NONLINEAR PRICING IN YELLOW PAGES

A Dissertation in
Economics
by
Yao Huang

© 2010 Yao Huang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2010
The dissertation of Yao Huang was reviewed and approved\textsuperscript{1} by the following:

Isabelle Perrigne  
Associate Professor of Economics  
Dissertation Co-advisor  
Co-Chair of Committee

Quang Vuong  
Professor of Economics  
Dissertation Co-advisor  
Co-Chair of Committee

Edward Green  
Professor of Economics

Mark J. Roberts  
Professor of Economics

Duncan K. H. Fong  
Professor of Marketing and Statistics

Barry W. Ickes  
Professor of Economics  
Acting Head of the Department of Economics

\textsuperscript{1}Signatures on file in the Graduate School.
CHAPTER 1: Nonlinear Pricing in Yellow Pages (with Isabelle Perrigne and Quang Vuong)

This paper analyzes nonlinear pricing in yellow page advertising. First, we develop a monopoly model that incorporates some features of the industry such as a 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 benefits of nonlinear pricing in presence of asymmetric information over other pricing rules.

CHAPTER 2: Competition and Nonlinear Pricing in Yellow Pages (with Gaurab Aryal)

This paper proposes a structural framework to analyze the nonlinear pricing strategies of two yellow page-advertising publishers. The data collected from the Yellow Page Association and the phone books suggest that the utility publisher is a leader in the market. Therefore, we consider a Stackelberg duopoly model of nonlinear pricing in which firms buying advertising are characterized by a bi-dimensional vector of tastes for the two directories. The model and the
econometric specification incorporate the features observed in the data such as the quadratic price schedules and a basic free advertisement offered to all firms by both publishers. Empirical results show substantial heterogeneity among firms’ willingness to pay. The estimated model is used to assess the welfare loss due to (i) asymmetric information, (ii) a merger between the two publishers and (iii) withdrawal of the non-utility publisher from the market.
# Table of Contents

List of Tables ........................................................................................................ vii
List of Figures ........................................................................................................ viii

Chapter 1. Nonlinear Pricing in Yellow Pages ....................................................... 1
  1.1 Introduction ..................................................................................................... 1
  1.2 Yellow Page Advertising Data ....................................................................... 5
  1.3 The Model ...................................................................................................... 9
  1.4 Nonparametric Identification and Estimation ............................................. 17
    1.4.1 Nonparametric Identification .................................................................. 17
    1.4.2 Nonparametric Estimation ...................................................................... 26
  1.5 Empirical Results ......................................................................................... 32
  1.6 Concluding Remarks ..................................................................................... 36
  1.7 Appendix A: Proofs in Section 1.3 ............................................................... 38
  1.8 Appendix B: Proofs in Section 1.4 ............................................................... 44
  1.9 Appendix C: Price Schedule ......................................................................... 56

Chapter 2. Competition and Nonlinear Pricing in Yellow Pages ....................... 68
  2.1 Introduction .................................................................................................... 68
  2.2 Nonlinear Pricing in Yellow Pages .............................................................. 73
    2.2.1 The Yellow Page Industry ..................................................................... 73
  2.3 The Model ..................................................................................................... 79
    2.3.1 The Demand Pattern ............................................................................ 81
    2.3.2 The Follower’s Problem ...................................................................... 82
    2.3.3 The leader’s problem .......................................................................... 89
  2.4 Identification and Estimation ....................................................................... 94
    2.4.1 Identification ....................................................................................... 94
    2.4.2 Estimation ........................................................................................... 101
  2.5 Empirical Results ....................................................................................... 105
  2.6 Concluding Remarks ................................................................................... 109
  2.7 Appendix A: Proofs .................................................................................... 112
  2.8 Appendix B: Merger .................................................................................. 119
  2.9 Appendix C: Exit ....................................................................................... 124
## List of Tables

1.1 Revenue Ranking by Industry Headings ........................................ 58  
1.2 Number of Purchases and Revenue by Sizes ................................ 59  
1.3 Summary Statistics of the Estimates ........................................... 62  
1.4 Examples of Estimated Firms’ Utility for No Color ......................... 63  
1.5 Summary Statistics in the Third Degree Discrimination ................... 63  

