown following (say) Tirole (1988, p156) that the price schedule is strictly concave in q when A4-(iv) is strengthened to a hazard rate $\rho(\theta)$ increasing in θ.

This result is not surprising since the model involves the cost function only through the marginal cost at the total amount produced. The next lemma formalizes this result.

Lemma 3: *The cost function is not identified except for the marginal cost at the total amount produced, which satisfies*

$$C'(Q) = T'(\overline{q}). \tag{15}$$

Evaluating (13) and (14) at $\overline{\theta}$ gives the result since $F(\overline{\theta}) = 1$ and $q(\overline{\theta}) = \overline{q}$ following Lemma 2. In other words, the marginal cost at the total amount produced equals the marginal tariff paid by the highest type or equivalently for the largest advertising quantity offered. Since we observe both the price schedule $T(\cdot)$ and $\overline{q}$, we are able to recover $C'(Q)$, i.e. the latter is identified. Miravete and Roller (2004) note a similar result to estimate the marginal cost in their study. Because they do not observe individual cellular phone usages in their data, they do not observe $\overline{q}$ and have to choose an arbitrary value for it. In Section 5, we discuss how identification of the cost function can be improved using data from different phone books providing different values for Q.

The structural elements left to identify are $v_0(\cdot)$ and $F(\cdot)$. We first show that a scale normalization is necessary. Intuitively, this normalization is needed because both the firm's type θ and the function $v_0(\cdot)$ are unknown. We can find two observationally equivalent structures by transforming the unknown firm's type θ into a new type $\tilde{\theta} = \alpha\theta$ and the unknown function $v_0(\cdot)$ into a new function $\frac{1}{\alpha}v_0(\cdot)$ for any positive number α. The next lemma formalizes this result.

Lemma 4: *Consider a structure* $S = [v_0(\cdot), F(\cdot), C'(\cdot)] \in \mathcal{S}$. *Define another structure* $\tilde{S} = [\tilde{v}_0(\cdot), \tilde{F}(\cdot), C'(\cdot)]$, *where* $\tilde{v}_0(\cdot) = \frac{1}{\alpha}v_0(\cdot)$ *and* $\tilde{F}(\cdot) = F(\cdot/\alpha)$ *for some* $\alpha > 0$. *Thus,* $\tilde{S} \in \mathcal{S}$ *and the two structures* S *and* $\tilde{S}$ *lead to the same set of observables* $[T(\cdot), G^*(\cdot), q_0, Q, F(\theta_0)]$, *i.e. the two structures are observationally equivalent.*

Several scale normalizations can be entertained. Three natural choices are to fix $\underline{\theta}$, θ_0 or $\overline{\theta}$. Before discussing the appropriate choice of normalization, we establish Lemma 5, in which the firm's marginal utility $v_0(\cdot)$ and its unobserved type θ are expressed as functions of the quantity purchased q and other observables $[T(q), G^*(q), q_0, F(\theta_0), Q]$. Based on Lemma 5, the choice of normalization and the nonparametric identification of $[v_0(\cdot), F(\cdot)]$ are readily established.

Lemma 5: *Let $[v_0(\cdot), F(\cdot), C'(\cdot)] \in \mathcal{S}$. Denote $\gamma \equiv C'(Q)$ and $\theta(\cdot) \equiv q^{-1}(\cdot)$. The necessary conditions (11) and (12) are equivalent to*

$$v_0(q) = \frac{T'(q)}{\theta_0}\left[1 - G^*(q)\right]^{1-\frac{\gamma}{T'(q)}} \exp\left\{-\gamma \int_{q_0}^{q} \frac{T''(x)}{T'(x)^2} \log\left[1 - G^*(x)\right] dx\right\} \tag{16}$$

$$\theta(q) = \theta_0 \left[1 - G^*(q)\right]^{\frac{\gamma}{T'(q)}-1} \exp\left\{\gamma \int_{q_0}^{q} \frac{T''(x)}{T'(x)^2} \log\left[1 - G^*(x)\right] dx\right\}, \tag{17}$$

for all $q \in (q_0, \overline{q}]$.

The proof of Lemma 5 exploits the unique mapping between the advertising quantity purchased q and the firm's type θ. In particular, Lemma 2 shows that $q(\cdot)$ is strictly increasing on $(\theta_0, \overline{\theta}]$. For each $q \in (q_0, \overline{q}]$, we observe the truncated distribution $G^*(q) = \Pr(\tilde{q} \leq q | \tilde{q} > q_0) = \Pr(\tilde{\theta} \leq \theta(q) | \tilde{\theta} > \theta(q_0)) = [F(\theta) - F(\theta_0)]/[1 - F(\theta_0)]$ with the corresponding density $g^*(q) = \theta'(q) f(\theta)/[1 - F(\theta_0)]$, where $\theta = \theta(q)$. We can then replace $[1 - F(\theta)]/f(\theta)$ in (13) by $\theta'(q)[1 - G*(q)]/g^*(q)$. This expression is further used to express the unknown base marginal utility function $v_0(\cdot)$ and the unobserved type θ in terms of observables, which are the corresponding quantity q, the truncated quantity distribution $G^*(\cdot)$, its density $g^*(\cdot)$ and the price schedule $T(\cdot)$ as well as $\gamma = C'(Q)$, which is identified by Lemma 3. There is a clear parallel here with auction models. Guerre, Perrigne and Vuong (2000) use the unique mapping between the bidder's private value and his equilibrium bid to rewrite the FOC of the bidder's optimization problem and express the unobserved private value in terms of the corresponding bid, the bid distribution and its density. In our problem, the firm's type θ can be interpreted as the unknown bidder's private value, while the firm's chosen quantity q can be interpreted as the observed bidder's bid. Our problem is, however, more involved because we have one more structural element in addition to the distribution of firms' type $F(\cdot)$ to recover, i.e. the base utility function $v_0(\cdot)$. To this end, we also exploit the relationship between the shape of the price schedule $T(\cdot)$ and the distribution of the unobserved firms' type $F(\cdot)$. The use of this relationship is made clear in the proof of Lemma 5.

In view of Lemma 5, it can be easily seen that if we normalize $\theta_0 = 1$, the base marginal utility function $v_0(\cdot)$ can be uniquely recovered on $(q_0, \overline{q}]$ from the observables $T'(\cdot)$, $T''(\cdot)$, $G^*(\cdot)$ and $\gamma = T'(\overline{q})$ and hence on q_0 by continuity of $v_0(\cdot)$. Similarly, the truncated type distribution $F^*(\cdot) \equiv [F(\cdot - F(\theta_0)]/[1 - F(\theta_0)]$ can be uniquely recovered on $[\theta_0, \overline{\theta}]$ from the same observables. The following assumption and proposition formalize this result.

Assumption B2: *We normalize* $\theta_0 = 1$.

Under such a normalization, $v_0(q)$ can be interpreted as the marginal utility function for the cutoff type.

Proposition 2: *Let* $[v_0(\cdot), F(\cdot), C'(\cdot)] \in \mathcal{S}$. *Under assumption B2, the marginal utility function* $v_0(\cdot)$ *and the truncated firms' type distribution* $F^*(\cdot)$ *are identified on* $[q_0, \overline{q}]$ *and* $[\theta_0, \overline{\theta}]$*, respectively.*

We note that $F(\cdot)$ can be recovered from $F^*(\cdot)$ on $[\theta_0, \overline{\theta}]$ since $F(\theta_0)$ is observed as the proportion of firms receiving q_0 at zero price. On the other hand, $v_0(\cdot)$ and $F(\cdot)$ are not identified on $[\underline{q}, q_0)$ and $[\underline{\theta}, \theta_0)$, respectively. Intuitively, the price and quantity data do not provide any variation to be able to identity these functions in these ranges of values as the quantity and price are fixed to q_0 and 0, respectively. For a similar reason, the constant additive term U_0 in (1) is not identified. Here again, we can make a parallel to auction models, in which a binding reserve price does not allow to identify the distribution of bidders' private values for private values below the reserve price because of the lack of bid data. See Guerre, Perrigne and Vuong (2000).

As indicated previously, alternative normalizations can be entertained. For instance, the normalization $\overline{\theta} = 1$ also allows to identify $v_0(\cdot)$ and $F^*(\cdot)$ on $[q_0, \overline{q}]$ and $[\theta_0, \overline{\theta}]$, respectively. In this case, (16) and (17) become

$$\begin{aligned} v_0(q) &= \frac{T'(q)}{\overline{\theta}} \left[1 - G^*(q)\right]^{1-\frac{\gamma}{T'(q)}} \exp\left\{\gamma \int_q^{\overline{q}} \frac{T''(x)}{T'(x)^2} \log\left[1 - G^*(x)\right]\right\} \\ \theta(q) &= \overline{\theta} \left[1 - G^*(q)\right]^{\frac{\gamma}{T'(q)}-1} \exp\left\{-\gamma \int_q^{\overline{q}} \frac{T''(x)}{T'(x)^2} \log\left[1 - G^*(x)\right]\right\}, \end{aligned}$$

for all $q \in (q_0, \overline{q}]$, which are obtained by evaluating (17) at $\overline{q}$, solving for θ_0 and substituting the solution in (16) and (17). Therefore, if $\overline{\theta} = 1$, identification is obtained as in Proposition 2. As a matter of fact, any normalization of $\theta \in [\theta_0, \overline{\theta}]$ would work. On the other hand, any normalization in $[\underline{\theta}, \theta_0)$ would not help in identifying the model.

It remains to show that the general model with a nonseparable multiplicative marginal utility function $v(\cdot, \cdot)$ is not identified.

3.2 Nonparametric Estimation

In view of our identification result, we adopt an indirect approach to estimate the model. Equations (16) and (17) provide expressions for the unknown base marginal utility $v_0(\cdot)$ and the firm's type $\theta(\cdot)$ as functions of $T'(\cdot)$, $T''(\cdot)$, $\gamma = T'(\overline{q})$ and the truncated quantity distribution $G^*(\cdot)$. The fonctions $T'(\cdot)$ and $T''(\cdot)$ come from the price schedule data. On the other hand, $G^*(\cdot)$, $\overline{q}$ and hence γ need to be estimated. We proceed with a two-step estimation procedure. In the first step, we estimate $\overline{q}$ and $G^*(\cdot)$ nonparametrically. This allows us to obtain an estimate for the base marginal utility $v_0(\cdot)$ using (16) and to construct a sample of pseudo types. In the second step, this pseudo sample is used to estimate nonparametrically the truncated density of the firms' type from which we can easily estimate the density of firms' type using the observed proportion of firms choosing q_0. Because we use data from a single market, our estimation procedure is not performed conditionally upon some exogenous variables Z, which capture market heterogeneity. If the data set contains data from several yellow pages directories, we can easily entertain this case by estimating $G^*(\cdot|\cdot)$ in the first step and $F(\cdot|\cdot)$ in the second step of the estimation procedure. The variables Z would include the median income, population size, etc.

