Nonlinear Pricing in Yellow Pages

Isabelle Perrigne
Pennsylvania State University

Quang Vuong
Pennsylvania State University

December 2010

* We are grateful to Yao Huang for her excellent research assitance and contributions to an earlier draft. We thank Dan Ackerberg, Don Andrews, Gaurab Aryal, Steve Berry, Philippe Fevrier, Jin Hahn, Phil Haile, Ken Hendricks, Guy Laroque, Pascal Lavergne, Vadim Marmer, Rosa Matzkin, Ariel Pakes, Daniel Quint, John Riley, Jean-Marc Robin and Marc Rysman for constructive comments. We also benefited from participants' comments at the Stanford Institute for Theoretical Economics 2008, the Cowles Summer Conference 2008, the Econometric Society European Meeting 2008, the Latin American Meeting of the Econometric Society 2008 and the IDEI Workshop on the Econometrics of Industrial Organization 2008 and at seminars at California Institute of Technology, Carnegie Mellon University, Massachussets Institute of Technology, New York University, Paris School of Economics, Simon Fraser University, University of Arizona, University of California at Davis, University of California at Los Angeles, University College London, University of Fortaleza, University of Maryland, University of Southern California and University of Wisconsin at Madison.

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

Abstract

This paper analyzes nonlinear pricing in yellow page advertising. First, we develop a model that incorporates some features of the industry such as a free minimal advertisement size offered to all businesses. The model structure is then defined by the distribution of businesses' types, the inverse demand function and the publisher's cost function. Under the assumption of a multiplicative inverse demand function, we show that the structure is nonparametrically identified up to the cost function, which is identified through its marginal cost at the total amount produced. Next, we propose a simple nonparametric procedure to estimate the type distribution and the inverse demand function. We establish the asymptotic properties of our two-step nonparametric estimator, whose first step converges at the parametric rate. The method is applied to analyze nonlinear pricing data in yellow page advertising. The empirical results show an important heterogeneity in businesses' tastes for advertising. Some counterfactuals assess the cost of asymmetric information and the gain or loss of nonlinear pricing in presence of asymmetric information relative other pricing rules.

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

Nonlinear Pricing in Yellow Pages

Isabelle Perrigne & Quang Vuong

1 Introduction

When firms face heterogeneous 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, celllular phone industry and advertising among others. See Wilson (1993) for additional examples. Economists analyze nonlinear pricing as an imperfect information model 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. Extensions to oligopoly competition and differentiated products include Oren, Smith and Wilson (1983), Ivaldi and Martimort (1994), Stole (1995), Rochet and Chone (1998), Armstrong and Vickers (2001), Rochet and Stole (2003) and Stole (2007).[1]

The economic importance of nonlinear pricing and the substantial theoretical developments have given rise to an increasing empirical literature. Early empirical studies by

[1]In these extensions, the optimal schedule becomes less tractable and closed form solutions can be obtained for some simple specifications only.

Lott and Roberts (1991), Shepard (1991), Clerides (2002) and Verboven (2002) focus on showing evidence of nonlinear pricing. Other studies by Borenstein (1991), Borenstein and Rose (1994) and Busse and Rysman (2005) document the impact of competition on patterns of nonlinear pricing. More recently, empirical studies evaluate the economic impact of nonlinear pricing on profits, consumer surplus and economic efficiency. Starting with Leslie (2004), a random utility discrete choice model for consumers' preferences is used to recover the consumers' taste distribution while considering an exogenous price schedule. See also McManus (2007) and Cohen (2008). A third trend in the empirical literature on nonlinear pricing endogeneizes the optimal price schedule to recover the demand and cost structure. See Ivaldi and Martimort (1994), Miravete (2002), Miravete and Roller (2004) and Crawford and Shum (2007) .

In this paper, we propose a structural analysis of nonlinear pricing in yellow page advertising. We first modify the Maskin and Riley (1984) monopoly model to incorporate an institutional feature, namely the publisher incorporates all the businesses by providing basic information such as their name, address and phone number at zero price.[2] This is equivalent to an optimal exclusion problem, i.e. an optimal threshold type below which businesses will be offered the standard listing at zero price. The inclusion of such businesses has, however, a cost to the publisher which should be taken into account. Moreover, in contrast to the theoretical and empirical literature which assumes a constant marginal cost, we consider a general cost function for the publisher. The previous empirical literature relies heavily on parametric specifications of the structure. In a nonlinear pricing model, the structure is defined by the business marginal payoff for buying advertising, the business type distribution and the publisher's cost function. The recent empirical industrial organization literature has documented that identification of models with incomplete information may or may not require particular functional forms.[3] In the spirit of the structural analysis of auction data, we investigate the nonparametric identification of the nonlinear pricing model from observables, which are individual advertising purchase

[2]This practice can be linked to the two-sided market/network effect of advertising. See Rysman (2004).

[3]See Laffont and Vuong (1996) and Athey and Haile (2007) for surveys on the nonparametric identification of auction models.

data and the tariffs offered by the publisher.

Our identification problem is reminiscent of Ekeland, Heckman and Neishem (2004) and Heckman, Matzkin and Neishem (2010) who study the nonparametric identification of hedonic price models. Both papers show that the marginal payoff function is not identified without further restrictions. In view of their results and Perrigne and Vuong (2011), we assume multiplicative separability of the marginal payoff function in the business type and the quantity of advertising purchased. Moreover, their identification results exploit some exogenous variables that are independent of the term of unobserved heterogeneity (the business type in our model). Because such variables are likely to be correlated with the business type, we adopt a different identifying strategy which exploits the first-order conditions of both sides of the market, i.e. optimal business choices and optimal publisher's tariff. Following Guerre, Perrigne and Vuong (2000), we 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 business type and its purchase. Because the structure contains multiple elements, the identification problem becomes more involved as in auctions with (say) risk averse bidders. See Perrigne and Vuong (2007) and Guerre, Perrigne and Vuong (2009). First, we show that the publisher's cost function is identified only at the margin for the total amount of advertising produced. Second, we show that the marginal payoff function and the business type distribution are identified above the truncation introduced by the optimal exclusion of firms. Below the threshold there is no variation in data that can allow identification of these functions. Third, relying on Matzkin (2003), we show that the type function aggregating the business observed and unobserved heterogeneity is also nonparametrically identified.

Based on this identification result, we propose natural and simple nonparametric estimators for the marginal cost at the total quantity produced, the marginal payoff function, the type distribution and the aggregation type function relying on empirical processes and kernel density estimation. We show the uniform consistency of each of our estimators and derive their respective consistency rates. In particular, the estimator of the marginal cost parameter is N consistent, while the estimator of the marginal payoff function is

$\sqrt{N}$-weakly convergent to a Gaussian process. In contrast to Guerre, Perrigne and Vuong (2000), the type density estimator achieves Stone (1982) optimal rate despite the unobservability of the types.

Next, we analyze a unique data set that we constructed from a phone directory in Pennsylvania and Yellow Page Association data. The data display a nonlinear pricing pattern as previously documented by Busse and Rysman (2005).[4] The resulting data set is unique because it contains (i) the full price schedule offered to businesses, (ii) the price and quantity chosen by each business and (iii) the whole population of businesses.[5] The price schedule provides several advertising categories and the purchase data show a large number of different price-quantity conmbinations chosen by businesses. This allows us to treat the price schedule as continuous as in the theoretical model. Our empirical results uncover an important heterogeneity in businesses' taste for advertising. The estimated marginal payoff function is decreasing as expected. Counterfactuals assess the cost of asymmetric information in terms of lost profit for the publisher relative to a complete information setting. We also simulate the gain or loss in publisher's revenue and firms' payoffs of nonlinear pricing realtive to other pricing schemes such a linear price and third-degree price discrimination based on the business headings.

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

[4]Busse and Rysman (2005) use the Yellow Page association data for the one-color category to show that larger competition is associated with a larger degree of curvature.

[5]Such nice data features do not exist in other sectors such as in 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.

2 Yellow Page Advertising Data

We collected data on yellow page advertising in 2006 for Centre County in Pennsylvania. The price schedule and the advertising options of the utility publisher (Verizon) were provided by the Yellow Page Association. The purchases by local businesses were constructed directly from the Verizon phone directory.[6] We collected a total of 6,823 advertisements over 1,193 industry headings. The advertisements bought by businesses generate a revenue of 6 million dollars. Table 1 displays the top 10 industry headings that represent 29% of the total revenue. Not surprisingly, we find professionals, household services and hospitality services.[7]

Other sources of advertising are available to businesses. First, the area is also covered by a non-utility publisher distributing 72,000 copies of its phone directory. This is the third of copies distributed by Verizon. Its directory is also much smaller than the Verizon one. This second publisher charges a significantly lower price and has a revenue of about 1 million dollars.[8] Second, an increasing competitor to printed yellow pages is the internet through search engines and internet yellow pages. Despite the predicted extinction of printed yellow pages, the industry remained strong until 2008. See the Newsletter by the Yellow Page Association (2008), which reports stable print usage and printed references for 2006 and 2007. Moreover, the printed phone directory was still used by about 87% of the U.S. population in 2007.[9] As a reasonable approximation, hereafter we assume that Verizon Communications acts as a monopoly in the area in 2006.

A notable feature of the offered advertising options is the socalled standard listing, which is free of charge to all businesses. The standard listing contains the name, address and phone number in the normal (smallest) font size.[10] In addition to the free standard

[6] After a careful check, the price schedule is strictly enforced.

[7] A similar ranking is obtained based on advertisement size.

[8] Aryal and Huang (2009) analyze the competition between the two publishers using a Stackelberg game and differentiated products within a parametric framework.

[9] Information on the generated revenue by Internet Yellow Pages in the State College area could not be found for 2006. A comparison for the top 10 industry headings in 2008 shows that only a small proportion bought advertising on the web.

[10] This practice became wide spread in the industry over the past ten years. In 2008, the twelve

listing, businesses can choose a large number of different advertising options. In particular, 4,671 businesses or 68.46% of all businesses decide to buy advertising.[11] Each advertising option is defined by the size, color and other special features.

In terms of size, the yellow page industry uses three categories, namely listing, space listing and display. Table 2 summarizes the various options chosen by businesses as well as the generated revenue for the publisher. The advertisement is measured in square picas, which is the unit commonly used in the publishing industry. One pica corresponds approximately to 1/6 inch. For instance, a standard listing is 12 square picas, and a full page is 3,020 square picas. The listing allows businesses to add extra lines and/or choose a larger font size to their standard listing. For instance, an extra line in normal font or 6 square picas is charged \$100.8. The listing category is chosen by 2,471 businesses, or 52.90% of those buying advertising, generating 10.77% of the total revenue with a size varying from 18 to 336 square picas. The space listing allocates a space within the column in addition to the listing. There are five sizes available within this category. Although the location of the advertisement on the page may contribute to its effectiveness, only its size matters in the publisher's price schedule in contrast to advertising in newspapers and on the web. The space listing category is chosen by 1,486 businesses, or 31.81% of those buying advertising, generating 18.82% of revenue with a size ranging from 54 to 612 square picas. The display provides a space beyond the column in which the listing is located. There are nine sizes available within this category, which can go up to two pages. The display category is chosen by 714 businesses, or 15.29% of those buying advertising, generating 70.41% of revenue with a size ranging from 174 to 6,147 square picas. Table 2 shows a striking inequality in terms of generated revenues. We also note the heterogeneity in demand. Conditional on buying, 91.61% of businesses buy a rather small advertising size, i.e. less than 10% of a page, while 1.65% of businesses buy more than 50% 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

publishers producing 82% of all printed directories in the U.S. proposed a standard listing free of charge to all business lines. This practice can be linked to the two-sided market/network effect of advertising.

[11] In contrast to urban phone books, our data do not show businesses carrying multiple names and/or name choices starting with an 'A' to be listed first in the heading as studied by McDevitt (2010).

plus one color, multiple colors including photos. These color options are not available for all categories. For instance, the multicolor option is offered for displays only. In our sample, only 0.7% of businesses choosing the listing category also choose some color. This number slightly increases to 4.5% in the space listing category. On the other hand, 54.9% of businesses choosing the display category have opted for some color. Color counts for an important difference in the price. For instance, one display page with no color costs $18,510 increasing up to $32,395 with multiple colors. Appendix C explains how the color options are introduced in our empirical analysis by adjusting the advertisement size.[12]

Regarding the special features, businesses can choose guide, anchor listing and trade marks.[13] Guide is offered to complement listing and space listing advertisements to indicate the business specialty. This option is chosen by 206 businesses, which can increase the price up to 30%. Anchor listing is provided for displays only. Under this option, a business can add a solid star to its listing to make the reference to the display advertisement more visible. This option is chosen by 105 businesses with a price ranging from $366 to $832. When a business carries a national brand, it can have the brand logo (trade mark) printed for a price ranging from $151 to $302.

To summarize, size, color and special features offer a large number of possible combinations for firms to choose from. We observe in the data 245 chosen different combinations leading to 245 different prices paid by firms. An interesting feature of the data is the curvature of the price schedule in terms of size for each given category. Specifically, the price paid by businesses per square pica decreases as the advertising size increases. For no color displays, the price per square pica varies from $9.41, $6.80 and $5.68 for the lowest size, half page and double page advertisements, respectively. This corresponds to a reduction of 66%. The same pattern is observed for other categories.[14] This corresponds to an important discount for quantity with a notable curvature for the tariff.

[12]Appendix C also justifies this procedure. Busse and Rysman (2005) consider only the price schedule for no color display advertising in their study.

[13]Regarding the cover pages and the coupons, these options concern only 22 businesses. The Yellow Page Association price schedule does not contain any information for these options. We have then excluded them from the empirical analysis.

[14]See also Table 2 though it mixes various categories for a given size.

3 The Model

We rely on Maskin and Riley (1984) nonlinear pricing model. The principal is the publishing company and the agents are the businesses buying advertising. The latter are characterized by a scalar taste parameter for advertising $\theta \in [\underline{\theta}, \overline{\theta}]$ with $0 \leq \underline{\theta} < \overline{\theta} < \infty$. This taste parameter is known to the business but unknown to the publisher. As indicated in Section 2, a norm in the industry is to propose a (standard) listing at zero price to all businesses. The problem relates to an optimal exclusion of consumers. The publisher chooses optimally a threshold level θ_0 below which $q_0 \geq \underline{q}$ is provided at zero price, where $\underline{q}$ is the minimal possible quantity of advertising. Such a quantity is exogenously determined by the printing technology.[15] Moreover, as in Riley (2011), we consider a general cost function instead of a constant marginal cost. Doing so allows us to examine to what extent the cost function can be identified from the observables in Section 4.

Each business has a payoff defined as

$$U(q, \theta) = \int_0^q v(x, \theta) dx - T(q) \qquad (1)$$

where q is the quantity of advertising purchased and $T(q)$ is the total payment for q units of advertising with $T(q) = 0$ if $q \leq q_0$. The function $v(q, \theta)$ expresses the θ business willingness to pay for the qth unit of advertising. It is also the marginal payoff for buying the qth unit of advertising or the inverse demand function for the business with type θ.[16] The type θ is distributed as $F(\cdot)$ with a continuous density $f(\cdot) > 0$ on $[\underline{\theta}, \overline{\theta}]$. The

[15] We show later that $q_0 = \underline{q}$.