2.1 Comparison of Two Partial Price Schedules ................................. 127  
2.2 Number of Purchases and Revenue by Sizes ................................ 128  
2.3 Summary Statistics by Consumption Pattern ............................... 128  
2.4 Estimation of the Parameters .................................................. 128  
2.5 OLS for Effect of Competition ............................................... 129  
2.6 Incomplete Information vs. Complete Information ....................... 129  
2.7 Summary Statistics of (ˆθ₁, ˆθ₂) ................................................. 129
List of Figures

1.1 Function \( \hat{v}_0(q) \) ................................................. 64
1.2 Function \( \hat{\theta}(q) \) .................................................. 64
1.3 Estimated Untruncated Density of Type .................................. 65
1.4 Estimated \( \theta - \frac{1-F(\theta)}{f(\theta)} \) ............................................ 66
1.5 Optimal Price Schedules under Third Degree Discrimination .......... 66
1.6 Optimal Price Schedules under Third Degree Discrimination .......... 67
1.7 Optimal Purchases under Third Degree Discrimination ................. 67

2.1 Demand Pattern ................................................................ 130
2.2 Multiproduct nonlinear pricing. ............................................ 130
2.3 Joint Density of \( (z_1, z_2) \) ................................................. 131
Chapter 1

Nonlinear Pricing in Yellow Pages

1.1 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, cellular 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), Armstrong and Vickers (2001), Rochet and Stole (2003) and Stole (2007).\footnote{In these extensions, the optimal schedule becomes less tractable and closed form solutions can be obtained for some simple specifications only.}

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. This is equivalent to an optimal exclusion problem, i.e. defining 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 assume 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

\[2^2\text{This practice can be linked to two-sided markets/network of advertising. See Rysman (2004).}\]
payoff for buying advertising, the business type distribution and the publisher’s cost function. The econometric literature and the recent empirical industrial organization literature have documented that identification of models may or may not require particular functional forms.\textsuperscript{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 mainly individual advertising purchase data and the tariffs offered by the publisher.

Our identification problem is reminiscent of Ekeland, Heckman ans Nesheim (2004) and Heckman, Matzkin and Nesheim (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 (2010), we consider multiplicative separability of the marginal payoff function in the business type and the quantity of advertising purchased. Moreover, their identification result exploits 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 is more involved than in auctions. 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

\textsuperscript{3}See Laffont and Vuong (1976), Athey and Haile for surveys on the nonparametric identification of auction models.
distribution are identified above the truncation introduced by the optimal exclusion of firms. Indeed, 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 counting processes and kernel density estimator. 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 $\sqrt{N}$ consistent. The latter result relies on stochastic processes. Moreover, we also derive the asymptotic variance for such an estimator. The type density estimator follows the Stone (1982) optimal rate. What about $r()$? to be completed.

Next, we analyze a unique data set that we constructed from a phone directory in Pennsylvania and the Yellow Page Association data base. The data display a nonlinear pricing pattern as previously documented by Busse and Rysman (2005). This data set is unique because it contains (i) the full price schedule offered to businesses, (ii) individual data on price and quantity chosen by businesses and (iii) the whole population of businesses. The price schedule provides a large number of advertising categories and the pruchase data shows a large number of different price-quantity combinations chosen by firms. This allows us to treat the price schedule as

---

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

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

1.2 Yellow Page Advertising Data

We collected data on yellow page advertising in 2006 for State College in Central 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. After a careful check, the price schedule is strictly enforced. Defining each phone number as an observation, we collected a total of 6,823 advertisements over about 1,200 industry headings. The advertisements bought by businesses generate a revenue of 6 million dollars. Table 1.1 presents the top 10 industry headings that represent 29% of the total revenue. Not surprisingly, we find professionals such as attorneys, dentists, physicians, veterinarians, household services such as plumbers and carpet cleaners and hospitality services.

---

6 After a careful check, the price schedule is strictly enforced.
7 A similar ranking is obtained based on advertisement size.
Other sources of advertising are available to businesses. First, the area is covered by another phone directory from a non-utility publisher. This publisher distributes 72,000 copies of such a directory, which 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 first 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 so-called 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 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. Each advertising option is defined by a vector of characteristics which are the size, the color and some other special features.