We denote by N^* the number of firms purchasing advertising space strictly larger than q_0, while $q_i, i = 1, 2, ..., N^*$ denotes the quantity purchased by each of those firms. From (15), we estimate γ by $\hat{\gamma} = T'(q_{\max})$, where $q_{max} = \max_{i=1,...,N^*} q_i$. We estimate $G^*(\cdot)$ as an empirical distribution using a counting process. We obtain

$$\hat{G}^*(q) = \frac{1}{N^*} \sum_{i=1}^{N^*} 1\!\!I(q_i \leq q), \tag{18}$$

where $1\!\!I(\cdot)$ is an indicator function, for $q \in [q_0, \overline{q}]$. Using (16), (18) and B2, the estimate for $v_0(\cdot)$ is

$$\hat{v}_0(q) = \begin{cases} T'(q) \left[1 - \hat{G}^*(q)\right]^{1-\frac{\hat{\gamma}}{T'(q)}} \exp\left\{-\hat{\gamma} \int_{q_0}^{q} \log\left[1 - \hat{G}^*(x)\right] \frac{T''(x)}{T'(x)^2} dx\right\} & \text{if } q \in (q_0, q_{max}) \\ \lim_{x \uparrow q_{\max}} \hat{v}_0(x) & \text{if } q \in [q_{\max}, \overline{q}]. \end{cases} \tag{19}$$

As a matter of fact, for $q \in [q_0, q_{\max}]$, the estimator $\hat{v}_0(q)$ is straightforward to compute as the integral in (19) can be decomposed as a finite sum. Specifically, because the empirical distribution $\hat{G}(\cdot)$ is a step function with steps at $q^1 < q^2 < \ldots < q^J$ in $(q_0, \overline{q}]$, we note that the integral from q_0 to q can be written as the sum of integrals from q^j to q^{j+1}. On each of these intervals, $\log[1 - \hat{G}^*(\cdot)]$ is constant, while the primitive of $-T''(\cdot)/T'(\cdot)$ is $1/T'(\cdot)$. Thus, for $q \in [q_0, q_{\max})$, (19) can be rewritten as

$$\begin{aligned} \hat{v}_0(q) \ &= T'(q) \left[1 - \hat{G}^*(q)\right]^{1-\frac{\hat{\gamma}}{T'(q)}} \\ &\exp\left\{\hat{\gamma} \sum_{t=0}^{j-1} \left[\left(\frac{1}{T'(q^{t+1})} - \frac{1}{T'(q^t)}\right) \log(1 - \hat{G}^*(q^t))\right] + \hat{\gamma} \left(\frac{1}{T'(q)} - \frac{1}{T'(q^j)}\right) \log(1 - \hat{G}^*(q^j))\right\} \end{aligned} \tag{20}$$

if $q \in [q^j, q^{j+1}), j = 0, \ldots, J-1$, where $q^0 = q_0$ and $q^J = q_{\max}$. In particular, if $q \in [q_0, q^1)$,

$\hat{v}_0(q) = T'(q)$ because $\hat{G}^*(q) = 0$. For $q \in [q_{\max}, \overline{q}]$, $\hat{v}_0(\cdot)$ is constant and equal to

$$\lim_{q \uparrow q_{\max}} \hat{v}_0(q) = T'(q_{\max}) \exp \left\{ \hat{\gamma} \sum_{t=0}^{J-1} \left[\left(\frac{1}{T'(q^{t+1})} - \frac{1}{T'(q^t)} \right) \log(1 - \hat{G}^*(q^t)) \right] \right\},$$

which is finite.[13] Thus, $\hat{v}_0(\cdot)$ is a well defined strictly positive *cadlag* (continue à droite, limites à gauche) function on $[q_0, \overline{q}]$ with steps at $q^1 < q^2 < \ldots < q^{J-1}$. To complete the first step, using (14) we construct a sample of pseudo firms' types

$$\hat{\theta}_i = \hat{\theta}(q_i) = \frac{T'(q_i)}{\hat{v}_0(q_i)} \tag{21}$$

for $i = 1, \ldots, N^*$. In particular, the highest type $\overline{\theta}$ is estimated by $\theta_{\max} = T'(q_{\max})/\hat{v}_0(q_{\max})$, where $\hat{v}_0(q_{\max})$ is given by the above limit. If $T'(\cdot)$ is strictly decreasing so that $T(\cdot)$ is concave, then it can be seen that $\theta_{\max} > \theta_0 = 1$, while $\hat{v}_0(\cdot)$ and $\hat{\theta}(\cdot)$ are strictly decreasing and increasing on $[q_0, q_{\max}]$, respectively.

In the second step, following Guerre, Perrigne and Vuong (2000), we estimate the truncated density of firms' type from the pseudo sample by using the kernel estimator

$$\hat{f}^*(\theta) = \frac{1}{N^* h} \sum_{i=1}^{N^*} K \left(\frac{\theta - \hat{\theta}_i}{h} \right), \tag{22}$$

for $\theta \in [\theta_0, \overline{\theta}] = [1, \overline{\theta}]$, where $K(\cdot)$ is a symmetric kernel function with compact support, h is a bandwidth and $\hat{\theta}_i$ is obtained from (21).[14]

The next proposition establishes the asymptotic properties of $\hat{v}_0(\cdot)$ as an estimator of $v_0(\cdot)$. We make the following assumption on the data generating process.

Assumption B2: *The types $\theta_i, i = 1, \ldots, N$, where N (total number of firms) is the sample size, are independent and identically distributed as $F(\cdot)$.*

---

[13]When $q \uparrow q_{\max}$, then $q \in [q^{J-1}, q^J]$. Thus $\hat{G}^*(q) = \hat{G}^*(q^{J-1}) \leq (N^* - 1)/N^*$ and $\lim_{q \uparrow q_{\max}} [1 - G^*(q)]^{1-(\hat{\gamma}/T'(q))} = 1$ as $\lim_{q \uparrow q_{\max}} T'(q) = \hat{\gamma}$.

[14]We prefer to estimate the type density instead the truncated type distribution as the latter is not as informative. Moreover, counterfactuals rely on the estimated density.

Following the empirical process literature popularized in econometrics by Andrews (1994), we view $\hat{v}_0(\cdot)$ as a stochastic process defined on $[q_0, \overline{q}]$ and hence as a random element of the space $D[q_0, \overline{q}]$ of cadlag functions on $[q_0, \overline{q}]$. We endow the latter space with the uniform metric $\|\psi_1 - \psi_2\| = \sup_{q \in [q_0, \overline{q}]} |\psi_1(q) - \psi_2(q)|$. For some technical reasons, we consider instead the space $D[q_0, q_1]$ with its uniform metric $\|\psi_1 - \psi_2\| = \sup_{q \in [q_0, q_1]} |\psi_1(q) - \psi_2(q)|$, where $q_1 \in [q_0, \overline{q})$.[15] Weak convergence on the space $D[q_0, q_1]$ is denoted by "$\Rightarrow$."

Proposition 3: *Under assumptions A1–A4, B1 and B2, for any fixed $q_1 \in [q_0, \overline{q})$, we have as N tends to infinity*

(i) $\|\hat{v}_0(\cdot) - v_0(\cdot)\| \xrightarrow{a.s.} 0$ *on* $[q_0, q_1]$,

(ii) as random functions in $D[q_0, q_1]$,

$$\sqrt{N}[\hat{v}_0(\cdot) - v_0(\cdot)] \Rightarrow \frac{v_0(\cdot)}{\sqrt{1 - F(\theta_0)}} \left\{ \left[1 - \frac{\gamma}{T'(\cdot)}\right] \frac{\mathcal{B}_{G^*}(\cdot)}{1 - G^*(\cdot)} - \gamma \int_{q_0}^{\cdot} \frac{T''(x)}{T'(x)^2} \frac{\mathcal{B}_{G^*}(x)}{1 - G^*(x)} dx \right\}, \quad (23)$$

where $\mathcal{B}_{G^*}(\cdot)$ is the G^*-Brownian bridge process on $[q_0, \overline{q}]$.[16]

The first part establishes the uniform almost sure convergence of $\hat{v}_0(\cdot)$ on any subset $[q_0, q_1]$ with $q_1 < \overline{q}$. The second part gives the asymptotic distribution of $\hat{v}_0(\cdot)$. It is worthnoting that its rate of convergence is the parametric rate $\sqrt{N}$ despite that $v_0(\cdot)$ is a function. This surprising result comes from that $\hat{v}_0(\cdot)$ is a functional of the empirical cdf $\hat{G}^*(\cdot)$. Let $Z(\cdot)$ be the Gaussian process appearing between the braces in (23). It can be shown that $Z(\cdot)$ has zero mean and covariances $\mathrm{E}[Z(q)Z(q')] = \omega(q)$ for $q_0 \le q \le q' \le q_1$

[15] As usual measurability issues are ignored below. This can be addressed by considering either the projection σ-field on $D[q_0, q_1]$ as in Pollard (1984) or outer probabilities as in van der Vaart (1998). Alternatively, we may use another metric such as the Skorohod metric as in Billingsley (1968).

[16] The G^*-Brownian bridge process on $[q_0, \overline{q}]$ is the limit of the empirical process $(1/\sqrt{N^*}) \sum_i \{ 1\!\!1(q_i \le \cdot) - G^*(\cdot)\}$ indexed by $[q_0, \overline{q}]$. See (say) van der Vaart (1998, p 266). It is a Gaussian process with mean 0 and covariance $G^*(q)[1 - G^*(q')]$, where $q_0 \le q \le q' \le \overline{q}$.

(see Appendix), where

$$\begin{aligned}\omega(q) &= \left(1-\frac{\gamma}{T'(q)}\right)^2 \frac{G^*(q)}{1-G^*(q)} + 2\gamma\left(1-\frac{\gamma}{T'(q)}\right) H(q) - 2\gamma^2 \int_{q_0}^{q} \frac{T''(x)}{T'(x)^2} H(x)dx \quad (24)\\ H(q) &= -\int_{q_0}^{q} \frac{T''(x)}{T'(x)^2}\frac{G^*(x)}{1-G^*(x)}dx.\end{aligned}$$

Note that the covariance $\mathrm{E}[Z(q)Z(q')]$ is independent of q. Thus, the covariances of the limiting process in (23) is $v_0(q)v_0(q')\omega(q)/[1-F(\theta_0)]$ for $q_0 \le q \le q' \le q_1$. In particular, Proposition 3-(ii) implies that

$$\sqrt{N}[\hat{v}_0(q) - v_0(q)] \xrightarrow{d} \mathcal{N}\left(0, \frac{v_0(q)^2}{1-F(\theta_0)}\omega(q)\right)$$

for every $q \in [q_0, q_1]$. The asymptotic variance of $\hat{v}(q)$ vanishes at $q = q_0$ as $\omega(q_0) = 0$, which is expected since $\hat{v}_0(q_0) = T'(q_0)$, while $\omega(q)$ increases as q increases to q_1 whenever $T''(\cdot) < 0$, i.e. $T(\cdot)$ is strictly concave.

In practice, the preceding asymptotic distribution is used to conduct large sample hypothesis tests or construct approximate "pointwise" confidence intervals for $v_0(q)$ provided the asymptotic variance is estimated consistently. A natural estimator is obtained by replacing $v_0(q)$, $1 - F(\theta_0)$ and $\omega(q)$ by $\hat{v}_0(q)$, N^*/N and $\hat{\omega}(q)$, respectively where $\hat{\omega}(q)$ is obtained from (24) by replacing γ and $G^*(\cdot)$ by their estimates $T'(q_{\max})$ and $\hat{G}^*(\cdot)$. Alternatively, the weak convergence result in Proposition 3 delivers an approximate "uniform" confidence interval of the form

$$\left[\hat{v}_0(\cdot)\left(1+\frac{1}{1+c/\sqrt{N^*}}\right), \hat{v}_0(\cdot)\left(1+\frac{1}{1-c/\sqrt{N^*}}\right)\right]$$

for $q \in [q_0, q_1]$, where c is the constant defined by $\Pr\left(\{|Z(q)| \le c, \forall q \in [q_0, q_1]\}\right) = 1 - \alpha$ with $0 < \alpha < 1$ and $Z(\cdot)$ is the Gaussian process introduced above.[17]

[17]Similar results apply to the estimation of $\theta(\cdot)$. For example, (21) and Proposition 3 imply that $\sqrt{N}[\hat{\theta}(\cdot) - \theta(\cdot)] \Rightarrow -\theta(\cdot)Z(\cdot)/\sqrt{1-F(\theta_0)}$. Thus, an approximate uniform confidence interval for $\theta(\cdot)$ is the same as that for $v_o(\cdot)$ with $\hat{v}_0(\cdot)$ replaced by $\hat{\theta}(\cdot)$.