[16] The marginal payoff for an additional unit of advertising can be the result of a game among businesses in the same subheading. For instance, with differentiated products and incomplete information, each business has a gross profit, which is a function of its type θ, its own price and advertising as well as its competitors' prices and advertising. Conditional on its own advertising and type, the Bayesian Nash equilibrium of the game in prices leads to a payoff, which is a function of its own advertising and type only. Such a payoff is $\int_0^q v(x, \theta) dx$. See Roberts and Samuelson (1988) and Gasmi, Laffont and Vuong (1992) for games in prices and advertising in complete information. Considering the full model with individual demand and cost functions in the econometric analysis would require data on prices and quantities of output for all businesses.

publisher does not know each business type but knows the distribution $F(\cdot)$.[17]

The following assumptions are made on $v(q,\theta)$.

Assumption A1: *The marginal payoff* $v(\cdot,\cdot)$ *is continuously differentiable on* $[0,+\infty)\times[\underline{\theta},\overline{\theta}]$, *and* $\forall q\geq 0, \forall\theta\in[\underline{\theta},\bar{\theta}]$

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

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

(iii) $v_2(q,\theta)>0$.[18]

Assumption A1-(i) says that the marginal payoff is strictly positive, while A1-(ii) says that the marginal payoff is decreasing in the quantity purchased. Assumption A1-(iii) says that businesses with a larger θ enjoy a larger marginal payoff 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 $q(\theta)>q_0$ for $\theta\in(\theta_0,\overline{\theta}]$. The payment $T(\cdot)$ is defined on $[0,q(\overline{\theta})]$ with $T(\cdot)=0$ on $[0,q_0]$. 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 show that with additional assumptions the optimal $q(\cdot)$ is strictly increasing.[19] 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

$$\Pi \;=\; \int_{\theta_0}^{\overline{\theta}} T(q(\theta))f(\theta)d\theta - C\left[q_0F(\theta_0)+\int_{\theta_0}^{\overline{\theta}} q(\theta)f(\theta)d\theta\right],$$

where the first term is the revenue collected from all businesses buying advertising and the second term expresses the cost for producing the total advertising quantity with a cost function $C(\cdot)$. Because businesses with types below θ_0 do not pay for their advertising quantity q_0, they do not show up in the publisher's revenue. On the other hand, this

[17]If the publisher was able to discriminate on some business characteristics such as the industry heading, such variables will show up in $F(\cdot)$ as conditioning variables and/or in $v(q,\theta)$ as additional variables.

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

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

production has a cost that the publisher needs to take into account. This explains the argument of the cost function in two parts: (i) $q_0F(\theta_0)$ represents the total quantity provided to businesses choosing q_0 and (ii) $\int_{\theta_0}^{\overline{\theta}} q(\theta)f(\theta)d\theta$ is the total quantity provided to other businesses. The cost function is assumed to be strictly increasing.

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

The publisher's profit is maximized subject to the individual rationality (IR) and the incentive compatibility (IC) constraints of the businesses. The latter is derived from the business optimization problem. For the IR constraint, consider first a business with type $\theta > \theta_0$. It must prefer to buy $q(\theta)$ rather than q_0, i.e.

$$U(q(\theta), \theta) \geq \int_0^{q_0} v(x, \theta)dx \equiv U_0(\theta) \quad \forall \theta \in (\theta_0, \overline{\theta}]. \tag{2}$$

For a business with a type $\theta \leq \theta_0$, it receives q_0 for free, which provides the payoff $U_0(\theta)$ satisfying trivially its individual rationality constraint. We remark that despite having the reservation payoff $U_0(\theta)$ depending on the business type, (2) does not lead to countervailing incentives as studied by Lewis and Sappington (1989) and Maggi and Rodriguez-Clare (1995). This is so because (2) is equivalent to $\int_{q_0}^{q} v(x, \theta)dx - T(q) \geq 0$, which is strictly increasing in θ by A1-(iii) for any given q.

For the IC constraint, we consider four cases. First, a business 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 $U(\tilde{\theta}, \theta) \equiv U(q(\tilde{\theta}), \theta) \quad \forall \theta, \tilde{\theta} \in (\theta_0, \overline{\theta}]$. This 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 $U(\tilde{\theta}, \theta) = \int_0^{q(\tilde{\theta})} v(x, \theta)dx - T(q(\tilde{\theta}))$. Thus $U_1(\theta, \theta) = [v(q(\theta), \theta) - T'(q(\theta))]$ $q'(\theta)$. 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 business 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 $U_0(\theta)$. Thus, the IC constraint is $U(q(\theta), \theta) \geq U_0(\theta)$, which is the IR constraint (2). Third, a business 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 $U_0(\theta)$. Thus, this constraint is trivially verified. Fourth, a business 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 $U_0(\theta) \geq U(\tilde{\theta}, \theta)$ leading to $T(q(\tilde{\theta})) \geq \int_{q_0}^{q(\tilde{\theta})} v(x, \theta)dx$ for $\theta \leq \theta_0$ and $\tilde{\theta} > \theta_0$. But, because $v(x, \cdot)$ is increasing in θ by A1-(iii), the latter is equivalent to $T(q(\tilde{\theta})) \geq \int_{q_0}^{q(\tilde{\theta})} v(x, \theta_0)dx$ for $\tilde{\theta} > \theta_0$, which is true.[20]

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 A1, (2) and $q'(\cdot) > 0$ on $(\theta_0, \overline{\theta}]$, the local FOC (4) is sufficient for all the IC constraints to hold globally.*

We can now solve the publisher's optimization problem:

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

subject to the IR constraint (2), the IC constraint (4), where $q(\cdot) = q_0 \geq \underline{q}$ for $\theta \leq \theta_0$ and $q(\theta) > q_0$ for $\theta > \theta_0$ with $q(\cdot)$ strictly increasing on $(\theta_0, \overline{\theta}]$. Because q_0 affects Π only through $q_0 F(\theta_0)$ in the cost function, q_0 needs to be set at the minimum $\underline{q}$. As usual (see Tirole (1988, Chapter 3)), we eliminate $T(\cdot)$ by $\tilde{U}(\cdot)$ in the optimization problem, where $\tilde{U}(\theta) = \int_{\underline{q}}^{q(\theta)} v(x, \theta)dx - T(q(\theta))$ for $\theta \in (\theta_0, \overline{\theta}]$. Taking the derivative with respect to θ gives $\tilde{U}'(\theta) = [v(q(\theta), \theta) - T'(q(\theta)]\, q'(\theta) + \int_{\underline{q}}^{q(\theta)} v_2(x, \theta)dx = \int_{\underline{q}}^{q(\theta)} v_2(x, \theta)dx$, where the second equality uses (4). Let $U_+ \equiv \lim_{\theta \downarrow \theta_0} \tilde{U}(\theta)$. Integrating the previous equation gives $\tilde{U}(\theta) = \int_{\theta_0}^{\theta} \left\{\int_{\underline{q}}^{q(u)} v_2(x, u)dx\right\} du + U_+$. Using the definition of $\tilde{U}(\theta)$, we obtain

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

[20] From (4), $T'(q(\tilde{\theta})) = v(q(\tilde{\theta}), \tilde{\theta}) \geq v(q(\tilde{\theta}), \theta_0)$ by A1-(iii) again. Thus, $T'(q) \geq v(q, \theta_0)$ for $q \geq q_0$. Integrating gives $\int_{q_0}^{q} T'(x)dx \geq \int_{q_0}^{q} v(x, \theta_0)dx$. Letting $T_+ = \lim q \downarrow q_0 T(q)$, it gives $T(q) - T_+ \geq \int_{q_0}^{q} v(x, \theta_0)dx$ for $q \geq q_0$. Evaluating the latter at $q(\tilde{\theta})$ gives $T(q(\tilde{\theta})) \geq \int_{q_0}^{q(\tilde{\theta})} v(x, \theta_0)dx + T_+$. Hence, $T(q(\tilde{\theta})) \geq \int_{q_0}^{q(\tilde{\theta})} v(x, \theta_0)dx$ for $\tilde{\theta} > \theta_0$ if $T_+ \geq 0$. We show later that $T_+ = 0$.

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

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

The second term becomes $\int_{\theta_0}^{\overline{\theta}} \left[\int_{\underline{q}}^{q(\theta)} v_2(x,\theta)dx\right] d\theta - \int_{\theta_0}^{\overline{\theta}} \left\{\left[\int_{\underline{q}}^{q(\theta)} v_2(x,\theta)dx\right] F(\theta)\right\} d\theta$ by integration by parts. After rearranging terms and noting that $\tilde{U}(\cdot)$ appears through U_+ only, the firm's problem becomes

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

Maximization of Π with respect to U_+ gives trivially $U_+ = 0$. The optimal control problem is nonstandard because θ_0 appears at the boundary of the integral. It 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*

$$v(q,\theta) = C'(Q) + \frac{1-F(\theta)}{f(\theta)} v_2(q,\theta) \quad \forall \theta \in (\theta_0, \overline{\theta}] \qquad (8)$$

$$\lim_{\theta \downarrow \theta_0} q(\theta) = \underline{q} \qquad (9)$$

$$T'(q) = v(q,\theta) \quad \forall \theta \in (\theta_0, \overline{\theta}] \qquad (10)$$

$$\lim_{q \downarrow \underline{q}} T(q) = 0, \qquad (11)$$

where $Q \equiv \underline{q}F(\theta_0) + \int_{\theta_0}^{\overline{\theta}} q(u)f(u)du$ 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 a constant c. The marginal payoff for each type then equals the marginal cost plus a nonnegative distortion term due to incomplete information. Hence, by A1-(iii) all businesses buy less than the efficient (first-best) quantity of advertising except for the business with type $\overline{\theta}$ for which there is no distortion. When the cost function is nonlinear, the publisher considers the marginal cost of the last unit of the total quantity produced Q. For the highest type, its marginal payoff 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 payoff for that type. Equations (9) and (11) imply the continuity of $q(\cdot)$ and $T(\cdot)$ at θ_0 and $\underline{q}$, respectively.

The next lemma shows that the optimal $q(\cdot)$ is strictly increasing as desired under the following assumption.

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

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

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

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

These assumptions are standard. Assumption A4-(i) says that the demand elasticity is nonincreasing in type. Maskin and Riley (1984) show that a large classes of preferences satisfy it.[21] Assumption A4-(ii) is more difficult to interpret. Lemma 2 shows, however, that A4-(ii) is implied by A4-(iii) and A4-(iv). Assumption A4-(iii) says that the increase in demand price is diminishing as θ increases, while A4-(iv) states that the hazard rate of the type distribution does not decline too rapidly as θ increases. A large class of

[21] In particular, demand functions of the forms $v(q,\theta) = \theta - \beta(\theta)q$ and $v(q,\theta) = \alpha(\theta)q^{-1/\eta}$, where $\eta > 0$, satisfy A4-(i) under suitable assumptions on $\beta(\theta)$ and $\alpha(\theta)$.

distribution functions satisfy the latter property.[22]

Lemma 2: *Under A1, A2, and A4-(i,ii) or under 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* $[\underline{q}, \overline{q}]$ *with* $T'(\cdot) > C'(Q)$ *on* $(\underline{q}, \overline{q})$*,* $T'(\underline{q}) = v(\underline{q}, \theta_0)$ *and* $T'(\overline{q}) = C'(Q)$ *where* $\overline{q} \equiv q(\overline{\theta})$*.*

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

4 Nonparametric Identification and Estimation

4.1 Nonparametric Identification

We define the game structure and the observables. Following Section 3, the model primitives are $[v(\cdot, \cdot), F(\cdot), C(\cdot)]$, which are the marginal payoff, the businesses' type distribution and the publisher's cost function. The data provide information on the price-advertising schedule, the minimum quantity at zero price, the proportion of businesses choosing this minimum quantity, the businesses advertising purchases and the total amount of advertising produced. Using our previous notations, the observables are $[T(\cdot), \underline{q}, F(\theta_0), G^*(\cdot), Q]$.[23] The function $G^*(\cdot)$ denotes the truncated distribution of businesses' purchases, i.e. $G^*(\cdot) = \Pr(q \leq \cdot)/\Pr(q > \underline{q})$. The data also provide some exogenous characteristics Z of businesses such as their industry heading. A natural question is how to introduce this observed heterogeneity in the model of Section 3. This is not straightforward here because the observed schedule $T(\cdot)$ must be independent of Z as the publisher cannot discriminate businesses based on their characteristics Z. Our proposal is

[22] These assumptions are generally sufficient for the second-order conditions of the optimization problem. As such, they might be weakened. In a different context, Perrigne and Vuong (2011) derive the sufficient and necessary conditions for the second-order conditions to hold in terms of observables. Such an exercise which is related to testing the model validity is left for future research.

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

to view θ as a scalar aggregation of the business' observed and unobserved heterogeneity. Let ϵ denote the latter. Formally, we make the following assumption.

Assumption B1: *The business' type satisfies* $\theta = r(Z, \epsilon)$ *with* ϵ *distributed as* $F_{\epsilon|Z}(\cdot|\cdot)$ *given* Z.

In particular, this allows for dependence between the business' type θ and characteristics Z. We remark that Z and ϵ need not be independent. Under B1, the marginal payoff $v(q, \theta)$ becomes $v^{\dagger}(q, Z, \epsilon) \equiv v(q, r(Z, \epsilon))$. In view of (1), the business' optimization problem remains the same and the optimal choice q is a function of θ only. Hence, the cutoff type θ_0 is independent of Z by (9). The publisher's optimization problem is still given by (5), while Proposition 1 and Lemma 2 hold.[24]

Hereafter, 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)–(11). The model primitives become $[v(\cdot,\cdot), r(\cdot,\cdot), F_{\epsilon|Z}(\cdot|\cdot), C(\cdot)]$. Identification investigates whether the primitives can be uniquely recovered from the observables.

Our identification problem is reminiscent of Ekeland, Heckman and Neishem (2004) and Heckman, Matzkin and Neishem (2010) who study the nonparametric identification of hedonic models. In these papers, consumers satisfy a first-order condition similar to (10), where the marginal payoff equals the marginal price, which is nonlinear in quantity. Both papers show that the marginal payoff function is nonidentified without any further restrictions. Ekeland, Heckman and Neishem (2004) establish identification of the marginal payoff and the distribution of unobserved heteregeneity up to location and scale under an additive separable marginal payoff by exploiting variations in some continuous exogenous variables that are independent of the term of unobserved heterogeneity. This result is obtained without the need to consider the firms' optimization problem. In addi-

[24] An alternative would be to consider a model with $v(q, \theta, Z)$ and $F(\theta|Z)$, while still imposing a price schedule independent of Z. This will lead to a publisher's optimization problem different from (5) in two aspects. First, the expected profit will have an additional integration with respect to z in both the revenue and the total amount produced. Second, the control function $q(\cdot,\cdot)$ will have two arguments (θ, z). The derivation of this model is left for future research.

tion, they show that considering the equilibrium conditions combining the consumers and the firms does not provide additional identifying information for either side of the market. Heckman, Matzkin and Nesheim (2010) consider alternative identifying assumptions on the functional form of the marginal payoff with exogenous variables possibly for single and multimarket data. In view of their results and Perrigne and Vuong (2011) in the context of a procurement model with both adverse selection and moral hazard, we assume that the marginal payoff $v(q,\theta)$ is multiplicatively separable in q and θ.