In terms of size, the yellow page industry uses three categories, namely listing, space listing and display. Table 1.2 summarizes the various options chosen by businesses as well as the

\(^8\) 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 widespread in the industry over the past ten years. In 2008, the twelve 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 markets/networks of advertising.
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. 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 1.2 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 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.\footnote{Appendix C also justifies this procedure. Busse and Rysman (2005) consider only the price schedule for no color display advertising in their study.}

Regarding the special features, businesses can choose guide, anchor listing and trade marks.\footnote{Regarding the cover pages and the coupons, these options concern 22 businesses. The Yellow Page Association price schedules do not contain any information for these options. We have then excluded these options from the empirical analysis.} 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 very 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 advertisement, a half page advertisement and a double page advertisement, respectively. This corresponds to a reduction
of 66%. The same pattern is observed for other categories.\textsuperscript{13} This corresponds to an important discount for quantity with a notable curvature for the tariff.

1.3 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 $\theta \in [\underline{\theta}, \bar{\theta}]$ with $0 \leq \underline{\theta} < \bar{\theta} < \infty$. 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 $\theta_0$ 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.\textsuperscript{14} Moreover, as in Riley (2009), 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

$$U(q, \theta) = \int_0^q v(x, \theta)dx - T(q)$$ (1.1)

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 $\theta$ business willingness to pay for the $q$th unit of advertising. It is also the marginal payoff for buying the $q$th unit

\textsuperscript{13}A similar pattern can be seen in Table 1.2. As Table 1.2 mixes various categories for a given size, the results might be misleading.

\textsuperscript{14}We show later that $q_0 = \underline{q}$. 
of advertising or the inverse demand function for the business with type $\theta$. The type $\theta$ is distributed as $F(\cdot)$ with a continuous density $f(\cdot) > 0$ on $[\underline{\theta}, \overline{\theta}]$. The publisher does not know each business type but knows the distribution $F(\cdot)$.

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 $\forall q \geq 0, \forall \theta \in [\underline{\theta}, \overline{\theta}]$

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

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

(iii) $v_2(q, \theta) > 0$.\(^{17}\)

Assumption A1-(i) says that the marginal payoff is always nonnegative, while A1-(ii) says that the marginal payoff is decreasing in the quantity purchased. Assumption A1-(iii) says that businesses with a larger $\theta$ 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

---

\(^{15}\)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 $\pi$, which is a function of its type $\theta$, its own price and advertising as well as its competitors’ prices and advertising. Conditional on its own advertising quantity 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.

\(^{16}\)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.

\(^{17}\)Whenever a function has more than one variable, we denote its derivative with respect to the $k$th argument by a subscript $k$. 

\( q(\cdot) \) is a strictly increasing function on \((\theta_0, \theta]\). Later, we show that with additional assumptions the optimal \( q(\cdot) \) is strictly increasing.\(^{18}\) 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

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

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 \( \theta_0 \) do not pay for their advertising quantity \( q_0 \), they do not show up in the publisher’s revenue. On the other hand, this production has a cost that the publisher needs to take into account. This explains the argument of the cost function in two parts: (i) \( q_0 F(\theta_0) \) represents the total quantity provided to businesses choosing \( q_0 \) and (ii) \( \int_{\theta_0}^{\theta} q(\theta) f(\theta) d\theta \) 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 \( C'(q) > 0 \quad \forall q \geq q_0 \).

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 constraints, consider first a business with type \( \theta > \theta_0 \). It must prefer to buy \( q(\theta) \) rather than \( q_0 \), i.e.

\[
U(q(\theta), \theta) \geq \int_0^{\theta_0} v(x, \theta) dx \equiv U_0(\theta) \quad \forall \theta \in (\theta_0, \theta].
\]

\(^{18}\)This result is based on the complete sorting optimum in Maskin and Riley (1984, Proposition 4).
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 $\theta$ by A1-(iii) for any given $q$.