We now turn to the second step, i.e. the estimation of the type density $f(\cdot)$. Following the two-step procedure of Guerre, Perrigne and Vuong (2000), but in contradistinction with that paper, the pseudo firms' types $\hat{\theta}_i$ obtained in the first step converge to the firms' types θ_i essentially at the parametric rate $\sqrt{N^*}$, more precisely at the rate $\sqrt{N^*/\log\log N^*}$, which is the uniform rate of convergence of $\sup_{q\in[q_0,q_1]}|\hat{G}^*(q)-G^*(q)|$ from e.g. van der Vaart (1998, p.268). Because this rate is larger than the maximal rate of convergence that can be achieved for estimating the density $f(\cdot)$, estimation of θ_i does not affect the second step. Consequently, the standard kernel estimator (22), which uses the pseudo firms' types, possesses the standard asymptotic properties of uniform convergence and limiting distribution, namely

(i) $\sup_{\theta\in C}|\hat{f}(\theta)-f(\theta)| \overset{a.s.}{\to} 0$ for any compact subset C of $(\theta_0,\overline{\theta})$ provided $h\to 0$ and $N^*h/\log N^*\to\infty$,

(ii) $\sqrt{N^*h}[\hat{f}(\theta)–f(\theta)] \overset{d}{\to} \mathcal{N}(0, f(\theta)\int K(x)^2dx)$ for every $\theta\in(\theta_0,\overline{\theta})$ provided $N^*h^5\to 0$, as $f(\cdot)$ is twice continuously differentiable and bounded away from zero on $[\theta_0,\overline{\theta}]$. See e.g. Silverman (1986). In particular, by choosing the bandwidth h proportional to $(\log N^*/N^*)^{1/5}$, the optimal uniform convergence rate obtained by Stone (1982) for estimating $f(\cdot)$ when the firms' types θ_i are observed can be achieved by our indirect two-step procedure when the θ_i are unobserved but firm's consumption choices q_i are observed.

4 Yellow Page Advertising Data

We collected data on yellow page advertising in 2006 for State College and Bellefonte in Central Pennsylvania. Data have been obtained from two sources. A first source consists in the full price schedule of the utility publisher (Verizon) provided by the Yellow Page Association, which is an industry trading group. This data contains detailed information on the different advertising options and their prices proposed by Verizon to businesses.

A second source consists in the phone directory published by Verizon Information Services. The demand for the different advertising categories is directly constructed from the directory.[18] Defining each phone number as an advertisement, we collected a total of 7,214 advertisements distributed over 1,189 different industry headings. A total of 215,414 copies of such a directory is distributed. The advertisements bought by firms generate a revenue of nearly 6 million dollars. Table 1 presents the top 10 industry headings in terms of generated revenue. These ten headings represent 29% of the total revenue. Not surprisingly, we find professionals such as attorneys, dentists, veterinarians, services to households such as plumbers, auto repairs, carpet cleaners, and restauration services such as hotels and restaurants.

The area of State College and Bellefonte is also covered by another directory from a non-utility publisher. This second directory covers the whole Centre County of Pennsylvania, which includes four additional boroughs. The publisher distributes only 72,000 copies of such a directory, and the volume of yellow pages is much smaller than the Verizon directory. The yellow page association also provides information on their price schedule. This second publisher proposes less options and charges a significantly lower price. As a first approximation, it is reasonable to assume that Verizon acts as a monopoly. The analysis of competition is left for future research.[19]

4.1 Description of the Price Schedule

The Verizon directory proposes a large number of different advertising options from which firms can choosein addition to the standard listing which is free of charge. The lowest

[18] After a careful check with Verizon Information Services, the price schedule provided by the Yellow Page Association is strictly enforced.

[19] Competition could be analyzed with a model in which Verizon would be the dominant firm on the market and the non-utility publisher a follower as in a Stackelberg game. Following Ivaldi and Martimort (1994), Miravete and Roller (2004) estimates a duopoly with differentiated products.

positive price is \$100.8 for an extra line in the standard listing in the normal font, while the largest price is \$60,002.4 for a double display page with multiple colors. Each advertising option is defined by a vector of characteristics which are the size, the color and some other special features. This subsection describes the different advertising options and their prices.

In terms of size, the yellow page industry is using three categories, namely listing, space listing and display. The listing refers to the name, address and phone number of the firm under appropriate industry heading. A listing is typically a line in a column. Four different font sizes are available. Firms can choose the normal font, which is the smallest size offered. Such a listing with the normal font is called the regular or standard listing. This service is proposed to firms at zero price by law as all the firms need to be included in the yellow pages. This option is chosen by 2,347 businesses, which accounts for 32.5% of all advertisements or firms. Firms may choose to add some extra lines to their listing thereby increasing the advertising size. They may also choose larger font sizes, which also increases the advertising size. These extra lines and larger fonts are paid by the firms.

The space listing allocates a space within the column under the corresponding heading, while display advertisement provides a listing under the heading and allocates another space somewhere else. This additional space can cover up to two pages. There are five sizes available within the category of space listing and nine sizes available within the category of display. Display advertisements are usually larger than space listing advertisements. Although one may argue that the location of the advertisement on the page may contribute to its effectiveness, only the advertising space matters for the publisher and in its price schedule. Firms which purchase displays are also entitled to have their name, address and phone numbers in the listing. They can increase their listing size by having extra lines and/or larger font sizes.

The different size options are measured in square picas, which is the unit commonly used in the publishing industry, namely one pica corresponds approximately to 1/6 inch. As an illustration, a standard listing is 12 square picas, and a full page is 3,020 square picas. Table 2 presents the advertising sizes purchased, the number of purchases and the generated revenue for all the sizes chosen by the firms. The sizes also include the listing size for firms purchasing display. This makes a total of 47 different advertising sizes bought by firms ranging from 12 to 6066 square picas. As noted previously, 32.5% of all firms choose the free advertising option and obtain 12 square picas of advertising size. Another 33.8% choose to add some extra lines and/or larger fonts to their listing generating 11.1% of the publisher's revenue. Space listing is chosen by 19.8% of the firms generating 17% of the publisher's revenue. The largest part of the publisher's revenue comes from displays, which generate the remaining 71.9% of the publisher's revenue while being chosen by 9.9% of the firms. We note that the latter always add some extra features to their listing. A Lorenz curve would show a striking inequality in terms of generated revenues. We also note the heterogeneity in demand. About 89% of the firms buy a rather small advertising size, i.e. less than 7% of a page, while 1.7% of the firms buy more than 25% of a page.

In addition to size, the different advertising options contain a color dimension. Five color categories are available: no color, one color, white background, white background plus one color, multiple colors including photos. These different color options are not available for all the above sizes. For instance, the multicolor option is offered for displays only. The colors black and yellow are the standard colors in the yellow page industry and are not counted as additional colors.[20] In our data, 99.5% of firms choosing the listing category also choose the no color option. This number slightly decreases to 96.2%

[20]Busse and Rysman (2004) consider only the price schedule for no color display advertising in their study.

for firms choosing the space listing category. On the other hand, only 44.1% of firms choosing the display category have opted for the no color option. Color counts for an important difference in the advertising price for a given size. For instance, one display page with no color costs \$18,510 increasing up to \$32,395 with multiple colors. Besides color, there are two other qualitive features, namely guide and anchor listing. Guide is offered to complement listing and space listing advertising only. Guide is placed at the end of each industry heading. Advertisements under guide are grouped by the specialty in which they practice and the name of the specialty is highlighted in blue color. For example, at the end of the industry heading Dentists, there are listings that group dentists by their specialties such as endodontics and the subheading endodontics is highlighted. The guide option may increase the advertising price up to 30%. For instance, a 108 square picas of space listing with no color is priced \$1,134, while the same size with the guide is priced \$1,462. If we consider the option with one additional color (red), the price varies from \$1,789 to \$2,306. The guide option is chosen by about 4.3% of firms for the listing and space listing categories. Anchor listing is provided for displays only. As mentioned above, each display is provided a listing under each heading, which is called the anchor listing. Firms can add extra features in their anchor listings, such as a solid star to make the reference to the display advertisement more visible. Three options are available with a price ranging from \$366 to \$832. This option is chosen by 14.6% of firms in the display category. Lastly, there are additional advertising features such as coupon pages, trade marks and advertising on the front and back covers of the phone directory. Our price schedule data provide incomplete information for these categories and we cannot assign a price for these three categories. These additional features concern a total of 283 observations out of the 7,214 firms, which is rather small. This leaves us with a total of 6,931 observations as displayed in Table 2. Because it concerns a rather small number of observations and because we have incomplete price data for these categories, we have chosen to exclude these 283 observations in our empirical study.

Size, color, guide and anchor listing offer a very large number of possible combinations for firms to choose from. We observe in the data 149 chosen combinations leading to 149 different prices paid by firms. Figure 1 displays the scatter plot of the various price schedules as a function of size measured in square picas for different categories. For a given category, we observe that the price per square pica decreases as the advertising size increases as predicted by the curvature of the optimal price schedule in nonlinear pricing models. For instance, from the price schedule data, a firm buying a double page with no color pays \$5.68 per square picas, a firm buying a page with no color pays \$6.13 per square picas, a firm buying a half page pays \$6.80 and so on. When comparing the price per square picas for no color display for the lowest and largest size offered, i.e. 174 and 6039, the price varies from \$9.41 down to \$5.68 corresponding to a reduction of 66%. The same decreasing pattern is observed for other categories.[21] This corresponds to an important discount for quantity. As shown in Section 3, this curvature is crucial to identify nonparametrically the model. Moreover, without such a curvature, there will be a mismatch between the model and the data. The advertisements vary through several characteristics, while our model in Section 2 has a one dimension continuous quantity q. The next subsection proposes a method to include the different advertising categories while still having a single dimension continuous quantity.

4.2 Quality-Adjusted Quantity

From the previous subsection, advertisements differ in size and other aspects. Information such as color is difficult to aggregate through standard methods as it seems heroic to attach an index or value to it. In this respect, it is interesting to review how quality has been

[21] A similar pattern can be seen in Table 2. As Table 2 mixes various categories for a given size, the results might be misleading.

integrated in theoretical models of nonlinear pricing. Maskin and Riley (1984) consider a situation where the monopolist can discriminate consumers by an optimal bundling of the quantity and quality levels of the product, both taking continuous values. They show that the optimal bundle of quantity and quality should line on a unique curve in the quantity-quality space. Moreover, quantity should increase with quality along this curve. We can reasonably expect that this result extends to a situation in which quality takes discrete values though some pooling will naturally arise. Nonetheless, if both quantity and quality are used to discriminate firms, the same quantity cannot be associated with two different qualities because of the increasing relationship between the two variables. First, we observe in our price schedule data that the publisher offers various qualities for each size. Second, we do not observe a perfect correlation between the number of colors and the advertising size chosen by firms. As a matter of fact, except for the 2 pages size, firms choose different color options for a given advertising size contradicting the prediction of the model that larger advertising sizes should pair with more colors. Another solution would be to consider a nonlinear pricing model in which every firm is characterized by several types dictating its taste for quantity and the various colors. This will give rise to a complex model with multidimensional screening, which is known to require restrictive assumptions to be solved including particular functional forms for the model structure.