Assumption B2: *The businesses' marginal payoff 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* $q \geq \underline{q}$ *and for all* $\theta \in [\underline{\theta}, \overline{\theta}] \subset (0, +\infty)$.[25]

We interpret $v_0(q)$ as the base marginal payoff. It can be easily seen that the assumptions on the marginal payoff A1, A3 and A4-(i,iii) are satisfied. The necessary conditions (8) and (10) then become

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

where $\theta = q^{-1}(q)$ and $\overline{q} = q(\overline{\theta})$ since $q(\cdot)$ is strictly increasing by Lemma 2. The one-to-one mapping between the unobserved type θ and the observed advertising quantity q is the key of our identification result. Following B2, the model structure becomes $[v_0(\cdot), r(\cdot,\cdot), F_{\epsilon|Z}(\cdot|\cdot), C(\cdot)]$. To establish identification, we proceed in two steps. First, we show that the model structure $[v_0(\cdot), F(\cdot), C(\cdot)]$ is identified. Second, we show that $r(\cdot,\cdot)$ and $F_{\epsilon|Z}(\cdot|\cdot)$ are identified. 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 B2, A2 and A4-(iv).

[25] We could consider a multiplicative function of the form $\psi(\theta)v_0(q)$. The function $\psi(\cdot)$ satisfies $\psi(\cdot) > 0$, $\psi'(\cdot) > 0$ 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 remains nonidentified. Also, under this multiplicative separability, it can be shown following Tirole (1988, p.156) that the price schedule is strictly concave in q when A4-(iv) is strengthened to a hazard rate $\rho(\theta)$ increasing in θ.

Our first identification result concerns the cost function, which is not identified except for the marginal cost at the total amount produced. This result is not surprising since the model involves the cost function only through the marginal cost at this amount. 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* $C'(Q) = T'(\overline{q})$.

This follows from Lemma 2. Since we observe both the price schedule $T(\cdot)$ and $\overline{q}$, we can recover $C'(Q)$, i.e. the latter is identified.[26] Identification of the cost function could be improved by using data from different markets providing variations in Q.

The structural elements left to identify are $v_0(\cdot)$ and $F(\cdot)$. In view of Ekeland, Heckman and Nesheim (2004) and Heckman, Matzkin and Nesheim (2010), a natural question is whether one can identify $v_0(\cdot)$ and $F(\cdot)$ by using the businesses' first-order condition (14) only. Their identication argument relies on the availability of some continuous exogenous variables that are independent of the term of unobserved heterogeneity.[27] In our case, the only variables that are available are discrete. Morever, these variables are likely to be correlated with the business' taste θ for advertising. Consequently, our identification argument uses instead both sides of the market.

We first show that a scale normalization is necessary. This normalization is needed because both the business' type θ and the base marginal payoff $v_0(\cdot)$ are unknown. 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* $[T(\cdot), \underline{q}, F(\theta_0), G^*(\cdot), Q]$, *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

[26] Miravete and Roller (2004) note a similar result. 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.

[27] Their separability assumption is also different from ours as they assume $v(q, z, \epsilon) = \epsilon v_0(q, z)$, while we have $v(q, z, \epsilon) = r(z, \epsilon) v_0(q)$.

5, in which the business' marginal payoff $v_0(\cdot)$ and its unobserved type θ are expressed as functions of the quantity purchased q and other observables $[T(\cdot), \underline{q}, F(\theta_0), G^*(q), Q]$. 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 (13) and (14) are equivalent to*

$$v_0(q) = \frac{T'(q)}{\theta_0 \xi(q)} \text{ and } \theta(q) = \theta_0 \xi(q), \tag{15}$$

for all $q \in (\underline{q}, \overline{q}]$, where

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

with $\xi(\underline{q}) = 1$ and $\xi(\overline{q}) = \lim_{q \uparrow \overline{q}} \xi(q) = \overline{\theta}/\theta_0$.

The proof of Lemma 5 exploits the one-to-one mapping between the advertising quantity q and the business' type $\theta \in (\theta_0, \overline{\theta}]$ obtained from Lemma 2. For each $q \in (\underline{q}, \overline{q}]$, we observe the truncated marginal distribution $G^*(q) = \Pr(\tilde{q} \leq q | \tilde{q} > \underline{q}) = \Pr(\tilde{\theta} \leq \theta(q) | \tilde{\theta} > \theta(\underline{q})) = [F(\theta) - F(\theta_0)]/[1 - F(\theta_0)]$ 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 payoff $v_0(\cdot)$ and the unobserved type θ in terms of observables, which are the corresponding quantity q, the truncated quantity distribution $G^*(\cdot)$, its density $g^*(\cdot)$, the price schedule $T(\cdot)$ as well as $\gamma = C'(Q)$, which is identified by Lemma 3. There is a parallel with auction models. Guerre, Perrigne and Vuong (2000) use the one-to-one 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 business' type θ can be interpreted as the unknown bidder's private value, while the business' 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 to recover in addition to the distribution of businesses' type $F(\cdot)$, namely the base marginal payoff $v_0(\cdot)$. To this end, we exploit the relationship between the shape of the price schedule $T(\cdot)$ and the distribution of the unobserved businesses' type $F(\cdot)$ as made clear in the proof of Lemma 5. We also note that the density $g^*(\cdot)$ does not appear in $\xi(\cdot)$. This feature will be important in deriving the consistency rate of our estimator of $\xi(\cdot)$ in Section 4.2.

In view of Lemma 5, a natural normalization is $\theta_0 = 1$. The base marginal payoff $v_0(\cdot)$ can be uniquely recovered on $(\underline{q}, \overline{q}]$ from the observables $T'(\cdot)$, $T''(\cdot)$, $G^*(\cdot)$ and $\gamma = T'(\overline{q})$ and hence at $\underline{q}$ 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 B3: *We normalize $\theta_0 = 1$.*

Under such a normalization, $v_0(q)$ is the marginal utility function for the cutoff type.

Proposition 2: *Let $[v_0(\cdot), F(\cdot), C(\cdot)] \in \mathcal{S}$. Under B2-B3, the marginal payoff $v_0(\cdot)$ and the truncated business' type distribution $F^*(\cdot)$ are identified on $[\underline{q}, \overline{q}]$ and $[\theta_0, \overline{\theta}]$, respectively. In particular, $v_0(\underline{q}) = T'(\underline{q})$ and $\overline{\theta} = \lim_{q \uparrow \overline{q}} \xi(q)$.*

We note that $F(\cdot)$ can be recovered from $F^*(\cdot)$ on $[\theta_0, \overline{\theta}]$ using $F(\cdot) = F(\theta_0) + (1 - F(\theta_0))F^*(\cdot)$ since $F(\theta_0)$ is observed as the proportion of businesses receiving $\underline{q}$ at zero price. On the other hand, $v_0(\cdot)$ and $F(\cdot)$ are not identified on $[0, \underline{q})$ and $[\underline{\theta}, \theta_0)$, respectively. Intuitively, the quantity and price data do not provide any variation to identify these functions in those ranges as the minimum observed quantity is $\underline{q}$. As in auctions, a binding reserve price does not allow to identify the distribution of bidders' private values below the reserve price. See Guerre, Perrigne and Vuong (2000).

Alternative normalizations can be entertained. The normalization $\overline{\theta} = 1$ allows to identify $v_0(\cdot)$ and $F^*(\cdot)$ on $[\underline{q}, \overline{q}]$ and $[\theta_0, \overline{\theta}]$, respectively. In this case, (15) and (16) give

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

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

We now turn to the second step, which addresses the identification of $r(\cdot,\cdot)$ and $F_{\epsilon|Z}(\cdot|\cdot)$ appearing in B1. Note that the first step identifies not only the truncated distribution $F^*(\cdot)$ of θ by Proposition 2 but also recovers the type θ for each business from its purchase $q > \underline{q}$ by (15) and (16). On the other hand, under B1, $\theta = r(z, \epsilon)$. Since θ is known only if $\theta > \theta_0$, this introduces a censoring, i.e. $\tilde{\theta} = \theta$ if $r(z, \epsilon) > \theta_0$ and $\tilde{\theta} = \theta_0$ otherwise, where $\theta_0 = 1$ by B3. Identification of $r(\cdot,\cdot)$ and $F_{\epsilon|Z}(\cdot|\cdot)$ extends Matzkin (2003) to a censored nonseparable model under the following assumption. Consider the partition $Z = (Z_1, Z_2)$. Let $\mathcal{S}_{Z_1}$ and $\mathcal{S}_{Z_2|z_1}$ be the supports of Z_1 and Z_2 given $Z_1 = z_1$, respectively. Let also $\mathcal{S}_{\epsilon,Z_2|z_1}$ be the support of (ϵ, Z_2) given $Z_1 = z_1$.

Assumption B4: *We have*

(i) The unobserved heterogeneity ϵ is independent of Z_1 given Z_2,

(ii) There exists a known value $z_1^o \in \mathcal{S}_{Z_1}$ such that $\mathcal{S}_{Z_2|z_1^o} = \mathcal{S}_{Z_2}$ and $r(z_1^o, z_2, \epsilon) = \epsilon$ for all $(\epsilon, z_2) \in \mathcal{S}_{\epsilon,Z_2|z_1^o}$.

(iii) For every $(z_1, z_2) \in \mathcal{S}_{Z_1,Z_2}$, the functions $r(z_1, z_2, \cdot)$ and $F_{\epsilon|Z_1,Z_2}(\cdot|z_1, z_2)$ are both strictly increasing on $\mathcal{S}_{\epsilon|z_1,z_2}$.

Assumption B4-(i) is weaker than independence between ϵ and Z. Assumption B4-(ii) follows Matzkin (2003) first normalization. In particular the first part of (ii) is a full support requirement on z_1^o. Assumption B4-(iii) underlies the use of quantiles for identification. For every $z \in \mathcal{S}_Z$, let $e(z)$ satisfy $r(z, e(z)) = \theta_0$.

Proposition 3: *Under B1, B3 and B4, the function $r(\cdot,\cdot)$ and the distribution $F_{\epsilon|Z}(\cdot|\cdot)$ are identified on $\{(z_1, z_2, e) \in \mathcal{S}_{Z_1,Z_2,\epsilon} : e \geq \max[e(z_1, z_2), \theta_0]\}$ and $\{(z_1, z_2, e) \in \mathcal{S}_{Z_1,Z_2,\epsilon} : e \geq \theta_0\}$, respectively, as*

$$
\begin{aligned}
r(z_1, z_2, e) &= F^{*-1}_{\theta|Z_1,Z_2}[F^*_{\epsilon|Z_1,Z_2}(e|z_1, z_2)|z_1, z_2] && (17)\\
F_{\epsilon|Z_1,Z_2}(e|z_1, z_2) &= \Pr[\tilde{\theta} = \theta_0|Z_1 = z_1^o, Z_2 = z_2] \\
&\quad +(1 - \Pr[\tilde{\theta} = \theta_0|Z_1 = z_1^o, Z_2 = z_2])F^*_{\theta|Z_1,Z_2}(e|z_1^o, z_2), && (18)
\end{aligned}
$$

where $\max[e(z_1, z_2), \theta_0]$ *is identified as a function of* $(z_1, z_2) \in \mathcal{S}_{Z_1,Z_2}$ *and*

$$
F^*_{\epsilon|Z_1,Z_2}(e|z_1, z_2) = 1 - \frac{1 - \Pr[\tilde{\theta} = \theta_0|Z_1 = z_1^o, Z_2 = z_2]}{1 - \Pr[\tilde{\theta} = \theta_0|Z_1 = z_1, Z_2 = z_2]}[1 - F^*_{\theta|Z_1,Z_2}(e|z_1^o, z_2)]. \quad (19)
$$

Proposition 3 establishes the identification of $r(\cdot,\cdot)$ and $F_{\epsilon|Z}(\cdot|\cdot)$ on some suitable subsets of their supports. We remark that $\Pr[\tilde{\theta} = \theta_0|Z_1 = z_1, Z_2 = z_2]$ is the probability that businesses with characteristics (z_1, z_2) choose not to buy advertising. Thus, this probability is observed from the data. Moreover, the distribution $F^*_{\theta|Z_1,Z_2}(\cdot|z_1, z_2)$ is the businesses' type distribution conditional on advertising purchase and characteristics (z_1, z_2). Since θ is recovered for all businesses buying advertising, this distribution is also available from the data. Relative to the non-censored case in Matzkin (2003), the expression for $r(\cdot,\cdot)$ differs in three aspects. First, our result involves the truncated distribution of θ given Z. Second, there is an adjustment due to censoring as $\Pr[\tilde{\theta} = \theta_0|Z = (z_1^o, z_2)] \neq \Pr[\tilde{\theta} = \theta_0|Z = (z_1, z_2)]$. Third, $r(z, e)$ is identified only for $e \geq \max[e(z), \theta_0]$. Lastly, the term θ_0 in the truncated supports of $r(\cdot,\cdot)$ and $F_{\epsilon|Z}(\cdot|\cdot)$ arises because $\theta_0 = r(z_1^o, z_2, e(z_1^o, z_2)) = e(z_1^o, z_2)$ by B4-(ii).

4.2 Nonparametric Estimation

Given B3, (15) provides the marginal payoff $v_0(\cdot)$ and the business type $\theta(\cdot)$ as functions of $T'(\cdot)$ and $\xi(\cdot)$. The latter depends on $T'(\cdot)$, $T''(\cdot)$, γ and the truncated quantity distribution $G^*(\cdot)$. The functions $T'(\cdot)$ and $T''(\cdot)$ come from the price schedule data implying that only $\gamma = T'(\overline{q})$ and $G^*(\cdot)$ need to be estimated. We propose a two-step estimation procedure. In the first step, we estimate $\xi(\cdot)$ nonparametrically using (16). This allows us to obtain an estimate for the marginal payoff $v_0(\cdot)$ and to construct a sample of pseudo types from (15). In the second step, this pseudo sample is used to estimate nonparametrically the truncated business' type density, the function $r(\cdot,\cdot)$ as well as the conditional distribution $F_{\epsilon|Z}(\cdot|\cdot)$. Because we use data from a single market, our estimation procedure is not performed conditionally upon some variables X capturing market heterogeneity such as the median income and population size. If data from several yellow page directories were available, we could extend our estimation procedure by conditioning on X and hence estimating $G^*(\cdot|\cdot)$ in the first step and $F(\cdot|\cdot)$ in the second step.

We denote by N^* the number of firms purchasing advertising space strictly larger than $\underline{q}$, while $q_i, i = 1, 2, ..., N^*$ denotes the quantity purchased by each of those firms. Following (16) we estimate $\xi(\cdot)$ by

$$\hat{\xi}(q) = \begin{cases} \left[1 - \hat{G}^*(q)\right]^{\frac{\hat{\gamma}}{T'(q)}-1} \exp\left\{\hat{\gamma} \int_{\underline{q}}^{q} \frac{T''(x)}{T'(x)^2} \log\left[1 - \hat{G}^*(x)\right] dx\right\} & \text{if } q \in [\underline{q}, q_{\max}) \\ \lim_{q \uparrow q_{\max}} \hat{\xi}(q) & \text{if } q \in [q_{\max}, \overline{q}], \end{cases} \tag{20}$$

where $q_{\max} = \max_{i=1,\ldots,N^*} q_i$, $\hat{\gamma} = T'(q_{\max})$, $\hat{G}^*(\cdot)$ is the empirical distribution

$$\hat{G}^*(q) = \frac{1}{N^*} \sum_{i=1}^{N^*} 1\!\!I(q_i \leq q), \text{ for } q \in [\underline{q}, \overline{q}], \tag{21}$$

and $1\!\!I(\cdot)$ is the indicator function. In particular, $q_{\max} \leq \overline{q}$.

As a matter of fact, for $q \in [\underline{q}, q_{\max}]$, $\hat{\xi}(q)$ is straightforward to compute. Specifically, because the empirical distribution $\hat{G}^*(\cdot)$ is a step function with steps at $q^1 < q^2 < \ldots < q^J$ in $(\underline{q}, \overline{q}]$, the integral in (20) can be written as the finite sum of integrals on $[q^j, q^{j+1})$. 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 $q \in [\underline{q}, q_{\max})$, we have

$$\begin{aligned} \hat{\xi}(q) \; = \; & \left[1 - \hat{G}^*(q)\right]^{\frac{\hat{\gamma}}{T'(q)}-1} \\ & \exp\left\{\hat{\gamma} \sum_{t=0}^{j-1} \left[\left(\frac{1}{T'(q^t)} - \frac{1}{T'(q^{t+1})}\right) \log(1 - \hat{G}^*(q^t))\right] + \hat{\gamma} \left(\frac{1}{T'(q^j)} - \frac{1}{T'(q)}\right) \log(1 - \hat{G}^*(q^j))\right\} \end{aligned} \tag{22}$$

if $q \in [q^j, q^{j+1}), j = 0, \ldots, J-1$, where $q^0 = \underline{q}$ and $q^J = q_{\max}$. In particular, if $q \in [\underline{q}, q^1)$, $\hat{\xi}(q) = 1$ because $\hat{G}^*(q) = 0$. For $q \in [q_{\max}, \overline{q}]$, $\hat{\xi}(\cdot)$ is constant and equal to

$$\lim_{q \uparrow q_{\max}} \hat{\xi}(q) = \exp\left\{\hat{\gamma} \sum_{t=0}^{J-1} \left[\left(\frac{1}{T'(q^t)} - \frac{1}{T'(q^{t+1})}\right) \log(1 - \hat{G}^*(q^t))\right]\right\}, \tag{23}$$

which is finite.[28] Thus, $\hat{\xi}(\cdot)$ is a well defined strictly positive *cadlag* (continue à droite, limites à gauche) function on $[\underline{q}, \overline{q}]$ with steps at $q^1 < q^2 < \ldots < q^{J-1}$.

To complete the first step, following (15) and B3 we estimate $v_0(\cdot)$ and $\theta(\cdot)$ by

$$\hat{v}_0(q) = \frac{T'(q)}{\hat{\xi}(q)}, \quad \hat{\theta}(q) = \hat{\xi}(q) \text{ for } q \in [\underline{q}, \overline{q}]. \tag{24}$$

[28]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 $\lim_{q \uparrow q_{\max}} [1 - \hat{G}^*(q)]^{1-(\hat{\gamma}/T'(q))} = 1$ as $\lim_{q \uparrow q_{\max}} T'(q) = \hat{\gamma}$.

In particular, $\hat{v}_0(\cdot)$ and $\hat{\theta}(\cdot)$ are also well-defined strictly positive cadlag functions with steps at $q^1 < q^2 < \ldots < q^{J-1}$. If $T'(\cdot)$ is strictly decreasing so that $T(\cdot)$ is concave, then it can be shown that $\hat{v}_0(\cdot)$ and $\hat{\theta}(\cdot)$ are strictly decreasing and increasing on $[\underline{q}, q_{\max}]$, respectively. The pseudo sample of businesses' types is $\hat{\theta}_i = \hat{\theta}(q_i)$ for $i = 1, \ldots, N^*$.

In the second step, we estimate the truncated business' type density from the pseudo sample by using the kernel estimator

$$\hat{f}^*(\theta) = \frac{1}{N^* h} \sum_{i=1}^{N^*} K\left(\frac{\theta - \hat{\theta}_i}{h}\right), \tag{25}$$

for $\theta \in (\theta_0, \overline{\theta}) = (1, \overline{\theta})$, where $K(\cdot)$ is a symmetric kernel function with compact support and h is a bandwidth. Moreover, from (18) we can estimate the conditional density $f_{\epsilon|Z_1,Z_2}(\cdot|z_1, z_2)$ by $(1 - \hat{\Pr}[\tilde{\theta} = \theta_0 | Z_1 = z_1^o, Z_2 = z_2]) \hat{f}^*_{\theta|Z_1,Z_2}(\cdot|z_1^o, z_2)$. Here, $\hat{f}^*_{\theta|Z_1,Z_2}(\cdot|z_1^o, z_2) = \hat{f}^*_{\theta,Z_1,Z_2}(\cdot|z_1^o, z_2)/\hat{f}_{Z_1,Z_2}(z_1^o, z_2)$ with numerator and denominator similar to (25) with additional kernels, while $\hat{\Pr}[\tilde{\theta} = \theta_0 | Z_1 = z_1^o, Z_2 = z_2]$ is obtained from the kernel estimate of the regression $\mathrm{E}[1\!\!1(q_i = \underline{q})|Z_{i1} = z_1^o, Z_{i2} = z_2]$ on $i = 1, \ldots, N$, where N is the total number of businesses. On the other hand, from (17) $r(z_1, z_2, e)$ is estimated by the $\hat{F}^*_{\epsilon|Z_1,Z_2}(e|z_1, z_2)$-quantile of $\hat{F}^*_{\theta|Z_1,Z_2}(\cdot|z_1, z_2)$. The latter conditional c.d.f. is estimated by combining a counting process on $\hat{\theta}$ as in (21) and kernels for (Z_1, Z_2), while the former is estimated using (19) which involves the nonparametric regression $\mathrm{E}[1\!\!1(q_i = \underline{q})|Z_{i1} = z_1, Z_{i2} = z_2]$.

We make the following assumption on the data generating process.

Assumption B5: *The observed and unobserved heterogeneity $(Z_i, \epsilon_i), i = 1, \ldots, N$, are independent and identically distributed (i.i.d.).*

Consequently, given B1 the businesses' types $\theta_i, i = 1, \ldots, N$ are i.i.d. as $F(\cdot)$. Similarly, since $q_i = q(\theta_i) 1\!\!1(\theta_i > \theta_0) + \underline{q} 1\!\!1(\theta_i \leq \theta_0)$, the observed advertising quantities q_i are i.i.d.

The next lemma establishes the strong consistency of $\hat{\gamma} = T'(q_{\max})$ for the marginal cost $\gamma = C'(Q)$ with a rate of convergence faster than $\sqrt{N}$. It also provides the asymptotic distribution of $\hat{\gamma}$. These properties follow from the delta method combined with known asymptotic properties of the highest order statistic $q_{\max}$ from e.g. Galambos (1978).

Lemma 6: *Under A2, A4-(iv), B1–B2 and B5, we have (i) $\hat{\gamma} = \gamma + O_{a.s.}[(\log \log N)/N]$,* *and (ii)* $N(\hat{\gamma}-\gamma) \xrightarrow{D} -T''(\overline{q})\mathcal{E}/[g^*(\overline{q})(1-F(\theta_0))]$, *where* $\mathcal{E}$ *is standard exponential distributed, as* $N \to \infty$.

Following Campo, Guerre, Perrigne and Vuong (2011), $g^*(\overline{q})$ can be estimated consistently by a one-sided kernel density estimator $\hat{g}^*(\cdot)$ evaluated at $q_{\max}$, while $T''(\overline{q})$ and $1-F(\theta_0)$ can be estimated as usual by $T''(q_{\max})$ and N^*/N, respectively. Thus, $N(\hat{\gamma}-\gamma)\hat{g}^*(q_{\max})/T''(q_{\max}) \xrightarrow{D} -\mathcal{E}$, which can be used for hypothesis testing on γ.

The next proposition establishes the asymptotic properties of $\hat{v}_0(\cdot)$ and $\hat{\theta}(\cdot)$. Following the empirical process literature introduced in econometrics by Andrews (1994), we view $\hat{v}_0(\cdot)$ and $\hat{\theta}(\cdot)$ as stochastic processes defined on $[\underline{q},\overline{q}]$ and hence as random elements of the space $\mathcal{D}[\underline{q},\overline{q}]$ of cadlag functions on $[\underline{q},\overline{q}]$. Because of $\log[1-G^*(q)]$ in (16), we consider instead the space $\mathcal{D}[\underline{q},q_\dagger]$ with its uniform metric $\|\psi_1-\psi_2\|_\dagger = \sup_{q\in[\underline{q},q_\dagger]}|\psi_1(q)-\psi_2(q)|$, where $q_\dagger \in (\underline{q},\overline{q})$.[29] Weak convergence on the space $\mathcal{D}[\underline{q},q_\dagger]$ is denoted by "$\Rightarrow$."

Proposition 4: *Under A2, A4-(iv), B1–B3 and B5, for any* $q_\dagger \in (\underline{q},\overline{q})$, *we have*

(i) $\|\hat{v}_0(\cdot)-v_0(\cdot)\|_\dagger \xrightarrow{a.s.} 0$ *and* $\|\hat{\theta}(\cdot)-\theta(\cdot)\|_\dagger \xrightarrow{a.s.} 0$ *as* $N\to\infty$,

(ii) as random functions in $\mathcal{D}[\underline{q},q_\dagger]$, $\sqrt{N}[\hat{v}_0(\cdot)-v_0(\cdot)] \Rightarrow -v_0(\cdot)Z(\cdot)/\sqrt{1-F(\theta_0)}$ *and* $\sqrt{N}[\hat{\theta}(\cdot)-\theta(\cdot)] \Rightarrow \theta(\cdot)Z(\cdot)/\sqrt{1-F(\theta_0)}$ *as* $N\to\infty$, *where* $Z(\cdot)$ *is a tight Gaussian process defined on* $[\underline{q},q_\dagger]$ *by*

$$Z(\cdot) = \left[1-\frac{\gamma}{T'(\cdot)}\right]\frac{\mathcal{B}_{G^*}(\cdot)}{1-G^*(\cdot)} - \gamma\int_{\underline{q}}^{\cdot}\frac{T''(x)}{T'(x)^2}\frac{\mathcal{B}_{G^*}(x)}{1-G^*(x)}dx, \qquad (26)$$

where $\mathcal{B}_{G^*}(\cdot)$ *is the* G^**-Brownian bridge process on* $[\underline{q},\overline{q}]$.[30]

The first part establishes the uniform almost sure convergence of $\hat{v}_0(\cdot)$ and $\hat{\theta}(\cdot)$ on any subset $[\underline{q},q_\dagger]$ with $q_\dagger < \overline{q}$. The second part gives the asymptotic distributions of $\hat{v}_0(\cdot)$ and $\hat{\theta}(\cdot)$. It is worthnoting that their rates of convergence are the parametric rate $\sqrt{N}$. This remarkable result comes from that $\hat{v}_0(\cdot)$ and $\hat{\theta}(\cdot)$ are smooth functionals of

[29] As usual measurability issues are ignored below. This can be addressed by considering either the projection σ-field on $\mathcal{D}[\underline{q},q_\dagger]$ 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).

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

the empirical c.d.f. $\hat{G}^*(\cdot)$. The appendix shows that $Z(\cdot)$ has zero mean and covariance $\mathrm{E}[Z(q)Z(q')] = \omega(q)$ for $\underline{q} \le q \le q' \le q_\dagger$, where

$$\omega(q) \quad = \quad \left(1 - \frac{\gamma}{T'(q)}\right)^2 \frac{G^*(q)}{1 - G^*(q)} - 2\gamma \int_{\underline{q}}^{q} \left(1 - \frac{\gamma}{T'(x)}\right) \frac{T''(x)}{T'(x)^2} \frac{G^*(x)}{1 - G^*(x)} dx. \qquad (27)$$

Note that the covariance $\mathrm{E}[Z(q)Z(q')]$ is independent of q'. Thus, the covariance of the limiting process of $\sqrt{N}[\hat{v}_0(\cdot) - v_0(\cdot)]$ is $v_0(q)v_0(q')\omega(q)/[1 - F(\theta_0)]$ for $\underline{q} \le q \le q' \le q_\dagger$. In particular, Proposition 4-(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 $q \in [\underline{q}, q_\dagger]$. The asymptotic variance of $\hat{v}(q)$ vanishes at $q = \underline{q}$ as $\omega(\underline{q}) = 0$, which is expected since $\hat{v}_0(\underline{q}) = T'(\underline{q})$, while $\omega(q)$ increases as q increases to $q_\dagger$ 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 (27) by replacing γ and $G^*(\cdot)$ by their estimates $T'(q_{\max})$ and $\hat{G}^*(\cdot)$.[31] Similar comments apply to $\hat{\theta}(\cdot)$.[32]

Regarding the second step, we note that the pseudo businesses' types $\hat{\theta}_i = \hat{\xi}(q_i), i = 1, \ldots, N^*$ converge uniformly to the businesses' types θ_i at the rate $\sqrt{N^*/\log\log N^*}$, which

[31] Recalling that $N^*/N \xrightarrow{a.s.} 1 - F(\theta_0)$, Proposition 4-(ii) can also be used to deliver an asymptotic "uniform" confidence interval for $v_0(\cdot)$ 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 $q \in [\underline{q}, q_\dagger]$, where c is the constant defined by $\Pr(||Z(\cdot)||_\dagger \le c) = 1 - \alpha$ with $0 < \alpha < 1$.

[32] Estimation of $\overline{\theta}$ is more delicate. A natural estimator is $\theta_{\max} = \hat{\xi}(q_{\max})$ as given by (23). To achieve consistency, a correction is needed. Specifically, let $\hat{\overline{\theta}} = \hat{\xi}(\overline{q}_N)$, where $\overline{q}_N \equiv q_{\max} - b_N$ with $b_N \downarrow 0$ and $Nb_N/\log\log N \to \infty$. In particular, $\overline{q}_N$ converges almost surely to $q_{\max}$ but at a slower rate than $N/\log\log N$. From Wellner (1978) Corollary 1, it follows that $\sup_{q\in[\underline{q},\overline{q}_N]} \left|[(1 - \hat{G}^*(q))/(1 - G^*(q))] - 1\right| \xrightarrow{a.s.} 0$ and hence $\sup_{q\in[\underline{q},\overline{q}_N]} \left|\log[(1 - \hat{G}^*(q))/(1 - G^*(q))]\right| \xrightarrow{a.s.} 0$ as $N \to \infty$. A proof similar to that of Lemma 8-(i) then establishes that $\sup_{q\in[\underline{q},\overline{q}_N]} \left|\hat{\xi}(q) - \xi(q)\right| \xrightarrow{a.s.} 0$. In particular, $\hat{\xi}(\overline{q}_N) \xrightarrow{a.s.} \xi(\overline{q})$ since $\xi(\overline{q}_N) \xrightarrow{a.s.} \xi(\overline{q})$.

is the uniform rate of convergence of $\sup_{q \in [\underline{q}, q_\dagger]} |\hat{G}^*(q) - G^*(q)|$ from e.g. van der Vaart (1998, p.268). Because this rate is faster than the optimal 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 (25), which uses the pseudo businesses' types, possesses the standard asymptotic properties of uniform convergence and limiting distribution, namely (i) $\sup_{\theta \in C} |\hat{f}^*(\theta) - f^*(\theta)| \stackrel{a.s.}{\to} 0$ for any compact subset C of $(\theta_0, \overline{\theta})$ provided $h \to 0$ and $N^*h/\log N^* \to \infty$ and (ii) $\sqrt{N^*h}[\hat{f}^*(\theta) - f^*(\theta)] \stackrel{d}{\to} \mathcal{N}(0, f^*(\theta) \int K(x)^2 dx)$ 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 is achieved by our two-step procedure when the θ_is are unobserved but businesses' advertisements q_i are observed. Estimation of $f(\cdot)$ is obtained from $f(\cdot) = (1 - F(\theta_0))f^*(\cdot)$ on $[\theta_0, \overline{\theta}]$, where $1 - F(\theta_0)$ is estimated by N^*/N at the $\sqrt{N}$ rate. For the same reason, the conditional density $f_{\epsilon|Z_1,Z_2}(\cdot|\cdot,\cdot)$ possesses the standard asymptotic properties on any compact subset of $\{(z_1, z_2, e) \in \mathcal{S}_{Z_1,Z_2,\epsilon} : e \geq \theta_0\}$ because $\hat{\Pr}[\tilde{\theta} = \theta_0 | Z_1 = z_1^o, Z_2 = z_2]$ converges faster than $\hat{f}^*_{\theta|Z_1,Z_2}(\cdot|z_1^o.z_2)$. Lastly, the estimator of $r(\cdot,\cdot,\cdot)$ achieves the same rate of convergence of a conditional quantile given (Z_1, Z_2) at every value in a compact subset of $\{(z_1, z_2, e) \in \mathcal{S}_{Z_1,Z_2,\epsilon} : e \geq \max[e(z_1, z_2), \theta_0]\}$.