We argue instead that the publisher does not provide different qualities to discriminate among firms. Technological constraints impose a limited number of advertising sizes that can be offered as displayed in Table 2 with 47 different sizes. As a result, the size cannot be offered on a continuous scale leaving some room for additional discrimination as it is reasonable to assume that firms' willingness to pay for advertising takes a continuous value. Thus, a good strategy for the publisher is to offer additional options for each available size to fill up the holes in the size scale.[22] As a result, the publisher offers a large

[22]Table 2 shows that the range 12-108 is relatively well covered with 22 different size options. This

number of different options to choose from such that each firm can find an advertising option corresponding to its willingness to pay. Without these additional options, there will be some pooling. Offering various qualities then improves the discriminatory process as it increases the number of options for firms. The price schedule data favor such an argument. Figure 1 shows some curvature in size suggesting that size is used as a discriminatory variable in the sense that an important discount is offered per unit of advertising to large buyers. If quality would have been used as a second discriminatory variable per se, we would have observed some variations in the price schedule for quality as well. For instance, the publisher could have taken advantage of the willingness to pay of some firms for the pair (large quality, large size) relative to other options. Figure 1 suggests the contrary as the publisher seems just to report the additional cost due to different qualities in the prices. Specifically, we observe that the ratio of prices for two different qualities remains the same across sizes. For instance, considering the no color and one color display price schedules, the price ratio is almost constant and equal to 1.5 across the 9 possible sizes. Similarly, when considering the no color and multicolor display price schedules, the ratio is almost constant and equal to 1.75 across sizes. In view of economic theory, this strategy might be not optimal as both quality and quantity could have been used to discriminate firms optimally. The data suggest that the publishing company has proposed various qualities to increase the number of options without using the quality as a discriminaotry variable per se.

Following this empirical evidence, we construct a quality-adjusted quantity index. To do so, we consider the price schedule for multi colors as the continuous price schedule offered by the publisher and we adjust the advertising sizes for other color options in view of the price schedule for the multi colors. Thus q will be the multi colors adjusted

number dramatically decreases for sizes above 108. Many range values such as 590-730, 790-1100 or 1150-1480 do not offer any size option.

size. The sizes for all other colors will be assigned lower values by using a fitted nonlinear function for the multi colors price schedule. As a result, a one page in one color will correspond to a lower quality-adjusted size. Some examples will be used later to illustrate our method. Our choice of the price schedule for multi colors is made without loss of generality. We could have chosen the price schedule for the no color as well though the fitted function would not have been estimated on the same range of values for the other price schedules. The multi colors price schedule avoids this problem. In particular, all quantities measured with less colors can readily correspond to quantities on that curve. We use a nonlinear function to fit the multi colors price schedule. The fitted nonlinear function is

$$\widehat{\log(T)} = 4.1602 + 0.7317 \times \log(q) + 0.0062 \times (\log(q))^2, \tag{25}$$

where T is the price in dollars and q is the advertising size measured in square picas. The coefficients are estimated using an ordinary least squares estimator. The R^2 of such a regression is 0.999, which is almost a perfect fit.[23] This function gives the price schedule $T(\cdot)$ in the theoretical model of Section 2. We note that without the technological constraint limiting the number of q offered, the price schedule would have provided a continuous function $T(\cdot)$ with (say) 1,000 different values of q. We need to fit $T(\cdot)$ because of the limited number of quantities offered. The fit we obtain is quite remarkable and can be considered as a reasonable approximation of the price schedule $T(\cdot)$. As such, we consider that the measurement error it may introduce in the econometric model cen be ignored.

The quality-adjusted quantities are constructed as follows. We plug on this curve all the observed prices for other qualities to solve for the quality-adjusted quantities. As an illustration, a one-page advertising display measuring 3,020 square picas with no color and

[23]A least squares estimator is appropriate since there is no endogeneity problem in this equation. The variable q is not endogenous because $T(q)$ is a deterministic function.

no additional feature now corresponds to a multi colors adjusted quantity of 1,470 picas. Similarly, a half-page advertising display of 1,485 square picas with one color corresponds now to a multi colors adjusted quantity of 1,153 square picas.[24]

This method has the main advantage of avoiding the difficulty if not the impossibility of aggregating multiple quantitative and qualitative characteristics. For instance, standard dimension reducing methods such as principal component analysis work poorly here since there is no obvious way to evaluate the relative importance of the characteristics and there is no strong pattern of correlation between the various characteristics.

5 Empirical Results

Using the quality-adjusted quantities, the price schedule and the firm's purchases, we apply the estimators (18) to (21). The values for $T'(\cdot)$ and $\hat{\gamma} = T'(q_{\max})$ are obtained using (25). From Table 2, $q_{\max}$ is equal to 6066 and we obtain $T'(6066) = 8.3159 = \hat{\gamma}$. This value, which is the marginal cost in dollars for an additional quality-adjusted square pica at the total production, seems to be reasonable for the publishing industry. This means that an additonal line of listing or 1.85 quality-adjusted square picas costs at the margin for the publisher \$15.38, which can be compared to \$100.8 charged to the firm. Data from a single phone book do not allow us to recover more of the cost function. If data from several phone books published by the same company would be collected satisfying some criteria of homogeneity such as population size, we could recover the cost function at several values of Q. We can then hope to identify $C(\cdot)$ partially.

The estimated marginal utility $\hat{v}_0(\cdot)$ and the firm's type $\hat{\theta}(\cdot)$ functions are displayed in Figures 2 and 3, which are obtained after standard smoothing for readability. Figure 2

[24] For $T = 0$ corresponding to the standard listing, we consider a 10^{-2} precision corresponding to 1 cent. It gives a quality-adjusted q_0 of 1.329×10^{-6}.

shows that $\hat{v}_0(\cdot)$ is strictly decreasing as assumed by the theoretical model, while Figure 3 shows that $\hat{\theta}(\cdot)$ is strictly increasing as predicted by the model.[25] The estimated (untruncated) density of types $\hat{f}(\cdot)$ obtained from the pseudo values (21) and using (22) is displayed in Figure 4. The estimated density indicates the existence of substantial heterogeneity among firms in terms of their taste for advertising. We note that the density displays two modes, a first one around 1.3 and a second one around 2.4. The first mode corresponds to firms which choose an advertising listing while adding a few lines and/or a larger font. The second mode corresponds to firms which choose the two smallest sizes of space listing. We also note the concentration of observations to the left of 3. The shape of the type density shows other irregularities such as a little bump at 6, which correspond to firms choosing between the third (around one fifth of a page) and fifth (around a quarter of a page) smallest display sizes.

With the estimates of $v_0(\cdot)$ and $\theta(\cdot)$, we can estimate the utility surplus for each firm using $U(q,\theta) = U_0 + \theta[\int_{q_0}^{q} v_0(x)dx] - T(q)$. The utility U_0 of being advertized is not identified. Given that U_0 is an additive constant to all firms' utilities, we can still make a comparison of utility across firms. Some estimated values of $\hat{U} - U_0$ for no color are given in Table 3. The firm's utility has been obtained while using the unsmoothed $\hat{v}_0(\cdot)$. Thus, it can be viewed as a lower bound for the firm's utility. We remark that the firm's utility increases with the advertising size purchased. Figure 5 displays the estimated firm's utility $\hat{U} - U_0$ as a function of the quality-adjusted quantity. We note that the firm's utility strictly increases with the quality-adjusted advertising size purchased. The firm's utility $\hat{U} - U_0$ takes values between \$-25.74 and $\$27,123.80$ and the mean value is equal to \$502.28. We note that these values are quite large relative to what firms pay for their advertisements. The economic interpretation of the firm's utility is the rent that the publisher leaves to the firm to reveal its type. In particular, in a world of complete

[25] The unsmoothed graphical representations of $\hat{v}_0(\cdot)$ and $\hat{\theta}(\cdot)$ are available upon request to the authors.

information, the publisher knows each firm's type and the price that the publisher charges to each firm should be equal to each firm's utility, i.e. $\theta \int_0^q v_0(x)dx$ leaving zero rent to the firm. Thus, by summing up the estimated informational rents across firms, we empirically assess the magnitude of the information rent or the cost of asymmetric information to the publisher. Such an estimate is equal to \$3,476,780.2 representing 37.71% of the publisher's revenue under complete information, which would be \$9,219,089 (\$3,476,780+\$5,742,309).

With structural estimates, we can perform some counterfactuals. For instance, we can simulate the publisher's revenue and firms' utility under alternative pricing rules. A standard pricing rule would be to apply linear prices, i.e. any size will be charged the same amount at the margin. We can rewrite the publisher's problem by considering $T(q) = pq$. Thus, $v_0(q) = p/\theta$. Since we do not have an estimate of the cost function except for the marginal cost at the total production, hereafter we assume that we have a constant marginal cost, which is equal to $\hat{\gamma}$. Because $f(\cdot)$ is not identified bewlo θ_0, we make abstraction of the problem of optimal exclusion and still consider the value of θ_0 as the threshold type to purchase any advertising. A grid search allows us to obtain an estimate for the linear price p that maximizes the publisher's revenue. We obtain \$20.20 per quality-adjusted square picas. We then need to simulate for each firm its demand when facing a linear price, i.e. $q = v_0^{-1}(p/\theta)$, where θ and $v_0(\cdot)$ are replaced by their estimated counterparts.[26] As expected, the firms purchasing small quantities under the current price schedule tend to buy larger ones. On the other hand, the firms purchasing large advertisements under the current price schedule tend to buy smaller ones. For instance, the maximum of the size purchased becomes 740 square picas instead of 6066 square picas unde the current price schedule. These correspond to a half a page with no color and a double page multi colors, respectively. In terms of publisher's revenue, a

[26]We consider that the quality-adjusted quantity q is continuous. We could have considered some rule such as assigning to the firm the lowest closest quantity observed in the data.

linear price would generate \$4,161,682, which represents a loss of 27.5%. For the firms, the mean value for utility $\hat{U} - U_0$ becomes \$672.02, which is larger than under nonlinear pricing. This result is due to the right skewness of firms' type density. In particular, linear pricing tends to favor the small buyers.

Another interesting policy evaluation would be to consider that the publisher is not required to include all businesses phone numbers. Thus, the publisher can choose optimally the minimal advertsing size it will propose, which is likely to be larger than q_0. The problem has the flavor of optimal exclusion but the minimum quantity is determined endogenously. *to be completed.*

6 Concluding Remarks

This paper extends the standard model of nonlinear pricing incorporating some features of the yellow page advertising industry such as the requirement to incorporate all the businesses' phone numbers at zero price. Nonparametric identification is established under a multiplicative decomposition of the firm's marginal utility function. The shape of the price schedule is crucial to identify the marginal base utility function and the firm's unknown type. Nonparametric identification is achieved by exploiting the first-order conditions of the publisher's profit maximization and the unique mapping between the firm's type and its purchased quantity. This result is in the spirit of the recent literature on the nonparametric identification of incomplete information models such as first-price auctions though our identification result is more involved as an additional function needs to be identified. Following the nonparametric identification, a nonparametric two-step estimation procedure is developed to estimate the marginal utility function and the type distribution. Asymptotic properties of our estimator are establshed. In particular, it is shown that the estimator of the marginal utility function is uniformly consistent while its asymptotic distribution converges at the parametric rate.

Data on advertising in yellow pages from a utility publisher in Central Pennsylvania are analyzed. Because the options offered to firms also integrate a quality dimension, we construct a quality-adjusted advertising size to incorporate the various qualities offered to firms. Empirical results show a decreasing marginal utility and an increasing type in the purchased quantity as expected from the model thereby suggesting that the model is not rejected by the data. Some counterfactuals provide an estimate for the cost of asymmetric information for the publisher and assess the loss in terms of publisher's revenue and firms' utility surplus or rents when a linear price is adopted and when the publisher does not have the obligation to include everyone in the phone book.