5 Empirical Results

Using the quality-adjusted quantities from (C.1), the price schedule and the firm's purchases, we apply the estimators (21)– (24), where $q_{\max} = 6,230$. Thus, the marginal cost in dollars for an additional quality-adjusted square pica at the total production is $\hat{\gamma} = T'(6,230) = 8.29$, which seems reasonable in the publishing industry. Thus, an additonal line of listing or 1.85 quality-adjusted square picas costs at the margin for the publisher \$15.34, which is charged \$100.80 to businesses. If one assumes a constant marginal cost, an additional no color full page or 1,470 quality-adjusted pica square would cost at the margin \$12,186, which is priced at \$18,513. Data from a single phone directory do not allow us to recover more of the cost function as explained in Section 4.1.

The estimated marginal payoff $\hat{v}_0(\cdot)$ and the business' type $\hat{\theta}(\cdot)$ as functions of q are displayed in Figures 1 and 2. These figures are based (20) and (24).[33] Figure 1 shows that $\hat{v}_0(\cdot)$ is strictly decreasing as assumed in B1, while Figure 2 shows that $\hat{\theta}(\cdot)$ is strictly increasing as predicted by the model. Using (25) with a triweight kernel and a rule-of-thumb bandwidth, the estimated (truncated) density of types $\hat{f}^*(\cdot)$ on $[1, \theta_{\max}] = [1, 13.95]$ is displayed in Figure 3. The estimated density reveals substantial heterogeneity among businesses in their taste for advertising. We note that the density displays two modes, a first one around 1.3 and a second one around 2.3. The first one corresponds to businesses adding a few more lines and/or a larger font to their standard listing, while the second mode corresponds to businesses choosing the smallest size of space. Figure 4 displays $\theta - [(1 - \hat{F}(\theta))/\hat{f}(\theta)]$, which is strictly increasing thereby satisfying A4-(iv).

With $\hat{v}_0(\cdot)$ and $\hat{\theta}(\cdot)$, we can assess the payoff or informational rent for each business buying advertising using $U(q_i, \theta_i) = U_0(\theta_i) + \theta_i \int_{\underline{q}}^{q_i} v_0(x)dx - T(q_i), i = 1, \ldots, 4,671$. The term $U_0(\theta) = \theta \int_0^{\underline{q}} v_0(x)dx$ is not identified since $v_0(\cdot)$ is not identified on $[0, \underline{q})$. On the other hand, we can estimate the additional rent for advertising beyond $\underline{q}$ by $\Delta\hat{U}_i \equiv \hat{U}(q_i, \hat{\theta}_i) - U_0(\hat{\theta}_i) = \hat{\theta}_i \int_{\underline{q}}^{q_i} \hat{v}_0(x)dx - T(q_i)$. As expected, this additional rent strictly increases with the quality-adjusted advertising size purchased. Table 3 gives summary statistics on $\Delta\hat{U}_i$ as well as on the payment T_i, quality-adjusted advertising size q_i, base marginal utility $\hat{v}_0(q_i)$ and type $\hat{\theta}_i$ for the 4,671 businesses buying advertising. Table 3 also gives this additional rent as a proportion of the gross additional rent $\Delta\hat{U} + T$. Given that $U_0(\theta_i)$ is not included, our results provide a lower bound for the informational rent left to businesses. Informational rents are thus quite substantial for businesses buying large quantities. Overall, the sum of the $\Delta\hat{U}_i$ is \$4,128,906, which can be compared to the total payment made by the businesses \$6,017,632. This gives an overall informational rent of the order of 40.69%, which measures the extent of the cost of asymmetric information. In particular, in a world of complete information, the publisher would know each business' type and the price that the publisher would charge would be equal to each business'

[33]To obtain smooth representations, we have used the smoothed empirical distribution with a triweight kernel density estimator and a rule-of-thumb bandwidth instead of (21).

benefit, i.e. $\theta \int_{\underline{q}}^{q} v_0(x)dx$ leaving the minimal rent $U_0(\theta)$ to each business.

In view of Proposition 3, we could estimate $r(\cdot,\cdot)$ and $F_{\epsilon|Z}(\cdot|\cdot)$. Our data provide the business heading and the corresponding number of businesses in the heading as potential Z. When decomposing the total variance of the $\hat{\theta}_i, i = 1, \ldots, 4671$, we find that only 27.08% of the total variance is due to variation between headings. This suggests that the variability of the businesses' types is mainly due to within factors. This is confirmed when considering a Tobit model due to the censoring on θ. With all the headings as dummy variables one obtains a pseudo R^2 of 0.1163 with only 180 heading dummies among 1,193 that are significant at the 10% level. Moreover, with only the number of businesses in each heading, the pseudo R^2 decreases to 0.0046. Lastly, the correlation between the business' type $\hat{\theta}$ and the mean of the businesses' types in the same heading is low, namely 0.3666 when considering the full sample and 0.3454 when considering the subsample of businesses buying advertising. These results suggest that (i) the decision of buying advertising and the amount of advertising purchased are mostly idiosyncratic and possibly due to variables that are not available such as the business' size, and (ii) competition in advertising within the same heading, if any, involves at most a few businesses.

With structural estimates, we can perform some counterfactuals. Hereafter, we assume a constant marginal cost, i.e. $C(Q) = \hat{\gamma}Q$. The estimated profit under the observed nonlinear pricing is then \$3,112,442. An alternative pricing rule consists in a linear price p, in which any size is charged the same amount at the margin so that $T(q) = pq$. The business' first-order condition (4) still holds and can be written as $p = \theta v_0(q)$ for $\theta > \theta_0^L$, where θ_0^L is the threshold level associated with a linear price given by $\theta_0^L = p/v_0(\underline{q})$. Equivalently, given p we have $q = v_0^{-1}(p/\theta)$ for $\theta > p/v_0(\underline{q})$ and $q = \underline{q}$ for $\theta \leq p/v_0(\underline{q})$. Using the estimate of $v_0(\cdot)$, we can estimate the publisher's expected profit by

$$p \int_{p/\hat{v}_0(\underline{q})}^{\overline{\theta}} \hat{v}_0^{-1}(p/\theta)\hat{f}(\theta)d\theta - \hat{\gamma}\left[\underline{q}\hat{F}(p/\hat{v}_0(\underline{q})) + \int_{p/\hat{v}_0(\underline{q})}^{\overline{\theta}} \hat{v}_0^{-1}(p/\theta)\hat{f}(\theta)d\theta\right],$$

provided $p \geq \hat{v}_0(\underline{q})$ or $\theta_0^L \geq 1$ since we identify $F(\cdot)$ and $v_0(\cdot)$ on $[1, \theta_{\max}]$ and $[\underline{q}, q_{\max}]$, respectively. From Table 3, the lowest possible value for p is \$38.72 per quality-adjusted pica square. The largest expected profit is \$1,739,502, which is obtained for $p = 38.72$. Thus, this value is a lower bound for the profit under an optimal linear price $p \leq 38.72$.

In particular, the optimal linear price may exclude fewer businesses or equivalently may induce businesses with $\theta \leq 1$ to buy advertising. On the other hand, businesses with large θ values will buy lower quantities than under the current nonlinear price because these large buyers will not benefit from quantity discounts. On aggregate, the businesses' total payoff may increase as businesses with low θ can buy larger quantities of advertising.

A second counterfactual considers third-degree price discrimination. We consider three possible categories: (i) businesses with households/individuals as primary customers, (ii) businesses with businesses as primary customers and (iii) non-profit organizations such as federal, state and local agencies, charities and churches. Table 4 provides summary statistics on $\hat{\theta}_i$, the number of businesses and the actual proportion of businesses buying advertising for each group. Businesses in the first group have the largest values and are also those which are more likely to buy advertsing, while businesses in the third group have the lowest values and are least likely to buy advertising. Figure 5 displays the estimated densities for the three groups. Assuming a constant marginal cost, third-degree price discrimination reduces to apply (5) for each group. In particular, with the estimated density and distribution for each group, we use (13) and (14) to obtain the quantity schedule $q(\cdot)$ and the optimal tariff $T(\cdot)$ for each group. Figures 6 and 7 display these functions. We observe that prices will be lower for the three groups, while quantities will be larger. For the first two groups, more businesses will be excluded as the optimal threshold type is slightly above 1 at 1.0396.[34] For the first (second) group, 67.92% (64.75%) of businesses will buy advertsing. Despite lower prices, the average amount spent on advertising will be larger for the three groups. For instance, while the actual average price paid by a business in the first group is $1,431, this amount will increase to $1,652. For the second group, the amount will increase to $1,216 from $1,067. Lastly, the expected profit for the publisher will be $3,716,134, which represents a 19.4% increase.

[34]The threshold level for the third group was set at 1 because the density is not identified for values below. Because the third group contains businesses with low types, one can expect that fewer businesses would be excluded. Thus, our estimated publisher's profit from this group is a lower bound.

6 Concluding Remarks

This paper studies the identification and estimation of the nonlinear pricing model. Identification is achieved by exploiting the first-order conditions of the publisher's profit maximization and the unique mapping between the business' type and its purchased quantity. This result is in the spirit of the recent literature on the nonparametric identification of incomplete information models. A nonparametric two-step estimation procedure is then developed to estimate the model primitives. Asymptotic properties of our estimator are established. Data on advertising in yellow pages from a utility publisher in Central Pennsylvania are analyzed. Our empirical results show an important heterogeneity in businesses' types and a significant loss due to asymmetric information for the publisher.

Several extensions can be entertained. When there is a discrete price schedule, pooling arises at the equilibrium and limits the extent to which one can identify the model primitives. Recent results by Aryal, Perrigne and Vuong (2009) in the context of insurance suggest that observations on repeated purchases allow us to identify the type distribution while avoiding the use of the first-order conditions of the principal. Another extension of interest is to test adverse selection in this market. To perform such a test, we need to derive all the restrictions imposed by the nonlinear pricing model under both incomplete and complete information. We then can test which set of restrictions are validated by the data. Lastly, the methodology developed in this paper in combination with Perrigne and Vuong (2011) can entertain other models with adverse selection in contract theory with a broad range of applications to retailing and labor to name a few.

Appendix A

This appendix collects the proofs of Lemmas and Proposition in Section 3.

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 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 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: Consider the change of variables $t = \overline{\theta} - \theta$ and $t_0 = \overline{\theta} - \theta_0$ in (7) so that θ_0 becomes the terminal time. Hence, using $q_0 = \underline{q}$, (7) becomes

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

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

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

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 problem (A.2) can be written as a standard free terminal time and free-end point control problem

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

where

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

The Hamiltonian function 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. From Kirk (1970, pp. 188 and 192), the necessary conditions for $X(\cdot),\overline{q}(\cdot)$ and $\lambda(\cdot)$ to be solutions of (A.3) are

$$\begin{aligned} \mathcal{H}_2\left[X(t),\overline{q}(t),\lambda(t),t\right] &= 0 \quad \forall t \in [0,\overline{\theta}-\theta_0) \\ \lambda'(t) &= -\mathcal{H}_1\left[X(t),\overline{q}(t),\lambda(t),t\right] \quad \forall t \in [0,\overline{\theta}-\theta_0) \\ \lim_{t\uparrow t_0}\lambda(t) &= \lim_{t\uparrow t_0} K_1\left[X(t),t\right] \\ \lim_{t\uparrow t_0}\mathcal{H}\left[X(t),\overline{q}(t),\lambda(t),t\right] &= -\lim_{t\uparrow t_0} K_2\left[X(t),t\right]. \end{aligned} \tag{A.5}$$

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

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

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

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

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

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

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

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

Plugging the optimal $\lambda(\cdot)$ into (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)

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

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

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

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

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

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 equality can be written as $\int_{\underline{q}}^{\overline{q}_-} \left[\Gamma(x,t_0) - \Gamma(\overline{q}_-,t_0)\right] dx = 0$, which implies $\overline{q}_- = \underline{q}$ by A3 when $\Gamma(x,t_0)$ is monotone in x. Alternatively, if $\Gamma(\cdot,t_0) = 0$, then the left-hand side of (A.11) is zero implying that $\overline{q}_- = \underline{q}$ because $C'(\cdot) > 0$ by A2. This establishes (9).

It suffices now to rewrite (A.10) in terms of $q(\cdot)$, $v(\cdot,\cdot)$ and $F(\cdot)$ to establish (8), while (10) is nothing else than the IC constraint (4). Equation (11) follows from $U_+ = 0$, which is equivalent to $\lim_{\theta \downarrow \theta_0} \int_{\underline{q}}^{q(\theta)} v(x,\theta)dx = \lim_{\theta \downarrow \theta_0} T(q(\theta))$. The left-hand side is equal to zero using (9) thereby establishing (11) since $q(\cdot)$ is strictly increasing on $(\theta, \overline{\theta}]$. □

Proof of Lemma 2: To simplify the exposition, we suppress the arguments of functions hereafter. 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

$$q' = \frac{v_2\left[(1/v_2) \times (\partial(v_2/\rho)/\partial\theta) - 1\right]}{v_1 - (v_{21}/\rho)}. \quad \text{(A.12)}$$

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

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), $[v^2/(qv_2)] \times [\partial(-qv_1/v)/\partial\theta] \leq 0$. Thus, $v_1 - (v_{21}/\rho) < 0$. Hence under A1-(iii), A2 and A4-(i,ii), $q' > 0$ on $(\theta_0, \overline{\theta}]$. Suppose $v_{21} \geq 0$. It is straightforward to see that the denominator of (A.12) 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 using (9). We now show that $q(\cdot)$ is continuously differentiable at θ_0, i.e. that $q'(\theta_0)$ exists and is strictly positive with $q'(\theta_0) = \lim_{\theta\downarrow\theta_0} q'(\theta) < \infty$. 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 $\theta_0 < \tilde{\theta} < \theta$. But, $\lim_{\theta\downarrow\theta_0} q'(\theta)$ is equal to the right-hand side of (A.12) 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 $(\theta_0, \overline{\theta}]$ from above, it follows that the IC constraint (4) can be written as $T'(q) = v(q, \theta(q))$ $\forall q \in (\underline{q}, \overline{q}]$. Recall that $v(\cdot, \cdot)$ is continuously differentiable on $[\underline{q}, +\infty) \times [\underline{\theta}, \overline{\theta}]$ by A1 while $\theta(\cdot)$ is continuously differentiable on $[\underline{q}, \overline{q}]$ 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 $(\underline{q}, \overline{q}]$, i.e. $T(\cdot)$ is twice continuously differentiable on $(\underline{q}, \overline{q}]$. We now show that $T(\cdot)$ is twice continuously differentiable at $\underline{q}$. We note that $T'(\cdot)$ exists and is continuous at $\underline{q}$. 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 $\underline{q}$ by (11) and continuously differentiable on $(\underline{q}, \overline{q}]$. Thus, (4) holds on $[\underline{q}, \overline{q}]$ showing that $T'(\cdot)$ is continuously differentiable on $[\underline{q}, \overline{q}]$ as $v(\cdot, \cdot)$ and $\theta(\cdot)$ are continuously differentiable on $[\underline{q}, +\infty) \times [\underline{\theta}, \overline{\theta}]$ and $[\underline{q}, \overline{q}]$, respectively. Hence, $T(\cdot)$ is twice continuously differentiable on $[\underline{q}, \overline{q}]$ and $T'(\underline{q}) = v(q_0, \theta_0) > 0$. Regarding the other 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 $(\underline{q}, \overline{q})$ by A1-(iii), while $T'(\overline{q}) = C'(Q)$. $\square$

Appendix B

This appendix collects the proofs of Lemmas and Propositions in Section 4.