As mentioned earlier, the quality-adjusted quantity may introduce some measurement error problem. An error term of zero mean could be added to the quantity purchased. Because this error term would enter nonlinearly in the model and at some bounds of integrals, the problem becomes complex and nonstandard. Though the large number of observed prices allows us to consider a continuous price schedule as a reasonable approximation, an alternative model would be to consider a discrete price schedule given an exogenous number of price options. Another extension of interest is to test adverse selection in this market. To perform a structural test of adverse selection, we would need to derive the restrictions imposed by the nonlinear pricing model under both incomplete and complete information environments. We then can test which set of restrictions are validated by the data. From a theoretical perspective, we have avoided the problem of countervailing incentives by considering an opportunity value independent of the firm's type in the firm's individual rationality constraint. Our identification and estimation results could be extended to incorportate this case. Another extension could study the role of competition by incorporating a second publisher in the model. In the county where we collected the data, a non-utility publisher distributes a second phone book though of less importance and distributed at a significantly lower number. We can collect data on the price schedule and demand data for this non-utility directory. The utility publisher seems to play the role of a leader in this market. A Stackelberg model could shed some light on the competition. Lastly, the methodology developed in this paper could be extended to other models with adverse selection in contract theory.

Appendix

Proof of Lemma 1: As discussed in the text, under A1 and (2), all the IC constraints hold globally except (4), which is defined only locally. To show that (4) also holds globally, we first show that the local second-order condition $U_{11}(\theta, \theta) \leq 0$ is satisfied. By differentiating the first-order condition (3) with respect to θ, we obtain $U_{12}(\theta, \theta) + U_{11}(\theta, \theta) = 0$. Hence, $U_{11}(\theta, \theta) \leq 0$ is equivalent to $U_{12}(\theta, \theta) \geq 0$. Since

$$U_{12}(\theta, \theta) = U_{21}(\theta, \theta) = v_2(q(\theta), \theta)q'(\theta) > 0,$$

the local second-order condition is satisfied under assumption A1-(iii) and $q'(\cdot) > 0$.

To show that the second-order condition also holds globally, we use a contradiction argument. Let θ_1 and θ_2 satisfy $\theta_0 < \theta_1 < \theta_2 \leq \overline{\theta}$. If $U(\theta_2, \theta_1) > U(\theta_1, \theta_1)$, we have

$$\int_{\theta_1}^{\theta_2} U_1(x, \theta_1)dx > 0. \tag{A.1}$$

We show that $U_1(x, \theta_1) \leq 0$ for $x \in [\theta_1, \theta_2]$ hence leading to a contradiction. By definition, $U(\tilde{\theta}, \theta) = \int_0^{q(\tilde{\theta})} v(x, \theta)dx - T(q(\tilde{\theta}))$. Thus we have

$$U_{12}(\tilde{\theta}, \theta) = U_{21}(\tilde{\theta}, \theta) = v_2(q(\tilde{\theta}), \theta)q'(\tilde{\theta}) > 0.$$

Hence, $U_1(x, \theta_1) \leq U_1(x, x) = 0 \quad \text{for} \quad x \geq \theta_1$, where the second equality results from the first-order condition of the IC constraint (3). This contradicts (A.1). Thus, $U(\theta_2, \theta_1) \leq U(\theta_1, \theta_1)$ for $\theta_0 < \theta_1 < \theta_2 \leq \overline{\theta}$.

Similarly, let θ_1 and θ_2 satisfy $\theta_0 < \theta_2 < \theta_1 \leq \overline{\theta}$. If $U(\theta_2, \theta_1) > U(\theta_1, \theta_1)$, we have

$$\int_{\theta_2}^{\theta_1} U_1(x, \theta_1)dx < 0.$$

But $U_1(x, \theta_1) \geq U_1(x, x) = 0 \quad \text{for} \quad x \leq \theta_1$ by a similar argument, leading to a contradiction. Thus the local second-order condition holds globally. □

Proof of Proposition 1: We make a change of variable in the publisher's problem (7) such that the choice variable θ_0 becomes the terminal time. We define $t = \overline{\theta} - \theta$ and $t_0 = \overline{\theta} - \theta_0$. The publisher's profit function (7) becomes

$$\begin{aligned}\Pi &= -\int_{t_0}^{0}\left\{\left[\int_{q_0}^{q(\overline{\theta}-t)} v(x,\overline{\theta}-t)dx\right]f(\overline{\theta}-t)-\left[1-F(\overline{\theta}-t)\right]\left[\int_{q_0}^{q(\overline{\theta}-t)} v_2(x,\overline{\theta}-t)dx\right]\right\}dt\\ &\quad -C\left[q_0F(\overline{\theta}-t_0)-\int_{t_0}^{q_0} q(\bar{\theta}-s)f(\bar{\theta}-s)ds\right].\end{aligned}$$

We further define $\overline{q}(t) \equiv q(\overline{\theta}-t)$, $\overline{v}(x,t) \equiv v(x,\overline{\theta}-t)$, $\overline{f}(t) \equiv f(\overline{\theta}-t)$, $\overline{F}(t) \equiv 1-F(\overline{\theta}-t)$ and $\overline{v}_2(x,t) = -v_2(x,\overline{\theta}-t)$ $\forall t \in \left[0,\overline{\theta}-\underline{\theta}\right]$. The publisher's problem becomes

$$\begin{aligned}\max_{\overline{q}(\cdot),t_0\in\left[0,\overline{\theta}-\underline{\theta}\right]} \Pi &= \int_0^{t_0}\left\{\left[\int_{q_0}^{\overline{q}(t)} \overline{v}(x,t)dx\right]\overline{f}(t)+\overline{F}(t)\left[\int_{q_0}^{\overline{q}(t)} \overline{v}_2(x,t)dx\right]\right\}dt\\ &\quad -C\left\{q_0\left[1-\overline{F}(t_0)\right]+\int_0^{t_0}\overline{q}(s)\overline{f}(s)ds\right\}.\end{aligned} \tag{A.2}$$

We treat $\overline{q}(t)$ as the control variable and $\int_0^t \overline{q}(s)\overline{f}(s)ds$ as the state variable. The maximization problem (A.2) can be written as a standard free terminal time and free-end point control problem.

To simplify the expression, we define

$$\begin{aligned}X(t) &= \int_0^t \overline{q}(s)\overline{f}(s)ds \quad \forall t \in \left[0,\overline{\theta}-\underline{\theta}\right]\\ \Psi\left[\overline{q}(t),t\right] &= \left[\int_{q_0}^{\overline{q}(t)} \overline{v}(x,t)dx\right]\overline{f}(t)+\overline{F}(t)\left[\int_0^{\overline{q}(t)} \overline{v}_2(x,t)dx\right] \quad \forall t \in \left[0,\overline{\theta}-\underline{\theta}\right]\\ K\left[X(t_0),t_0\right] &= -C\left\{q_0\left[1-\overline{F}(t_0)\right]+X(t_0)\right\}.\end{aligned}$$

The maximization problem (A.2) can be written as

$$\max_{\overline{q}(\cdot),t_0\in\left[0,\overline{\theta}-\underline{\theta}\right]} \Pi = \int_0^{t_0}\Psi\left[\overline{q}(t),t\right]dt+K\left[X(t_0),t_0\right]. \tag{A.3}$$

The Hamiltonian function for the above maximization problem is

$$\mathcal{H}\left[X(t),\overline{q}(t),\lambda(t),t\right] = \Psi\left[\overline{q}(t),t\right]+\lambda(t)\overline{q}(t)\overline{f}(t), \tag{A.4}$$

where $\lambda(t)$ is the multiplier function. According to the maximization principle in Kirk (1970, pp. 188 and 192), the necessary conditions for the functions $X(t), \overline{q}(t)$ and $\lambda(t)$ to be the solutions of (A.3) are

$$\mathcal{H}_2\left[X(t),\overline{q}(t),\lambda(t),t\right] = 0 \quad \forall t \in [0,\overline{\theta}-\theta_0)$$

$$
\begin{aligned}
\lambda'(t) &= -\mathcal{H}_1\left[X(t),\overline{q}(t),\lambda(t),t\right] \quad \forall t\in[0,\overline{\theta}-\theta_0)\\
\lim_{t\uparrow t_0}\lambda(t) &= \lim_{t\uparrow t_0}K_1\left[X(t),t\right]\\
\lim_{t\uparrow t_0}\mathcal{H}\left[X(t),\overline{q}(t),\lambda(t),t\right] &= -\lim_{t\uparrow t_0}K_2\left[X(t),t\right].
\end{aligned} \tag{A.5}
$$

By definition of the Hamiltonian function (A.4) and the function $K\left[X(t),t\right]$, the first three necessary conditions give

$$
\overline{v}(\overline{q}(t),t)\overline{f}(t)+\overline{F}(t)\overline{v}_2(\overline{q}(t),t)+\lambda(t)\overline{f}(t)=0 \quad \forall t\in[0,\overline{\theta}-\theta_0) \tag{A.6}
$$

$$
\lambda'(t)=0 \quad \forall t\in[0,\overline{\theta}-\theta_0) \tag{A.7}
$$

$$
\lim_{t\uparrow t_0}\lambda(t)=-C'\left\{q_0\left[1-\overline{F}(t_0)\right]+X(t_0)\right\}. \tag{A.8}
$$

Equations (A.7) and (A.8) give the optimal $\lambda(\cdot)$

$$
\lambda(t)=-C'\left\{q_0\left[1-\overline{F}(t_0)\right]+X(t_0)\right\} \quad \forall t\in[0,\overline{\theta}-\theta_0). \tag{A.9}
$$

Plugging (A.9) into (A.6) gives the optimal $\overline{q}(\cdot)$

$$
\overline{v}(\overline{q}(t),t)=C'\left\{q_0\left[1-\overline{F}(t_0)\right]+\int_0^{t_0}\overline{q}(s)\overline{f}(s)ds\right\}-\frac{\overline{F}(t)}{\overline{f}(t)}\overline{v}_2(\overline{q}(t),t) \quad \forall t\in[0,\overline{\theta}-\theta_0). \tag{A.10}
$$

Plugging the optimal $\lambda(\cdot)$ into the Hamiltonian function (A.4) and letting $\lim_{t\uparrow t_0}\overline{q}(t)=\overline{q}_-$ gives the following expression for the left-hand side of (A.5)

$$
\left[\int_{q_0}^{\overline{q}_-}\overline{v}(x,t_0)dx\right]\overline{f}(t_0)+\overline{F}(t_0)\left[\int_{q_0}^{\overline{q}_-}\overline{v}_2(x,t_0)dx\right]-C'\left\{q_0\left[1-\overline{F}(t_0)\right]+\int_0^{t_0}\overline{q}(s)\overline{f}(s)ds\right\}\overline{q}_-\overline{f}(t_0).
$$

By definition of $K[X(t_0),t_0]$, the right-hand side of (A.5) is $-C'\left\{q_0\left[1-\overline{F}(t_0)\right]+X(t_0)\right\}q_0\overline{f}(t_0)$. After equating the two terms and rearranging, we obtain

$$
\int_{q_0}^{\overline{q}_-}\left[\overline{v}(x,t_0)+\frac{\overline{F}(t_0)}{\overline{f}(t_0)}\overline{v}_2(x,t_0)\right]dx=\left[\overline{q}_- - q_0\right]C'\left\{q_0\left[1-\overline{F}(t_0)\right]+\int_0^{t_0}\overline{q}(s)\overline{f}(s)ds\right\}. \tag{A.11}
$$

Plugging (A.10) evaluated at t_0 into (A.11) gives

$$
\int_{q_0}^{\overline{q}_-}\left[\overline{v}(x,t_0)+\frac{\overline{F}(t_0)}{\overline{f}(t_0)}\overline{v}_2(x,t_0)\right]dx=\left[\overline{q}_- - q_0\right]\left[\overline{v}(\overline{q}_-,t_0)+\frac{\overline{F}(t_0)}{\overline{f}(t_0)}\overline{v}_2(\overline{q}_-,t_0)\right].
$$

Letting $\Gamma(x,t_0) \equiv \overline{v}(x,t_0) + [\overline{F}(t_0)/\overline{f}(t_0)]\overline{v}_2(x,t_0) = v(x,\theta_0) - [(1-F(\theta_0))/f(\theta_0)]v_2(x,\theta_0)$, the previous expression can be written as

$$\int_{q_0}^{\overline{q}_-} [\Gamma(x,t_0) - \Gamma(\overline{q}_-,t_0)]\, dx = 0. \tag{A.12}$$

We now invoke A3. If $\Gamma(x,t_0)$ is monotone in x, the integrand is strictly monotone. Thus, (A.12) implies $\overline{q}_- = q_0$. Alternatively, if $\Gamma(\cdot,t_0) = 0$, then the left-hand side of (A.11) is zero implying that $\overline{q}_- = q_0$ because $C'(\cdot) > 0$ by A2.

It suffices now to rewrite (A.10) and (A.13) in terms of $q(\cdot)$, $v(\cdot,\cdot)$ and $F(\cdot)$ to establish (8) and (9). Equation (10) is nothing else than the IC constraint (4). Equation (11) is obtained by using $U_+ = U_0$, which is equivalent to $\lim_{\theta\downarrow\theta_0}\int_{q_0}^{q(\theta)} v(x,\theta)dx = \lim_{\theta\downarrow\theta_0} T(q(\theta))$. The left-hand side is equal to zero using (9) thereby establishing (11) since $q(\cdot)$ is strictly increasing on $(\theta,\overline{\theta}]$ and (9) holds. □

Proof of Lemma 2: To simplify the exposition, we suppress the arguments of functions in the following proof. Taking the total derivative of (8) with respect to θ gives

$$v_1 q' + v_2 = \frac{\partial(v_2/\rho)}{\partial\theta} + \frac{v_{21}}{\rho}q',$$

where $\partial(v_2/\rho)/\partial\theta = (v_{22}/\rho) + v_2(\partial(1/\rho)/\partial\theta)$. Rearranging terms gives

$$q' = \frac{v_2\left[(1/v_2)\times(\partial(v_2/\rho)/\partial\theta) - 1\right]}{v_1 - (v_{21}/\rho)}. \tag{A.13}$$

The numerator is negative by A1-(iii) and A4-(ii). We want to show that the denominator is also negative. Consider now the denominator.

Suppose $v_{21} < 0$. From (8), A2 and A1-(iii), we have $(v/v_2) > (1/\rho) > 0$ implying $v_1 - (v_{21}/\rho) < v_1 - [(v_{21}v)/v_2]$. Moreover, $v_1 - [(v_{21}v)/v_2] = (v^2/qv_2)\times(\partial(-qv_1/v)/\partial\theta)$ since

$$\frac{\partial(-qv_1/v)}{\partial\theta} = q\frac{v_1v_2 - v_{12}v}{v^2} = \frac{qv_2}{v^2}\left(v_1 - \frac{v_{12}v}{v_2}\right).$$

By A1-(iii) and A4-(i), $[v^2/(qv_2)]\times[\partial(-qv_1/v)/\partial\theta] \le 0$. Thus, $v_1 - (v_{21}/\rho) < 0$. Hence under A1-(iii), A2 and A4-(i,ii), $q' > 0$ on $(\theta_0,\overline{\theta}]$. Suppose $v_{21} \ge 0$. It is straightforward to see that the denominator of (A.14) is strictly negative by A1-(ii). Hence, under A1-(ii,iii), A2 and

A4-(ii), $q' > 0$ on $(\theta_0, \overline{\theta}]$. That $q(\cdot)$ is strictly increasing and continuous on $[\theta_0, \overline{\theta}]$ is obvious. We now show that $q(\cdot)$ is continuously differentiable at θ_0, i.e. $q'(\theta_0) = \lim_{\theta \downarrow \theta_0} q'(\theta) < \infty$ and strictly positive, and that $q'(\theta_0) > 0$. By the Mean Value Theorem, we have $q'(\theta_0) \equiv \lim_{\theta \downarrow \theta_0}[q(\theta) - q(\theta_0)]/(\theta - \theta_0) = q'(\tilde{\theta})$, where $\theta_0 < \tilde{\theta} < \theta$. Hence, $\lim_{\theta \downarrow \theta_0} q'(\theta)$ is equal to the right-hand side of (A.13) evaluated at θ_0, which is finite and strictly positive under A4-(i,ii). It remains to show that under A1-(iii), A4-(iii) and A4-(iv) imply A4-(ii). Assumption A4-(ii) is equivalent to

$$\frac{1}{v_2}\frac{\partial(v_2/\rho)}{\partial\theta} - 1 = \frac{v_{22}}{v_2\rho} - \left(1 + \frac{\rho'}{\rho^2}\right) < 0.$$

Since $1 + (\rho'/\rho^2) > 0$ by A4-(iv) and $v_{22}/(v_2\rho) \leq 0$ by A1-(iii) and A4-(iii), the above expression is negative as desired.

Regarding the second statement, because $q(\cdot)$ is strictly increasing on $(\theta_0, \overline{\theta}]$ by Lemma 2, it follows that the IC constraint (4) can be written as $T'(q) = v(q, \theta(q))$ $\forall q \in (q_0, \overline{q}]$. Recall that $v(\cdot,\cdot)$ is continuously differentiable on $[q_0, +\infty) \times [\underline{\theta}, \overline{\theta}]$ by A1 while $\theta(\cdot)$ is continuously differentiable on $[q_0, \overline{q}]$ as $\theta'(q) = 1/q'(\theta)$ with $q'(\cdot)$ strictly positive and continuous on $[\theta_0, \overline{\theta}]$ as noted above. Thus, $T'(\cdot)$ is continuously differentiable on $(q_0, \overline{q}]$, i.e. $T(\cdot)$ is twice continuously differentiable on $(q_0, \overline{q}]$. We now show that $T(\cdot)$ is twice continuously differentiable at q_0 with $T'(q_0) > 0$. We note that $T'(\cdot)$ exists and is continuous at q_0. This follows by the same Mean Value Theorem argument used above replacing $q(\cdot)$ by $T(\cdot)$ and using (4) since $T(\cdot)$ is continuous at q_0 by (11) and continuously differentiable on $(q_0, \overline{q}]$. Thus, $T'(\cdot)$ is continuously differentiable on $[q_0, \overline{q}]$ as $v(\cdot,\cdot)$ and $\theta(\cdot)$ are continuously differentiable on $[q_0, +\infty) \times [\underline{\theta}, \overline{\theta}]$ and $[q_0, \overline{q}]$, respectively. Hence, $T(\cdot)$ is twice continuously differentiable on $[q_0, \overline{q}]$. It remains to show that the assertions on $T'(\cdot)$. Combining (8) and (10) gives

$$T'(q) = C'(Q) + \frac{1 - F(\theta)}{f(\theta)} v_2(q, \theta) \quad \forall \theta \in (\theta_0, \overline{\theta}].$$

This establishes $T'(\cdot) > C'(Q)$ on $(q_0, \overline{q}]$ by A1-(iii), while taking the limit as $q \downarrow q_0$ gives $T'(q_0) = C'(Q)$. □

Proof of Lemma 4: Let $\tilde{\theta} = \alpha\theta$, which is distributed as $\tilde{F}(\cdot)$ on $[\underline{\tilde{\theta}}, \overline{\tilde{\theta}}] = [\alpha\underline{\theta}, \alpha\overline{\theta}]$. Let $\tilde{T}(\cdot) \equiv T(\cdot)$, $\tilde{q}(\cdot) \equiv q(\cdot/\alpha)$, $\tilde{\theta}_0 = \alpha\theta_0$ and $\tilde{Q} \equiv q_0\tilde{F}(\tilde{\theta}_0) + \int_{\tilde{\theta}_0}^{\overline{\tilde{\theta}}} \tilde{q}(u)\tilde{f}(u)du$. First we show that

$\tilde{T}(\cdot)$, $\tilde{q}(\cdot)$ and $\tilde{\theta}_0$ satisfy the necessary conditions (8), (9) (10) and (11). We then show that $\tilde{G}^*(\cdot) = G^*(\cdot)$, where $\tilde{G}^*(\cdot)$ is the truncated distribution of $\tilde{q}$, $\tilde{Q} = Q$, and $\tilde{F}(\tilde{\theta}_0) = F(\theta_0)$. Hence, the observables $\left[\tilde{T}(\cdot), \tilde{G}^*(\cdot), q_0, \tilde{Q}, \tilde{F}(\tilde{\theta}^*)\right]$ generated by the structure $\tilde{S}$ are the same observables $[T(\cdot), G^*(\cdot), q_0, Q, F(\theta_0)]$ generated by the structure S. To complete the proof, we show $\tilde{S} \in \mathcal{S}$.

To show $\tilde{T}'(\tilde{q}(\tilde{\theta})) = \tilde{\theta}\tilde{v}_0(\tilde{q}(\tilde{\theta}))$ for all $\tilde{\theta} \in (\tilde{\theta}_0, \overline{\tilde{\theta}}]$, we rewrite this equation using the definition of $\tilde{T}(\cdot)$, $\tilde{v}_0(\cdot)$ and $\tilde{q}(\cdot)$. This gives $T'(q(\tilde{\theta}/\alpha)) = (\tilde{\theta}/\alpha)v_0(q(\tilde{\theta}/\alpha))$ for all $\tilde{\theta} \in (\tilde{\theta}_0, \overline{\tilde{\theta}}]$, which is true because of (14) with $\theta = (\tilde{\theta}/\alpha) \in [\theta_0, \overline{\theta}]$. To show $\tilde{\theta}\tilde{v}_0(\tilde{q}(\tilde{\theta})) = C'(\tilde{Q}) + [(1 - \tilde{F}(\tilde{\theta})/\tilde{f}(\tilde{\theta})]\tilde{v}_0(\tilde{q}(\tilde{\theta})$ for all $\tilde{\theta} \in (\tilde{\theta}_0, \overline{\tilde{\theta}}]$, we rewrite this equation using the definition of $\tilde{v}_0(\cdot)$, $\tilde{q}(\cdot)$ and $\tilde{F}(\cdot)$:

$$\frac{\tilde{\theta}}{\alpha}v_0(q(\tilde{\theta}/\alpha)) = C'(\tilde{Q}) + \frac{1 - F(\tilde{\theta}/\alpha)}{f(\tilde{\theta}/\alpha)}v_0(q(\tilde{\theta}/\alpha))$$

for all $\tilde{\theta} \in (\tilde{\theta}_0, \overline{\tilde{\theta}}]$. If $\tilde{Q} = Q$, the above equation holds for all $\theta = \tilde{\theta}/\alpha \in (\theta_0, \overline{\theta}]$ in view of (13). Conditions (9) and (11) can be derived trivially.

Next, we show that the observables coincide. First, we show $\tilde{Q} = Q$. Using the definitions of $\tilde{F}(\cdot)$, $\tilde{q}(\cdot)$, $\tilde{\theta}_0$, $\overline{\tilde{\theta}}$ and $\tilde{f}(\cdot)$, we have

$$\tilde{Q} = q_0 F(\alpha\theta_0/\alpha) + \int_{\alpha\theta_0}^{\alpha\overline{\theta}} q(u/\alpha)\frac{1}{\alpha}f(u/\alpha)du = q_0 F(\theta_0) + \int_{\theta_0}^{\overline{\theta}} q(\theta)f(\theta)d\theta = Q.$$

Second, we show $\tilde{G}^*(\cdot) = G^*(\cdot)$. Namely,

$$\begin{aligned}
\tilde{G}^*(y) = \Pr(\tilde{q}(\tilde{\theta}) \leq y | \tilde{q}(\tilde{\theta}) > q_0) &= \Pr(\tilde{\theta} \leq \tilde{q}^{-1}(y) | \tilde{\theta} > \tilde{q}^{-1}(q_0)) \\
&= \Pr(\alpha\theta \leq \alpha q^{-1}(y) | \alpha\theta > \alpha q^{-1}(q_0)) \\
&= \Pr(\theta \leq q^{-1}(y) | \theta > q^{-1}(q_0)) \\
&= \Pr(q(\theta) \leq y | q(\theta) > q_0) = G^*(y),
\end{aligned}$$

using the monotonicity of $\tilde{q}(\cdot)$ and $q(\cdot)$. Third, $\tilde{F}(\tilde{\theta}^*) = F(\alpha\theta_0/\alpha) = F(\theta_0)$.