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 $\tilde{Q} \equiv \underline{q}\tilde{F}(\tilde{\theta}_0) + \int_{\tilde{\theta}_0}^{\overline{\tilde{\theta}}} \tilde{q}(u)\tilde{f}(u)du$. 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 $\left[\tilde{T}(\cdot), \underline{q}, \tilde{F}(\tilde{\theta}_0), \tilde{G}^*(\cdot), \tilde{Q}\right]$ generated by the structure $\tilde{S}$ are the same observables $\left[T(\cdot), \underline{q}, F(\theta_0), G^*(\cdot), Q\right]$ generated by the structure S. Lastly, 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

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

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

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

using the monotonicity of $\tilde{q}(\cdot)$ and $q(\cdot)$. Third, $\tilde{F}(\tilde{\theta}_0) = F(\alpha\theta_0/\alpha) = F(\theta_0)$.

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 $q'(\cdot) > 0$ on $[\theta_0, \overline{\theta}]$, we have $G^*(q) = [F(\theta)-F(\theta_0)]/[1-F(\theta_0)]$ with a density $g^*(q) = \theta'(q)f(\theta)/[1-F(\theta_0)] > 0$ on $[\underline{q}, \overline{q}]$, where $\theta = \theta(q)$. Elementary algebra gives $[1-F(\theta)]/f(\theta) = \theta'(q)[1-G^*(q)]/g^*(q)$. Equations (13) and (14) lead to

$$T'(q) = \gamma + \frac{1-G^*(q)}{g^*(q)}\theta'(q)v_0(q) \text{ for } q \in (\underline{q}, \overline{q}]. \tag{B.1}$$

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.13) gives after some algebra

$$\theta(q)v_0'(q) = T''(q) - \frac{g^*(q)}{1-G^*(q)}\left[T'(q) - \gamma\right] \text{ for } q \in (\underline{q}, \overline{q}),$$

and hence for $q \in (\underline{q}, \overline{q}]$ by continuity. Dividing the left-hand side by $\theta(q)v_0(q)$ and the right hand side by $T'(q) = \theta v_0(q)$ gives

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

Integrating both sides of the above equation from $\underline{q}$ to q gives

$$\log\left(\frac{v_0(q)}{v_0(\underline{q})}\right) = \log\left(\frac{T'(q)}{T'(\underline{q})}\right) - \int_{\underline{q}}^{q} \frac{g^*(x)}{1-G^*(x)}\left(1 - \frac{\gamma}{T'(x)}\right)dx \text{ for } q \in (\underline{q}, \overline{q}]. \tag{B.2}$$

Taking the exponential gives

$$\frac{v_0(q)}{v_0(\underline{q})} = \frac{T'(q)}{T'(\underline{q})}\exp\left[-\int_{\underline{q}}^{q} \frac{g^*(x)}{1-G^*(x)}\left(1 - \frac{\gamma}{T'(x)}\right)dx\right] \text{ for } q \in (\underline{q}, \overline{q}]. \tag{B.3}$$

Condition (14) evaluated at $\underline{q}$ gives $T'(\underline{q}) = \theta_0 v_0(\underline{q})$. Multiplying the right-hand side of (A.15) by $T'(\underline{q})$ and the left-hand side by $\theta_0 v_0(\underline{q})$ gives for $q \in (\underline{q}, \overline{q}]$

$$\begin{aligned}
v_0(q) &= \frac{T'(q)}{\theta_0} \exp\left[-\int_{\underline{q}}^{q} \frac{g^*(x)}{1-G^*(x)}\left(1-\frac{\gamma}{T'(x)}\right)dx\right] \\
&= \frac{T'(q)}{\theta_0} \exp\left[-\int_{\underline{q}}^{q} \frac{g^*(x)}{1-G^*(x)}dx\right] \exp\left[-\gamma\int_{\underline{q}}^{q} \frac{-g^*(x)}{1-G^*(x)}\frac{1}{T'(x)}dx\right] \\
&= \frac{T'(q)}{\theta_0}[1-G^*(q)] \exp\left[-\gamma\left\{\frac{\log(1-G^*(x))}{T'(x)}\Big|_{\underline{q}}^{q} + \int_{q}^{\underline{q}} \log(1-G^*(x))\frac{T''(x)}{T'(x)}dx\right\}\right] \\
&= \frac{T'(q)}{\theta_0}\left[1-G^*(q)\right]^{1-\frac{\gamma}{T'(q)}} \exp\left\{-\gamma\int_{\underline{q}}^{q} \log\left[1-G^*(x)\right]\frac{T''(x)}{T'(x)^2}dx\right\},
\end{aligned}$$

where the third equality is obtained using integration by parts. This establishes (15) as $\theta(q) = T'(q)/v_0(q)$ by (14). Moreover, we have $\xi(\underline{q}) = 1$ and $\lim_{q\uparrow\overline{q}} \xi(q) = \overline{\theta}/\theta_0$ since $\overline{\theta} v_0(\overline{q}) = T'(\overline{q})$. All the derivations in the above proof are reversible, so the proof of sufficiency is omitted. □

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 $[\tilde{\underline{\theta}}, \tilde{\overline{\theta}}]$ and $\tilde{\theta}_0 = 1$. Both structures are assumed to be in $\mathcal{S}$ and to generate the same observables $[T(\cdot), \underline{q}, F(\theta_0), G^*(\cdot), Q]$. By Lemma 3, we note $C'(Q) = \tilde{C}'(Q) = \gamma$. In view of Lemma 5, the structure $\tilde{S}$ has to satisfy

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

By B2, $\theta_0 = \tilde{\theta}_0$ showing $\tilde{v}_0(\cdot) = v_0(\cdot)$ and $\theta(\cdot) = \tilde{\theta}(\cdot)$ on $(\underline{q}, \overline{q}]$ and hence on $[\underline{q}, \overline{q}]$ by continuity at $\underline{q}$. 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 $[\underline{q}, \overline{q}]$ and $[\theta_0, \overline{\theta}]$, respectively. Moreover, $\overline{\theta} = \tilde{\overline{\theta}}$ is identified as they are both equal to $\lim_{q\uparrow\overline{q}} \xi(q)$, while $v_0(\underline{q}) = \tilde{v}_0(\underline{q}) = T'(\underline{q})$ by Lemma 2 and B1-B2. □

Proof of Proposition 3: We begin by noting that for every $z = (z_1, z_2) \in \mathcal{S}_Z = \mathcal{S}_{Z_1, Z_2}$

$$\begin{aligned}
F^*_{\epsilon|Z}(e|z) &\equiv \Pr[\epsilon \leq e | Z = z, \epsilon > e(z)] \\
&= \Pr[\theta \leq r(z, e) | Z = z, \theta > 1] \\
&= F^*_{\theta|Z}(r(z, e)|z),
\end{aligned} \tag{B.4}$$

for every $e \in \mathcal{S}_{\epsilon|z}$, where the second equality follows from B4-(iii). Moreover,

$$F^*_{\epsilon|Z}(e|z) = \frac{F_{\epsilon|Z}(e|z) - F_{\epsilon|Z}(e(z)|z)}{1 - F_{\epsilon|Z}(e(z)|z)}, \tag{B.5}$$

for every $e \in [e(z), +\infty)$. We note that $F_{\epsilon|Z}(e(z)|z) = \Pr[\epsilon \leq e(z)|Z = z] = \Pr[r(z,\epsilon) \leq 1|Z = z] = \Pr[\tilde{\theta} = \theta_0|Z = z]$, thereby establishing that $F_{\epsilon|Z}(e(z)|z)$ is identified for every $z \in \mathcal{S}_Z$.

We now turn to the identification of $F_{\epsilon|Z}(\cdot|\cdot)$. Applying (B.4) and (B.5) at (z_1^o, z_2) with $z_2 \in \mathcal{S}_{Z_2|z_1^o}$ gives

$$\frac{F_{\epsilon|Z_1,Z_2}(e|z_1^o, z_2) - F_{\epsilon|Z_1,Z_2}(e(z_1^o, z_2)|z_1^o, z_2)}{1 - F_{\epsilon|Z_1,Z_2}(e(z_1^o, z_2)|z_1^o, z_2)} = F^*_{\theta|Z_1,Z_2}(e|z_1^o, z_2), \tag{B.6}$$

for every $e \in [e(z_1^o, z_2), +\infty) \cap \mathcal{S}_{\epsilon|z_1^o,z_2} = [\theta_0, +\infty) \cap \mathcal{S}_{\epsilon|z_1^o,z_2}$ since $e(z_1^o, z_2) = \theta_0$ as explained in the text. By B4-(i), we have $\mathcal{S}_{\epsilon|z_1^o,z_2} = \mathcal{S}_{\epsilon|z_1,z_2}$ and $F_{\epsilon|Z_1,Z_2}(e|z_1^o, z_2) = F_{\epsilon|Z_1,Z_2}(e|z_1, z_2)$. Using $\mathcal{S}_{Z_2|z_1^o} = \mathcal{S}_{Z_2}$ by B4-(ii), simple algebra establishes the expression for $F_{\epsilon|Z_1,Z_2}(e|z_1, z_2)$ in Proposition 3, and hence its identification on $\{(z_1, z_2, e) \in \mathcal{S}_{Z_1,Z_2,\epsilon} : e \geq \theta_0\}$.

Turning to the identification of $r(z, e)$, we note that combining (B.5) and (B.6) leads to the expression for $F^*_{\epsilon|Z}(\cdot|\cdot)$ in Proposition 3, thereby establishing its identification on $\{(z, e) \in \mathcal{S}_{Z,\epsilon} : e \geq \max[e(z), \theta_0]\}$. Inverting (B.4) establishes the expression for $r(z, e)$ in Proposition 3, and hence its identification on $\{(z, e) \in \mathcal{S}_{Z,\epsilon} : e \geq \max[e(z), \theta_0]\}$. Moreover, $\max[e(z), \theta_0]$ as a function of $z \in \mathcal{S}_Z$ is identified by

$$\begin{aligned} \max[e(z), \theta_0] \quad = \quad & F^{-1}_{\epsilon|Z}[\Pr(\tilde{\theta} = 1|Z = z)|z]\, 1\!\!1[\Pr(\tilde{\theta} = 1|Z = z) > \Pr(\tilde{\theta} = 1|Z = (z_1^o, z_2))] \\ & + \theta_0\, 1\!\!1[\Pr(\tilde{\theta} = 1|Z = z) \leq \Pr(\tilde{\theta} = 1|Z = (z_1^o, z_2))]. \end{aligned}$$

This follows from $e(z) > \theta_0$ if and only if $\Pr(\tilde{\theta} = 1|Z = z) > \Pr(\tilde{\theta} = 1|Z = (z_1^o, z_2))$, while $e(z) = F^{-1}_{\epsilon|Z}[\Pr(\tilde{\theta} = 1|Z = z)|z]$ whenever $e(z) > \theta_0$. □

Proof of Lemma 6: Given A2, A4-(iv) and B2, the assumptions of Lemma 2 are satisfied. Moreover, under B1 and B5 the observations $\{q_i, i = 1, \ldots, N^*\}$ are i.i.d. as $G^*(\cdot)$. This implies (i) $q_{\max} = \overline{q} + O_{a.s.}[(\log\log N^*)/N^*]$, and (ii) $N^*(q_{\max} - \overline{q}) \xrightarrow{D} -\mathcal{E}/g^*(\overline{q})$ as $N^* \to \infty$. These properties of $q_{\max}$ follow from e.g. Galambos (1978) noting that $g^*(\cdot)$ is continuous and bounded away from zero on $[\underline{q}, \overline{q}]$ as shown at the beginning of the proof of Lemma 5. Specifically, (i) follows from Galambos (1978) Theorem 4.3.1 and Example 4.3.2 by letting $u_{N^*} = \overline{q} - \kappa(\log\log N^*)/N^*$ for any $\kappa > 1$ so that $\sum_{N^*=2}^{\infty}[1 - G^*(u_{N^*})]\exp\{-N^*[1 - G^*(u_{N^*})]\} \equiv \sum_{N^*=2}^{\infty} v_{N^*} < \infty$ as $v_{N^*} \sim \tilde{v}_{N^*} \equiv \kappa g^*(\overline{q})[\log\log n^*]/[N^*(\log n^*)^{\kappa g^*(\overline{q})}]$ as $N^* \to$

∞ with $\sum_{N^*=2}^{\infty} \tilde{v}_{N^*} < \infty$. Thus, $\Pr[\overline{q} - q_{\max} \geq \kappa(\log\log N^*)/N^* \text{ i.o.}] = 0$, i.e. $\Pr[0 \leq (N^*/\log\log N^*)(\overline{q} - q_{\max}) \leq \kappa$ for N^* sufficiently large$] = 1$. Similarly, (ii) follows from Galambos (1978) Theorem 2.1.2 and Section 2.3.1 with $a_n = \overline{q}$ and $b_n = \overline{q} - G^{*-1}(1 - 1/N^*)$. Specifically, since $\lim_{t\to\infty}[1 - G^*(\overline{q} - 1/(tx))]/[1 - G^*(\overline{q} - 1/t)] = 1/x$ for $x > 0$, we obtain $(q_{\max} - \overline{q})/b_n \xrightarrow{D} -\mathcal{E}$, i.e. (ii) as $b_n \sim 1/[g^*(\overline{q})N^*]$ as $N^* \to \infty$. The lemma then follow from the standard delta method. Namely, $\hat{\gamma} - \gamma = T'(q_{\max}) - T'(\overline{q}) = T''(\tilde{q})(q_{\max} - \overline{q})$, where $q_{\max} < \tilde{q} < \overline{q}$ using a Taylor expansion and the continuous differentiability of $T'(\cdot)$ by Lemma 2. Moreover, $N^*/N \xrightarrow{a.s.} 1 - F(\theta_0) > 0$. □

To establish Proposition 3, we use two lemmas.

Lemma 7: *Under B1-B5, as $N^* \to \infty$ we have*

$$(i) \left\| \log\left(\frac{1 - \hat{G}^*(\cdot)}{1 - G^*(\cdot)}\right) \right\|_\dagger \xrightarrow{a.s.} 0, \ \ (ii)\sqrt{N^*}\log\left(\frac{1 - \hat{G}^*(\cdot)}{1 - G^*(\cdot)}\right) \Rightarrow \frac{-\mathcal{B}_{G^*}(\cdot)}{1 - G^*(\cdot)}.$$

on $[\underline{q}, q_\dagger]$ for every $q_\dagger \in (\underline{q}, \overline{q})$.

Proof of Lemma 7: Under B1-B5 the observations $\{q_i, i = 1, \ldots, N^*\}$ are independent and identically distributed as $G^*(\cdot)$. Let $\mathcal{D} = \mathcal{D}[\underline{q}, q_\dagger]$ with uniform metric $\|\cdot\|_\dagger$.