Lastly, we verify that the structure $\tilde{S}$ belongs to $\mathcal{S}$. Assumptions B1 and A2 are trivially satisfied. Regarding A3, we have

$$\tilde{\theta}\tilde{v}_0(\tilde{q}) - \frac{1 - \tilde{F}(\tilde{\theta})}{\tilde{f}(\tilde{\theta})}\tilde{v}_0(\tilde{q}) = \theta v_0(\tilde{q}) - \frac{1 - F(\theta)}{f(\theta)}v_0(\tilde{q}),$$

which is strictly monotone in $\tilde{q}$ or identically equal to zero for all $\theta \in (\theta_0, \overline{\theta}]$. Regarding A4-(iv), we have

$$\tilde{\theta} - \frac{1 - \tilde{F}(\tilde{\theta})}{\tilde{f}(\tilde{\theta})} = \tilde{\theta} - \frac{1 - F(\tilde{\theta}/\alpha)}{(1/\alpha) f(\tilde{\theta}/\alpha)} = \alpha \left[\frac{\tilde{\theta}}{\alpha} - \frac{1 - F(\tilde{\theta}/\alpha)}{f(\tilde{\theta}/\alpha)} \right],$$

which is strictly increasing in $\tilde{\theta}/\alpha$ and hence in $\tilde{\theta}$. □

Proof of Lemma 5: We first prove necessity. As explained in the text, because $q(\cdot)$ is strictly increasing in θ, we have $G^*(q) = [F(\theta) - F(\theta_0)]/[1 - F(\theta_0)]$ with a density $g^*(q) = \theta'(q) f(\theta)/[1 - F(\theta_0)]$, where $\theta = \theta(q)$. Elementary algebra gives $[1 - F(\theta)]/f(\theta) = \theta'(q)[1 - G^*(q)]/g^*(q)$. Equations (13) and (14) give

$$T'(q) = \gamma + \frac{1 - G^*(q)}{g^*(q)} \theta'(q) v_0(q). \tag{A.14}$$

Differentiating (14) with respect to q gives $T''(q) = \theta(q) v_0'(q) + \theta'(q) v_0(q)$, i.e. $\theta'(q) v_0(q) = T''(q) - \theta(q) - v_0'(q)$. Substituting the latter in (A.14) gives after some algebra

$$\theta(q) v_0'(q) = T''(q) - \frac{g^*(q)}{1 - G^*(q)} \left[T'(q) - \gamma \right].$$

Dividing the left-hand side by $\theta(q) v_0(q)$ and the right hand side by $T'(q)$ since $T'(q) = \theta v_0(q)$ by (14) gives

$$\frac{v_0'(q)}{v_0(q)} = \frac{T''(q)}{T'(q)} - \frac{g^*(q)}{1 - G^*(q)} \left[1 - \frac{\gamma}{T'(q)} \right].$$

Integrating both sides of the above equation from q_0 to q gives

$$\log \left(\frac{v_0(q)}{v_0(q_0)} \right) = \log \left(\frac{T'(q)}{T'(q_0)} \right) - \int_{q_0}^{q} \frac{g^*(x)}{1 - G^*(x)} \left(1 - \frac{\gamma}{T'(x)} \right) dx. \tag{A.15}$$

Taking the exponential gives

$$\frac{v_0(q)}{v_0(q_0)} = \frac{T'(q)}{T'(q_0)} \exp \left[- \int_{q_0}^{q} \frac{g^*(x)}{1 - G^*(x)} \left(1 - \frac{\gamma}{T'(x)} \right) dx \right]. \tag{A.16}$$

Condition (14) evaluated at q_0 gives $T'(q_0) = \theta_0 v_0(q_0)$. Multiplying the right-hand side of (A.16) by $T'(q_0)$ and the left-hand side by $\theta_0 v_0(q_0)$ gives

$$v_0(q) \quad = \quad \frac{T'(q)}{\theta_0} \exp \left[- \int_{q_0}^{q} \frac{g^*(x)}{1 - G^*(x)} \left(1 - \frac{\gamma}{T'(x)} \right) dx \right]$$

$$
\begin{aligned}
&= \frac{T'(q)}{\theta_0} \exp\left[-\int_{q_0}^{q} \frac{g^*(x)}{1-G^*(x)} dx\right] \exp\left[-\gamma \int_{q_0}^{q} \frac{-g^*(x)}{1-G^*(x)} \frac{1}{T'(x)} dx\right] \\
&= \frac{T'(q)}{\theta_0} [1-G^*(q)] \exp\left[-\gamma \left\{ \frac{\log(1-G^*(x))}{T'(x)} \Big|_{q_0}^{q} + \int_{q}^{q_0} \log(1-G^*(x)) \frac{T''(x)}{T'(x)} dx \right\}\right] \\
&= \frac{T'(q)}{\theta_0} \left[1-G^*(q)\right]^{1-\frac{\gamma}{T'(q)}} \exp\left\{-\gamma \int_{q_0}^{q} \log\left[1-G^*(x)\right] \frac{T''(x)}{T'(x)^2} dx\right\},
\end{aligned}
$$

where the third equality is obtained using integration by parts. Equation (17) follows from $\theta(q) = T'(q)/v_0(q)$ by (14). All the derivations in the above proof are reversible, so the proof of sufficiency is omitted. $\square$

Proof of Proposition 2: We consider two different structures $S = [v_0(\cdot), F(\cdot), C'(\cdot)]$ and $\tilde{S} = [\tilde{v}_0(\cdot), \tilde{F}(\cdot), \tilde{C}'(\cdot)]$, where $F(\cdot)$ is defined on $[\underline{\theta}, \overline{\theta}]$ with $\theta_0 = 1$ and $\tilde{F}(\cdot)$ is defined on $[\underline{\tilde{\theta}}, \overline{\tilde{\theta}}]$ and $\tilde{\theta}_0 = 1$. Both structures are assumed to be in $\mathcal{S}$ and to generate the same observables $[T(\cdot), q_0, F(\theta_0), G^*(\cdot), Q]$. By Lemma 3, we note $C'(Q) = \tilde{C}'(Q) = \gamma$. In view of Lemma 5, the structure $\tilde{S}$ has to satisfy

$$
\begin{aligned}
\tilde{v}_0(q) &= \frac{T'(q)}{\tilde{\theta}_0} \left[1-G^*(q)\right]^{1-\frac{-\gamma}{T'(q)}} \exp\left\{-\gamma \int_{q_0}^{q} \log\left[1-G^*(x)\right] \frac{T''(x)}{T'(x)^2} dx\right\} \quad \forall q \in (q_0, \overline{q}] \\
\tilde{\theta}(q) &= \tilde{\theta}_0 \left[1-G^*(q)\right]^{\frac{-\gamma}{T'(q)}-1} \exp\left\{\gamma \int_{q_0}^{q} \log\left[1-G^*(x)\right] \frac{T''(x)}{T'(x)^2} dx\right\} \quad \forall q \in (q_0, \overline{q}].
\end{aligned}
$$

By B2, $\theta_0 = \tilde{\theta}_0$ showing $\tilde{v}_0(\cdot) = v_0(\cdot)$ and $\theta(\cdot) = \tilde{\theta}(\cdot)$ on $(q_0, \overline{q}]$ and hence on $[q_0, \overline{q}]$ by continuity at q_0. Thus, $\tilde{F}^*(\cdot) = G^*(\tilde{q}(\cdot)) = G^*(q(\cdot)) = F^*(\cdot)$ on $[\theta_0, \overline{\theta}]$. Thus, $v_0(\cdot)$ and $F^*(\cdot)$ are uniquely determined on $[q_0, \overline{q}]$ and $[\theta_0, \overline{\theta}]$, respectively.$\square$

Proof of Proposition 3: *to be completed.*

References

Andrews, D. (1994): "Empirical Process in Econometrics," in in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics, Volume IV*, North Holland.

Armstrong, M. and **J. Vickers** (2001): "Competitive Price Discrimination," *Rand Journal of Economics*, 32, 579-605.

Billingsley, D. (1968): *Convergence of Probability Measures*, John Wiley and Sons.

Borenstein, S. (1991): "Selling Costs and Switching Costs: Explaining Retail Gasoline Margins," *Rand Journal of Economics*, 22, 354-369.

Borenstein, S. and **N. Rose** (1994): "Competition and Price Dispersion in the U.S. Airline Industry," *Journal of Political Economy*, 102, 653-683.

Busse, M. and **M. Rysman** (2005): "Competition and Price Discrimination in Yellow Pages Advertising," *Rand Journal of Economics*, 36, 378-390.

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.

Cohen, A. (2007): "Package Size and Price Discrimination in Paper Towels," *International Journal of Industrial Organization*, forthcoming.

Crawford, G.S. and **M. Shum** (2007): "Monopoly Quality Choice in Cable Television," *Journal of Law and Economics*, 50, 181-209.

Guerre, E., **I. Perrigne** and **Q. Vuong** (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525-574.

Ivaldi, M. and **D. Martimort** (1994): "Competition under Nonlinear Pricing," *Annales d'Economie et de Statistique*, 34, 71-114.

Katz, M. (1984): "Price Discrimination and Monopolistic Competition," *Econometrica*, 52, 1453-1472.

Kirk, D. (1970): *Optimal Control Theory: An Introduction*, Prentice-Hall, New Jersey.

Laffont, J.J. and **Q. Vuong** (1996): "Structural Analysis of Auction Data," *American Economic Review, Papers and Proceedings*, 86, 414-420.

Leslie, P. (2004): "Price Discrimination in Broadway Theatre," *Rand Journal of Economics*, 35, 520-541.

Lewis, T. and **D. Sappington** (1989): "Countervailing Incentives in Agency Problems," *Journal of Economic Theory*, 49, 294-313.

Li, T., **I. Perrigne** and **Q. Vuong** (2000): "Conditionally Independent Private Information in OCS Wildcat Auctions," *Journal of Econometrics*, 98, 129-161.

Lott, J. and **R. Roberts** (1991): "A Guide to the Pitfalls of Identifying Price Discrimination," *Economic Inquiry*, 29, 14-23.

Maggi, G. and **A. Rodriguez-Clare** (1995): "On Countervailing Incentives," *Journal of Economic Theory*, 66, 238-263.

Maskin, E. and **J. Riley** (1984): "Monopoly with Incomplete Information," *The Rand Journal of Economics*, 15, 171-196.

McManus, B. (2007): "Nonlinear Pricing in an Oligopoly Market: The Case of Specialty Coffee," *Rand Journal of Economics*, forthcoming.

Miravete, E. (2002): "Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans," *Review of Economic Studies*, 69, 943-971.

Miravete, E. and **L.H. Roller** (2004): "Competitive Nonlinear Pricing in Duopoly Equilibrium: The Early U.S. Cellular Telephone Industry," working paper , University of Pennsylvania.

Mussa, M. and **S. Rosen** (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301-317.

Oren, S., **S. Smith** and **R. Wilson** (1983): "Competitive Nonlinear Tariffs," *Journal of Economic Theory*, 29, 49-71.

Perrigne, I. and **Q. Vuong** (2007): "Nonparametric Identification of Incentive Regulation Models," working paper, Pennsylvania State University.

Pollard, D. (1984): *Convergence of Stochastic Processes*, Springer Verlag.