(i) Let $\mathcal{C}$ be the set of continuous functions on $[\underline{q}, q_\dagger]$ with uniform norms strictly smaller than one. Let $h(\cdot)$ map $\psi \in \mathcal{D}$ to $h(\psi) \in \mathcal{D}$, where $h(\psi)(\cdot) = \log[1 - \psi(\cdot)]$ if $\|\psi\|_\dagger < 1$ and $h(\psi)(\cdot) = 0$ otherwise. In particular, if $\psi_{N^*} \in \mathcal{D} \to \psi \in \mathcal{C}$, then for N^* sufficiently large $\|\psi_{N^*}\| < 1$ and $h(\psi_{N^*}) \to h(\psi)$. Now, $\|\hat{G}^* - G^*\|_\dagger \xrightarrow{a.s.} 0$ by the Glivenko-Cantelli theorem as $N^* \to \infty$. It follows from van der Vaart (1998) Theorem 18.11-(iii) that $\|h(\hat{G}^*) - h(G^*)\|_\dagger \xrightarrow{a.s.} 0$, where $h(G^*) = \log(1 - G^*)$ since $G^* \in \mathcal{C}$ as $\|G^*\|_\dagger \leq G^*(q_\dagger) < 1$. Noting that $h(\hat{G}^*) = \log(1 - \hat{G}^*)$ because $\|\hat{G}^*\|_\dagger \leq \hat{G}^*(q_\dagger) \leq (N^* - 1)/N^* < 1$, the desired property follows.

(ii) Let $h_{N^*}(\cdot)$ maps $\psi \in \mathcal{D}$ to $h_{N^*}(\psi) \in \mathcal{D}$, where $h_{N^*}(\psi)(\cdot) = \sqrt{N^*}\log\left[1 - \psi(\cdot)/\sqrt{N^*}\right]$ if $\|\psi\|_\dagger < \sqrt{N^*}$ and $h(\psi)(\cdot) = 0$ otherwise. Let $h_0(\cdot)$ be minus the identity. In particular, if $\|\psi_{N^*} - \psi\|_\dagger \to 0$ with $\psi \in \mathcal{D}$, we have $\|\psi_{N^*}\|_\dagger/N^{*\alpha} \to 0$ for any $\alpha > 0$ because $\|\psi\|_\dagger < \infty$. Hence, $h_{N^*}(\psi_{N^*})(\cdot) = \sqrt{N^*}\log\left[1 - \psi_{N^*}(\cdot)/\sqrt{N^*}\right]$ for N^* sufficiently large by taking $\alpha = 1/2$. Thus, using $\log(1 + x) = x - x^2/[2(1 + \tilde{x})^2]$ with $0 < |\tilde{x}| < |x|$ from a second-order Taylor expansion, we obtain for $x = -\psi_{N^*}(q)/\sqrt{N^*}$ and $0 < |\tilde{\psi}_{N^*}(q)| < |\psi_{N^*}(q)|/\sqrt{N^*}$

$$\|h_{N^*}(\psi_{N^*}) - h_0(\psi)\|_\dagger = \sup_{q\in[\underline{q}, q_\dagger]} \left|\sqrt{N^*}\log\left[1 - \psi_{N^*}(q)/\sqrt{N^*}\right] + \psi(q)\right|$$

$$
\begin{aligned}
&= \sup_{q\in[\underline{q},q_\dagger]}\left|\psi(q)-\psi_{N^*}(q)-\frac{\psi_{N^*}(q)^2}{2\sqrt{N^*}[1+\tilde{\psi}_{N^*}(q)]^2}\right| \\
&\le \|\psi_{N^*}-\psi\|_\dagger+\frac{1}{2}\frac{\sup_{q\in[\underline{q},q_\dagger]}\psi_{N^*}(q)^2/\sqrt{N^*}}{\inf_{q\in[\underline{q},q_\dagger]}[1+\tilde{\psi}_{N^*}(q)]^2} \\
&\le \|\psi_{N^*}-\psi\|_\dagger+\frac{1}{2}\frac{\|\psi_{N^*}\|_\dagger^2/\sqrt{N^*}}{[1-\|\psi_{N^*}\|_\dagger/\sqrt{N^*}]^2}.
\end{aligned}
$$

Thus, $\|h_{N^*}(\psi_{N^*})-h_0(\psi)\|_\dagger\to 0$ as $\|\psi_{N^*}-\psi\|_\dagger\to 0$ and $\|\psi_{N^*}\|_\dagger/N^{*\alpha}\to 0$ for $\alpha=1/2$ and 1. On the other hand, $\sqrt{N^*}(\hat{G}^*-G^*)/(1-G^*)\Rightarrow \mathcal{B}_{G^*}/(1-G^*)$ on $[\underline{q},q_\dagger]$ since $\sqrt{N^*}(\hat{G}^*-G^*)\Rightarrow \mathcal{B}_{G^*}$ as $N^*\to\infty$ by the Functional Central Limit Theorem. See e.g. van der Vaart (1998) Theorem 19.3. Applying van der Vaart (1998) Theorem 18.11-(i), it follows that

$$
h_{N^*}\left(\sqrt{N^*}\frac{\hat{G}^*-G^*}{1-G^*}\right)\Rightarrow h_0\left(\frac{\mathcal{B}_{G^*}}{1-G^*}\right)=\frac{-\mathcal{B}_{G^*}}{1-G^*}.
$$

But, as $N^*\to\infty$, $\|\sqrt{N^*}(\hat{G}^*-G^*)/(1-G^*)\|_\dagger<\sqrt{N^*}$ since $\|(\hat{G}^*-G^*)/(1-G^*\|_\dagger \xrightarrow{a.s.} 0$. Thus,

$$
h_{N^*}\left(\sqrt{N^*}\frac{\hat{G}^*-G^*}{1-G^*}\right)=\sqrt{N^*}\log\left[1-\frac{\hat{G}^*-G^*}{1-G^*}\right]=\sqrt{N^*}\log\left(\frac{1-\hat{G}^*}{1-G^*}\right).
$$

This establishes the desired result. □

Lemma 8: *Under A2-(iv), A4, B1–B3 and B5, for any* $q_\dagger\in(\underline{q},\overline{q})$, *we have as* $N\to\infty$

(i) $\|\hat{\xi}(\cdot)-\xi(\cdot)\|_\dagger\xrightarrow{a.s.}0$,

(ii) as a random function in $\mathcal{D}[\underline{q},q_\dagger]$, $\sqrt{N}[\hat{\xi}(\cdot)-\xi(\cdot)]\Rightarrow\xi(\cdot)Z(\cdot)/\sqrt{1-F(\theta_0)}$, *where* $Z(\cdot)$ *is the tight Gaussian process defined in (26).*

Proof of Lemma 8: By (16) and (20), $\xi(q)>0$ and $\hat{\xi}(q)>0$ for all $q\in[\underline{q},q_{\max})$. Thus,

$$
\begin{aligned}
\log\frac{\hat{\xi}(q)}{\xi(q)} &= \left[\frac{\gamma}{T'(q)}-1\right]\log\left[\frac{1-\hat{G}^*(q)}{1-G^*(q)}\right]+\gamma\int_{\underline{q}}^{q}\frac{T''(x)}{T'(x)^2}\log\left[\frac{1-\hat{G}^*(x)}{1-G^*(x)}\right]dx \\
&\quad+\frac{\hat{\gamma}-\gamma}{T'(q)}\log\left[1-\hat{G}^*(q)\right]+(\hat{\gamma}-\gamma)\int_{\underline{q}}^{q}\frac{T''(x)}{T'(x)^2}\log\left[1-\hat{G}^*(x)\right]dx, \qquad \text{(B.7)}
\end{aligned}
$$

for $q\in[\underline{q},q_{\max})$. Moreover, $q_\dagger<q_{\max}$ almost surely as $q_{\max}\xrightarrow{a.s.}\overline{q}$ by the proof of Lemma 6. Note also that $T'(\cdot)\ge C'(Q)>0$ on $[\underline{q},\overline{q}]$ so that $T'(\cdot)$ is bounded away from zero on $[\underline{q},\overline{q}]$.

(i) From (B.7), Lemma 6-(i) and Lemma 7-(i), we have $\|\log(\hat{\xi}/\xi)\|_\dagger\xrightarrow{a.s.}0$. By the continuous mapping theorem, we obtain $\|\hat{\xi}/\xi)\|_\dagger\xrightarrow{a.s.}1$ and hence $\|\hat{\xi}-\xi)\|_\dagger\xrightarrow{a.s.}0$ as $0<\|\xi\|_\dagger<\infty$.

(ii) Using Lemmas 6 and 7, it follows from (B.7) that

$$\sqrt{N}\log\frac{\hat{\xi}(q)}{\xi(q)} = \sqrt{\frac{N}{N^*}}\left\{\left[\frac{\gamma}{T'(q)}-1\right]\sqrt{N^*}\log\left[\frac{1-\hat{G}^*(q)}{1-G^*(q)}\right]+\gamma\int_{\underline{q}}^{q}\frac{T''(x)}{T'(x)^2}\sqrt{N^*}\log\left[\frac{1-\hat{G}^*(x)}{1-G^*(x)}\right]dx\right\} +o_{a.s.}(1)$$

uniformly in $q \in [\underline{q}, q_\dagger]$. Hence, using $N^*/N \xrightarrow{a.s} 1 - F(\theta_0) > 0$, it follows from Lemma 7-(ii) and the continuous mapping theorem that

$$\sqrt{N}\log\frac{\hat{\xi}(\cdot)}{\xi(\cdot)} \Rightarrow \frac{Z(\cdot)}{\sqrt{1-F(\theta_0)}} \tag{B.8}$$

by (26). From a first-order Taylor expansion we have $\sqrt{N}\log\left[\hat{\xi}(q)/\xi(q)\right] = \sqrt{N}\left[\hat{\xi}(q)-\xi(q)\right]/\tilde{\xi}(q)$, where $\xi(q) < \tilde{\xi}(q) < \hat{\xi}(q)$ for any $q \in [\underline{q}, q_\dagger]$. Thus, $\sqrt{N}\left[\hat{\xi}(q)-\xi(q)\right] = \tilde{\xi}(q)\sqrt{N}\log\left[\hat{\xi}(q)/\xi(q)\right]$. The desired result follows from (B.10) and $\|\tilde{\xi}-\xi\|_\dagger \xrightarrow{a.s.} 0$. Lastly, $Z(\cdot)$ is a tight Gaussian process as it is a bounded linear transformation of $\mathcal{B}_{G^*}(\cdot)$. □

Proof of Proposition 4: Using B3, we have $v_0(q) = T'(q)/\xi(q) > 0$ and $\theta(q) = \xi(q) > 0$ by (15), while $\hat{v}_0(q) = T'(q)/\hat{\xi}(q) > 0$ and $\hat{\theta}(q) = \hat{\xi}(q) > 0$ by (24) for all $q \in [\underline{q}, q_\dagger]$ as $T'(\cdot) > 0$ by Lemma 2. Moreover, $0 < \|\xi\|_\dagger < \infty$. Thus, (i) follows from Lemma 8-(i). Regarding (ii), we have $\sqrt{N}\log[\hat{v}_0(\cdot)/v_0(\cdot)] = \sqrt{N}\log[\theta(\cdot)/\hat{\theta}(\cdot)] = \sqrt{N}\log[\xi(\cdot)/\hat{\xi}(\cdot)] \Rightarrow -Z(\cdot)/\sqrt{1-F(\theta_0)}$ on $\mathcal{D}[\underline{q}, q_\dagger]$. The desired result follows by the argument at the end of the proof of Lemma 8. □

Proof of (27): We have $\mathrm{E}[\mathcal{B}_{G^*}(q)\mathcal{B}_{G^*}(q')] = G^*(q)[1-G^*(q')]$ for $\underline{q} \le q \le q' \le \overline{q}$. See e.g. van der Vaart (1998, p.266). Thus, for $\underline{q} \le q \le q' < \overline{q}$

$$\mathrm{E}\left[\frac{\mathcal{B}_{G^*}(q)}{1-G^*(q)}\frac{\mathcal{B}_{G^*}(q')}{1-G^*(q')}\right] = \frac{G^*(q)}{1-G^*(q)},$$

which is independent of q'. Hence, from the definition of $Z(\cdot)$ we have for $\underline{q} \le q \le q' < \overline{q}$

$$\begin{aligned}
\mathrm{E}[Z(q)Z(q')] &= \left[1-\frac{\gamma}{T'(q)}\right]\left[1-\frac{\gamma}{T'(q')}\right]\mathrm{E}\left[\frac{G^*(q)}{1-G^*(q)}\frac{G^*(q')}{1-G^*(q')}\right] \\
&\quad -\gamma\left[1-\frac{\gamma}{T'(q)}\right]\int_{\underline{q}}^{q'}\frac{T''(x)}{T'(x)^2}\mathrm{E}\left[\frac{\mathcal{B}_{G^*}(q)}{1-G^*(q)}\frac{\mathcal{B}_{G^*}(x)}{1-G^*(x)}\right]dx \\
&\quad -\gamma\left[1-\frac{\gamma}{T'(q')}\right]\int_{\underline{q}}^{q}\frac{T''(x)}{T'(x)^2}\mathrm{E}\left[\frac{\mathcal{B}_{G^*}(x)}{1-G^*(x)}\frac{\mathcal{B}_{G^*}(q')}{1-G^*(q')}\right]dx \\
&\quad +\gamma^2\int_{\underline{q}}^{q}\left\{\int_{\underline{q}}^{q'}\frac{T''(x)}{T'(x)^2}\frac{T''(x')}{T'(x')^2}\mathrm{E}\left[\frac{\mathcal{B}_{G^*}(x)}{1-G^*(x)}\frac{\mathcal{B}_{G^*}(x')}{1-G^*(x')}\right]dx'\right\}dx
\end{aligned}$$

$$
\begin{aligned}
= &\left[1-\frac{\gamma}{T'(q)}\right]\left[1-\frac{\gamma}{T'(q')}\right]\frac{G^*(q)}{1-G^*(q)} \\
&-\gamma\left[1-\frac{\gamma}{T'(q)}\right]\left\{\int_{\underline{q}}^{q}\frac{T''(x)}{T'(x)^2}\frac{G^*(x)}{1-G^*(x)}dx+\frac{G^*(q)}{1-G^*(q)}\int_{q}^{q'}\frac{T''(x)}{T'(x)^2}dx\right\} \\
&-\gamma\left[1-\frac{\gamma}{T'(q')}\right]\int_{\underline{q}}^{q}\frac{T''(x)}{T'(x)^2}\frac{G^*(x)}{1-G^*(x)}dx \\
&+\gamma^2\int_{\underline{q}}^{q}\frac{T''(x)}{T'(x)^2}\left\{\int_{\underline{q}}^{x}\frac{T''(x')}{T'(x')^2}\frac{G^*(x')}{1-G^*(x')}dx'+\frac{G^*(x)}{1-G^*(x)}\int_{x}^{q'}\frac{T''(x')}{T'(x')^2}dx'\right\}dx.
\end{aligned}
$$

Using $\int_a^b T''(x)/T'(x)^2 dx = [1/T'(a)] - [1/T'(b)]$ and

$$
\begin{aligned}
&\int_{\underline{q}}^{q}\frac{T''(x)}{T'(x)^2}\int_{\underline{q}}^{x}\frac{T''(x')}{T'(x')^2}\frac{G^*(x')}{1-G^*(x')}dx'dx \\
= &\frac{-1}{T'(q)}\int_{\underline{q}}^{q}\frac{T''(x)}{T'(x)^2}\frac{G^*(x)}{1-G^*(x)}dx+\int_{\underline{q}}^{q}\frac{1}{T'(x)}\frac{T''(x)}{T'(x)^2}\frac{G^*(x)}{1-G^*(x)}dx,
\end{aligned}
$$

which is obtained by integration by parts, the desired result follows after some algebra. □

Appendix C

As noted in Section 2, advertisements differ in size and quality (color and other features). It may be useful to review how quality can be integrated in nonlinear pricing models. Maskin and Riley (1984) consider a a monopolist who discriminates consumers by offering bundles of quantity and quality, both taking continuous values. With a one dimensional parameter of adverse selection, the optimal bundle should lay on an increasing curve in the quantity-quality space. Our data, however, display a weak correlation between the number of colors and the advertising size chosen by businesses. In particular, businesses choose several color options for a given advertising size. An alternative but more model would be to consider a multidimensional parameter of adverse selection. This would lead to multidimensional screening, which is known to require restrictive parametric assumptions on the model primitives to be solved. See Rochet and Chone (1998) and Rochet and Stole (2003) .

Instead, we observe that the ratio of prices for two different color options remains constant across sizes. For instance, considering the no color and one color display price schedules, this ratio equals 1.5 across the 9 possible advertisement sizes. Similarly, when considering the no color and multicolor display price schedules, the ratio equals 1.75 across advertisement sizes. Based on the Yellow Page Association data, such a pricing strategy for colors is used by Verizon across all areas in the United States. This pricing strategy for colors suggests that Verizon does not use the quality dimension to discriminate businesses. On the other hand, because of technological constraints, advertisement sizes cannot be offered on a continuous scale. The various color options can then be viewed as filling up the size scale.

In view of this empirical evidence, we construct a quality-adjusted quantity index. We consider the price schedule for multicolor displays and adjust the advertising sizes for other color options accordingly.[35] Using this schedule from the Yellow Page Association, we obtain

$$\widehat{\log(T)} = 4.1602 + 0.7317 \times \log(q) + 0.0062 \times (\log(q))^2, \tag{C.1}$$

where T is the price in dollars and q is the advertising size measured in square picas. The coefficients are estimated by ordinary least squares. The R^2 of such a regression is 0.999, which is an almost perfect fit.[36] Equation (C.1) gives the price schedule $T(\cdot)$.

The quality-adjusted quantity q is constructed as follows. For every business, its quality-adjusted quantity is obtained by solving (C.1) for q_i given its payment T_i. As an illustration, a one-page advertising display measuring 3,020 square picas with no color becomes equivalent to 1,470 picas in multicolor. Similarly, a half-page advertising display of 1,485 square picas with one color corresponds to a multicolor-adjusted quantity of 1,153 square picas. Table 3 provides some summary statistics on the prices paid and the quality-adjusted quantities bought by the 4,671 businesses. We note that an additional line to the standard listing is 1.8512 adjusted-quality square picas. Given that a standard listing is over one or two lines, the standard listing $\underline{q}$ offered at zero price is measured as $1.8512/2 = 0.9256$.

[35] Our choice of the multicolor display schedule is without loss of generality. We could have chosen another price schedule though the observed size range would be narrower.

[36] There is no endogeneity problem in this equation as q is not endogenous.

References

Andrews, D. (1994): "Empirical Process Methods 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.

Aryal, G. and **Y. Huang** (2009): "Competition and Nonlinear Pricing in Yellow Pages," Working Paper, Pennsylvania State University.

Aryal, G., **I. Perrigne** and **Q. Vuong** (2009): "Nonparametric Identification of Insurance Models with Multidimensional Screening," Working Paper, Pennsylvania State University.

Athey, S. and **P. Haile** (2007): "Nonparametric Approaches to Auctions," in J. Heckman and E. Leamer, eds., *Handbook of Econometrics, Volume VI*, Amsterdam: North Holland.

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.

Campo, S., **E. Guerre**, **I. Perrigne** and **Q. Vuong** (2011): "Semiparametric Estimation of First-Price Auctions with Risk Averse Bidders," *Review of Economic Studies*, forthcoming.

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. (2008): "Package Size and Price Discrimination in Paper Towels," *International Journal of Industrial Organization*, 26, 502-516.

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

Ekeland, I., **J. Heckman** and **L. Neishem** (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112, 60-109.

Galambos, J. (1978): *The Asymptotic Theory of Extreme Order Statistics,* New York: Wiley.

Gasmi, F., **J.J. Laffont** and **Q. Vuong** (1992): "Econometric Analysis of Collusive Behavior in a Soft-Drink Market," *Journal of Economics and Management Strategy*, 1, 277-311.

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

Guerre, E., **I. Perrigne** and **Q. Vuong** (2009): "Nonparametric Identification of Risk Aversion in First-Price Sealed-Bid Auctions Under Exclusion Restrictions," *Econometrica*, 77, 1193-1227.

Heckman, J., **R. Matzkin** and **L. Neishem** (2010): "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78, 1569-1591.

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

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.

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.

Matzkin, R. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339-1375.

McDevitt, R. (2010): "'A' Business by Any Other Name: Firm Name Choice as a Signal of Firm Quality," Working Paper, University of Rochester.

McManus, B. (2007): "Nonlinear Pricing in an Oligopoly Market: The Case of Specialty Coffee," *Rand Journal of Economics*, 38, 512-532.

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): "Identification and Estimation of Bidders' Risk Aversion in First-Price Auctions," *American Economic Review, Papers and Proceedings,* 97, 444-448.

Perrigne, I. and **Q. Vuong** (2011): "Nonparametric Identification of a Contract Model with Adverse Selection and Moral Hazard," *Econometrica*, forthcoming.

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

Riley, J. (2011): *Essential Microeconomics*, Cambridge University Press, forthcoming.

Roberts, M. and **L. Samuelson** (1988): "An Empirical Aanalysis of Dynamic Non-Price Competition in an Oligopolistic Industry," *Rand Journal of Economics*, 19, 200-220.

Rochet, J.C. and **P. Chone** (1998): "Ironing, Sweeping and Multidimensional Screening," *Econometrica*, 66, 783-826.

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.

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

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.

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 Discrimination and Tax Incidence-the Market for Gasoline and Diesel Cars in Europe," *Rand Journal of Economics*, 33, 275-97.

Wellner, J. (1978): "Limit Theorems for the Ratio of the Empirical Distribution Function to the True Distribution Function," *Probability Theory and Related Fields*, 45, 73-88.

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

Yellow Page Association (2008): "New Research Shows Overall Yellow Pages Usage Growing," http://www.ypassociation.org

Table 1. Revenue Ranking by Industry Headings

Industry heading	Revenue	Percentage
Attorneys	$512,587	8.52%
Dentists	$222,324	3.69%
Insurance	$188,302	3.13%
Storage household & commercial	$141,063	2.34%
Restaurants	$130,673	2.17%
Hotels	$130,655	2.17%
Physicians and Surgeons	$126,070	2.10%
Plumbing contractors	$101,840	1.69%
Carpet & rug cleaners	$99,861	1.66%
Veterinarians	$91,607	1.52%

Table 2: Number of Purchases and Revenue by Sizes

$Picas^2$	Percent of a page	# Purchases	Revenue	Percent
Listing				
12	0.4%	2,152	$0	0.0%
18	0.6%	844	$127,613	2.12%
24	0.79%	223	$22,478	0.37%
27	0.89%	802	$238,361	3.96%
30	0.99%	117	$29,333	0.49%
36	1.19%	127	$32,530	0.54%
39	1.29%	133	$48,899	0.81%
42	1.39%	31	$10,937	0.18%
45	1.49%	8	$3,721	0.06%
46	1.52%	12	$5,501	0.09%
48	1.59%	9	$3,617	0.06%
51	1.69%	45	$20,145	0.33%
54	1.79%	23	$12,423	0.21%
57	1.89%	1	$618	0.01%
58	1.92%	1	$517	0.01%
60	1.99%	4	$2,029	0.03%
63	2.09%	11	$5,638	0.09%
66	2.19%	15	$10,217	0.17%
72	2.38%	10	$7,415	0.12%
75	2.48%	1	$593	0.01%
78	2.58%	1	$808	0.01%
81	2.68%	10	$8,744	0.15%
84	2.78%	3	$2,671	0.04%
87	2.88%	3	$2,257	0.04%
93	3.08%	8	$6,774	0.11%
96	3.18%	2	$1,411	0.02%
99	3.28%	2	$2,207	0.04%
105	3.48%	2	$1,817	0.03%
108	3.58%	3	$3,635	0.06%
111	3.68%	1	$1,072	0.02%
114	3.77%	1	$958	0.02%
117	3.87%	3	$3,684	0.06%
120	3.97%	1	$934	0.02%
132	4.37%	1	$1,009	0.02%
135	4.47%	1	$1,754	0.03%
138	4.57%	1	$907	0.02%
144	4.77%	1	$3,176	0.05%
159	5.26%	1	$1,375	0.02%
162	5.36%	1	$1,590	0.03%
189	6.26%	1	$1,426	0.02%
192	6.36%	1	$2,218	0.04%
198	6.56%	1	$1,915	0.03%
225	7.45%	1	$2,713	0.05%
228	7.55%	1	$2,117	0.04%
282	9.34%	2	$5,572	0.09%
336	11.13%	1	$2,801	0.05%
Total		4,623	$648,127	10.77%

Table 2 (continued)

Picas^2	Percent of a page	# Purchases	Revenue	Percent
Space Listing				
54	1.79%	806	$406,224	6.75%
66	2.19%	88	$53,222	0.88%
72	2.38%	215	$167,814	2.79%
78	2.58%	24	$16,934	0.28%
81	2.68%	5	$4,350	0.07%
84	2.78%	26	$22,927	0.38%
90	2.98%	3	$2,419	0.04%
96	3.18%	11	$10,809	0.18%
99	3.28%	5	$5,735	0.10%
102	3.38%	1	$907	0.02%
108	3.58%	108	$154,351	2.56%
111	3.68%	2	$2,130	0.04%
120	3.97%	17	$23,940	0.40%
126	4.17%	6	$8,268	0.14%
132	4.37%	9	$13,759	0.23%
135	4.47%	2	$2,572	0.04%
144	4.77%	122	$149,089	2.48%
153	5.07%	1	$1,349	0.02%
156	5.17%	6	$7,651	0.13%
162	5.36%	4	$7,686	0.13%
168	5.56%	3	$6,440	0.11%
183	6.06%	1	$1,954	0.03%
189	6.26%	2	$3,770	0.06%
198	6.56%	1	$2,898	0.05%
204	6.75%	1	$1,626	0.03%
216	7.15%	7	$15,960	0.27%
228	7.55%	1	$2,180	0.04%
234	7.75%	1	$2,772	0.05%
243	8.05%	3	$9,188	0.15%
297	9.83%	1	$2,207	0.04%
324	10.73%	1	$6,731	0.11%
351	11.62%	2	$8,294	0.14%
612	20.26%	1	$6,376	0.11%
Total		1,486	$1,132,532	18.82%

Table 2 (continued)

Picas^2	Percent of a page	# Purchases	Revenue	Percent
Display				
174	5.76%	9	$18,182	0.30%
201	6.66%	137	$287,260	4.77%
208	6.90%	11	$27,192	0.45%
213	7.05%	11	$21,559	0.36%
220	7.29%	2	$5,443	0.09%
225	7.45%	6	$12,650	0.21%
235	7.79%	126	$320,074	5.32%
237	7.85%	2	$3,881	0.06%
246	8.15%	1	$2,180	0.04%
247	8.19%	14	$36,982	0.61%
259	8.58%	2	$5,645	0.09%
262	8.68%	1	$2,281	0.04%
271	8.98%	2	$5,381	0.09%
289	9.58%	2	$7,184	0.12%
301	9.98%	1	$4,082	0.07%
355	11.76%	6	$29,209	0.49%
379	12.56%	1	$3,276	0.05%
382	12.66%	135	$603,636	10.03%
391	12.96%	2	$9,715	0.16%
394	13.05%	19	$93,626	1.56%
406	13.45%	5	$21,500	0.36%
409	13.55%	2	$9,565	0.16%
421	13.95%	1	$3,541	0.06%
436	14.45%	3	$18,616	0.31%
517	17.13%	1	$6,743	0.11%
537	17.76%	2	$8,947	0.15%
544	18.02%	1	$5,860	0.10%
549	18.16%	3	$17,062	0.28%
562	18.62%	1	$7,523	0.13%
564	18.66%	70	$472,910	7.86%
576	19.06%	9	$64,538	1.07%
582	19.25%	1	$7,321	0.12%
588	19.45%	1	$7,346	0.12%
600	19.85%	2	$9,653	0.16%
621	20.55%	1	$8,190	0.14%
663	21.94%	1	$9,161	0.15%
735	24.34%	1	$5,872	0.10%
762	25.23%	24	$221,752	3.69%
774	25.63%	6	$50,512	0.84%
783	25.93%	1	$11,214	0.19%
786	26.03%	1	$10,458	0.17%
798	26.42%	1	$6,174	0.10%
978	32.38%	1	$13,385	0.22%

Table 2 (continued)

Picas^2	Percent of a page	# Purchases	Revenue	Percent
Display				
1137	37.65%	6	$82,877	1.38%
1485	49.17%	2	$35,280	0.59%
1512	50.07%	47	$759,594	12.62%
1521	50.36%	1	$15,990	0.27%
1524	50.46%	4	$56,248	0.93%
1536	50.86%	2	$36,061	0.60%
1566	51.85%	1	$10,748	0.18%
1593	52.75%	1	$19,039	0.32%
1623	53.74%	1	$19.090	0.32%
1794	59.40%	1	$21,017	0.35%
3020	99.98%	1	$32,395	0.54%
3047	100.88%	8	$261,298	4.34%
3071	101.67%	1	$32,597	0.54%
3074	101.77%	2	$62,737	1.04%
3083	102.07%	1	$33,076	0.55%
3119	103.26%	1	$33,378	0.55%
3182	105.35%	1	$34,733	0.58%
3275	108.43%	1	$37,667	0.63%
3311	109.62%	1	$34,008	0.57%
6066	200.86%	1	$60,380	1.00%
6147	203.54%	1	$61,478	1.02%
Total		714	$4,236,973	70.41%
Grand Total		6,823	$6,017,632	100.00%

Table 3: Some summary statistics

	Mean	Min	Max	STD
T	1,288.30	100.80	61,478.40	3,289.85
q	74.61	1.85	6,229.90	277.53
$\hat{v}_0$	16.07	0.59	38.72	10.14
$\hat{\theta}$	2.41	1.04	13.95	1.69
$\Delta\hat{U}$	883.94	3.99	30,708.00	2,370.80
$\Delta\hat{U}/(\Delta\hat{U}+T)$	0.295	0.039	0.478	0.127

Table 4: Some summary statistics on the three groups (the mean, min, max and standard deviation values are given for the businesses buying advertising)

	Number	Buying	Mean	Min	Max	STD
Group 1 $\hat{\theta}$	4,439	70.96%	2.49	1.04	13.95	1.80
Group 2 $\hat{\theta}$	1,858	67.71%	2.30	1.04	12.36	1.50
Group 3 $\hat{\theta}$	526	50.00%	1.96	1.04	8.37	1.10

Figure 1: $\hat{v}_0(\cdot)$

Figure 2: $\hat{\theta}(\cdot)$

Figure 3: Estimated Type Density

Figure 4: $\theta - [1 - \hat{F}(\theta)]/\hat{f}(\theta)$

Figure 5: Estimated Type Densities

Figure 6: Simulated $q_1(\cdot), q_2(\cdot), q_3(\cdot)$

Figure 7: Simulated $T_1(\cdot), T_2(\cdot), T_3(\cdot)$