Rochet, J.C. and **L.A. Stole** (2003): "The Economics of Multidimensional Screening," in *Advances in Economic Theory: Eighth World Congress*, Dewatripont, M., L.P. Hansen, and S.J. Turnovsky, eds, Cambridge University Press.

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

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

Spence, A.M. (1977): "Nonlinear Prices and Welfare," *Journal of Public Economics*, 1, 1-18.

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

Stole, L.A (2007): "Price Discrimination in Competitive Environment," in *Handbook of Industrial Organization, Volume 3*, M. Armstrong and R. Porter, eds., North Holland.

Silverman, B.W. (1986): *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Stone, C.J. (1982): "Optimal Global Rates of Convergence for Nonparametric Estimators," *The Annals of Statistics*, 10, 1040-1053.

Tirole, J. (1988): *The Theory of Industrial Organization*, MIT Press.

van der Vaart, A. (1998): *Asymptotic Statistics*, Cambridge University Press.

Verboven, F. (2002): "Quality-based Price Discriminatin and Tax Incidence-the Market for Gasoline and Diesel Cars in Europe," *Rand Journal of Economics*, 33, 275-97.

Wilson, R. (1993): *Nonlinear Pricing*, Oxford University Press.

Table 1. Revenue Ranking by Industry Headings

Industry heading	Revenue	Percentage
Attorneys	\$500,347	8.71%
Dentists	\$190,667	3.32%
Insurance	\$187,496	3.27%
Storage household & commercial	\$139,425	2.43%
Hotels	\$131,108	2.28%
Restaurants	\$127,777	2.23%
Veterinarians	\$102,077	1.78%
Plumbing contractors	\$101,749	1.77%
Carpet & rug cleaners	\$100,086	1.74%
Automobile repair & service	\$84,327	1.47%

Table 2: Number of Purchases and Revenue by Sizes

Picas^2	Percent of a page	# Purchases	Revenue	Percent
Listing				
12	0.4%	2,347	$0	0.0%
18	0.6%	820	$126,050	2.2%
24	0.8%	109	$10,987	0.2%
27	0.9%	672	$200,416	3.5%
30	1.0%	273	$69,623	1.2%
36	1.2%	50	$15,557	0.3%
39	1.3%	287	$112,340	2.0%
42	1.4%	52	$18,652	0.3%
48	1.6%	8	$2,609	0.0%
51	1.7%	131	$60,899	1.1%
54	1.8%	7	$3,175	0.0%
60	2.0%	3	$1,740	0.0%
63	2.1%	20	$10,596	0.2%
66	2.2%	2	$1,109	0.0%
72	2.4%	1	$492	0.0%
75	2.5%	3	$1,804	0.0%
87	2.9%	1	$794	0.0%
96	3.2%	1	$706	0.0%
108	3.6%	1	$806	0.0%
Space Listing				
54	1.8%	979	$500,069	8.7%
72	2.4%	277	$218,005	3.8%
108	3.6%	137	$184,732	3.2%
144	4.8%	26	$47,513	0.8%
216	6.9%	10	$27,686	0.5%

Table 2: Number of Purchases and Revenue by Sizes (continued)

Picas^2	Percent of a page	# Purchases	Revenue	Percent
Display				
201	6.7%	143	$294,474	5.1%
213	7.1%	16	$34,048	0.6%
225	7.5%	12	$26,146	0.5%
235	7.8%	132	$335,261	5.8%
247	8.2%	12	$35,748	0.6%
259	8.6%	16	$44,860	0.8%
382	12.7%	123	$551,136	9.6%
394	13.1%	45	$207,640	3.6%
406	13.5%	2	$8,858	0.2%
418	13.9%	1	$6,175	0.1%
564	18.8%	57	$365,947	6.4%
575	19.1%	24	$174,767	3.0%
587	19.6%	7	$54,094	0.9%
762	25.4%	35	$323,424	5.6%
774	25.8%	4	$33,025	0.6%
786	25.2%	1	$10,950	0.2%
1137	37.9%	3	$44,102	0.8%
1149	38.3%	5	$68,750	1.2%
1512	50.4%	59	$947,088	16.5%
1524	50.8%	1	$10,194	0.2%
1536	51.2%	1	$18,208	0.3%
3047	101.6%	13	$410,318	7.1%
6066	202.2%	2	$120,737	2.1%
Total		6,931	$5,742,309	

Table 3: Some Examples of Estimated Firms' Utility for No Color

q	Adjusted q	# Obs.	$\hat{U} - U_0$
18	3.20	738	\$-18.057
54	15.71	935	\$352.643
201	71.85	87	\$866.143
236	87.67	66	\$1,139.743
382	159.47	39	\$1,780.943
564	255.26	23	\$2,542.743
1512	699.10	8	\$6,518.143
3047	1469.60	1	\$18,177.340

Figure 1

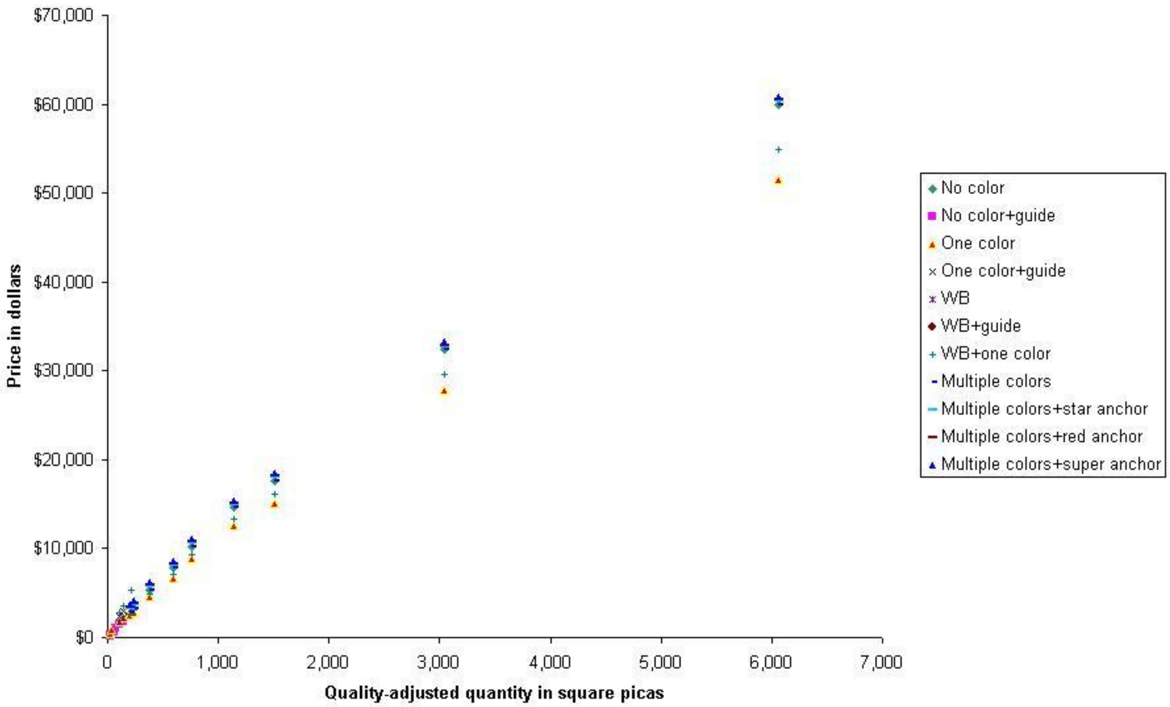
$70,000
$60,000
$50,000
$40,000
$30,000
$20,000
$10,000
$0
Price in dollars
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
Quality-adjusted quantity in square picas
No color
No color+guide
One color
One color+guide
WB
WB+guide
WB+one color
Multiple colors
Multiple colors+star anchor
Multiple colors+red anchor
Multiple colors+super anchor

Figure 2

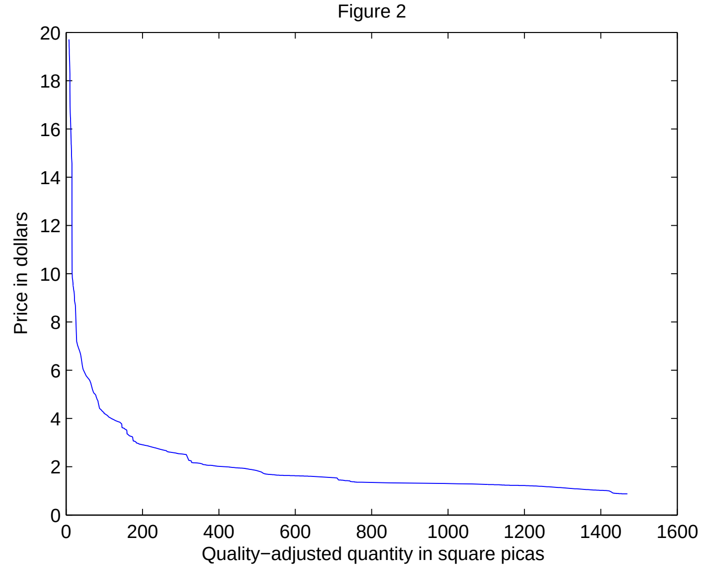

Figure 3

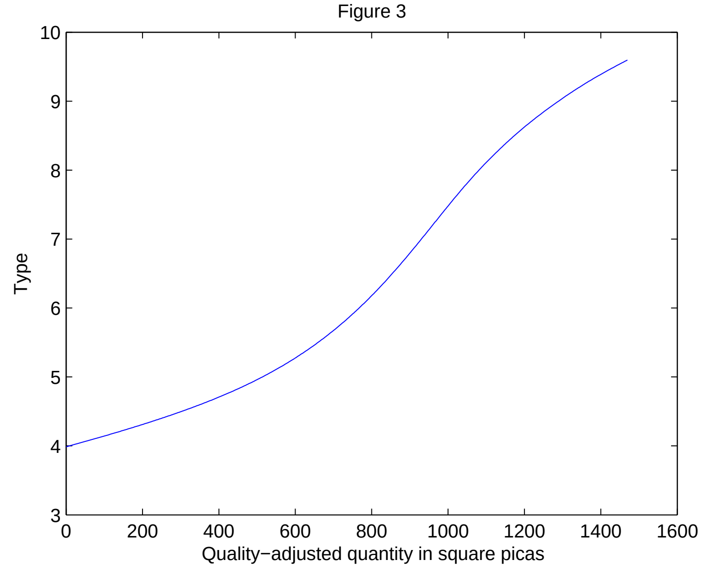

Figure 4

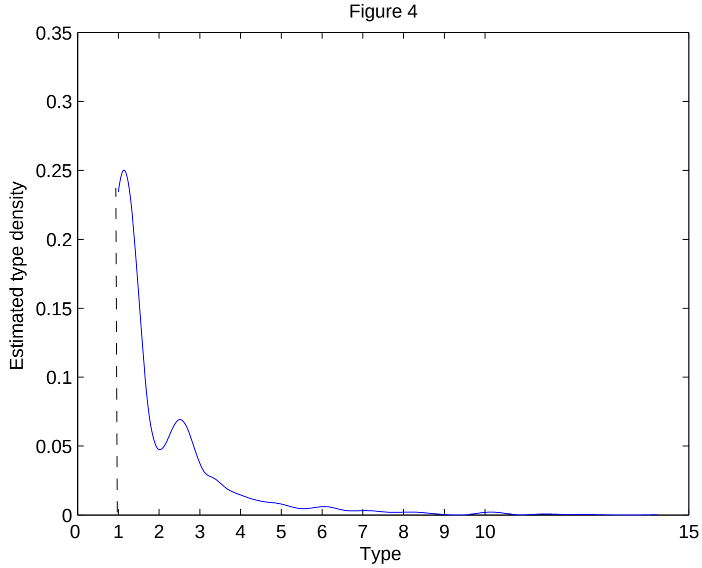

Figure 5