For the IC constraints, 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 $\tilde{\theta} \in (\theta_0, \theta_0)$, i.e. it must not pretend to be another type in $(\theta_0, \theta_0)$. Formally, let $U(\tilde{\theta}, \theta) \equiv U(q(\tilde{\theta}), \theta) \quad \forall \theta, \tilde{\theta} \in (\theta_0, \theta_0)$. This IC constraint can be written as $U(\theta, \theta) \geq U(\tilde{\theta}, \theta) \quad \forall \theta, \tilde{\theta} \in (\theta_0, \theta_0)$. The local first-order condition for the IC constraint to hold is

$$U_1(\theta, \theta) = 0 \quad \forall \theta \in (\theta_0, \theta_0). \tag{1.3}$$

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 $(\theta_0, \theta_0)$, (3) is equivalent to

$$v(q(\theta), \theta) = T'(q(\theta)) \quad \forall \theta \in (\theta_0, \theta_0). \tag{1.4}$$

Second, a business with type $\theta > \theta_0$ must prefer to buy $q(\theta)$ rather than $q(\tilde{\theta})$ for $\tilde{\theta} \in [\theta, \theta_0]$, i.e. it must not pretend to be another type in $[\theta, \theta_0]$. 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 $\tilde{\theta} \in [\theta, \theta_0]$, i.e. it must not pretend to be another type in $[\theta, \theta_0]$. 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 \( \tilde{\theta} \in (\theta_0, \bar{\theta}] \), i.e. it must not pretend to be another type in \((\theta_0, \bar{\theta}]\). Thus the IC constraint is \( U_0(\theta) \geq U(\tilde{\theta}, \theta) \) leading to \( T(q(\tilde{\theta})) \geq \int_{q_0}^{\tilde{q}} v(x, \theta)dx \) for \( \theta \leq \theta_0 \) and \( \tilde{\theta} > \theta_0 \). But, because \( v(x, \cdot) \) is increasing in \( \theta \) by A1-(iii), the latter is equivalent to \( T(q(\tilde{\theta})) \geq \int_{q_0}^{\tilde{q}} v(x, \theta_0)dx \) for \( \tilde{\theta} > \theta_0 \), which is true.\footnote{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 \( \int_{q_0}^{\tilde{q}} T'(x)dx \geq \int_{q_0}^{\tilde{q}} v(x, \theta_0)dx \). Letting \( T_+ = \lim q \downarrow q_0 T(q) \), it gives \( T(q) - T_+ \geq \int_{q_0}^{\tilde{q}} v(x, \theta_0)dx \) for \( q \geq q_0 \). Evaluating the latter at \( q(\tilde{\theta}) \) gives \( T(q(\tilde{\theta})) \geq \int_{q_0}^{\tilde{q}} v(x, \theta_0)dx + T_+ \). Hence, \( T(q(\tilde{\theta})) \geq \int_{q_0}^{\tilde{q}} v(x, \theta_0)dx \) for \( \tilde{\theta} > \theta_0 \) if \( T_+ \geq 0 \). We show later that \( T_+ = 0 \).}

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 \((\theta_0, \bar{\theta}]\), the local FOC (4) is sufficient for all the IC constraints to hold globally.

We can now solve the publisher’s optimization problem:

\[
\max_{q(\cdot), T(\cdot), \theta_0} \Pi = \int_{\theta_0}^{\tilde{\theta}} T(q(\theta))f(\theta)d\theta - C \left[ q_0 F(\theta_0) + \int_{\theta_0}^{\tilde{\theta}} q(\theta)f(\theta)d\theta \right], \tag{1.5}
\]

subject to the IR constraint (2), the IC constraint (4), where \( q(\cdot) = q_0 \geq \tilde{q} \) for \( \theta \leq \theta_0 \) and \( q(\theta) > q_0 \) for \( \theta > \theta_0 \) with \( q(\cdot) \) strictly increasing on \((\theta_0, \bar{\theta}]\). Because \( q_0 \) affects \( \Pi \) only through \( q_0 F(\theta_0) \) in the cost function, \( q_0 \) needs to be set at the minimum \( \tilde{q} \). As usual (see Tirole (1988, Chapter 3)), we eliminate \( T(\cdot) \) by \( \tilde{U}(\cdot) \) in the optimization problem, where \( \tilde{U}(\theta) = \int_{\tilde{q}}^{q(\theta)} v(x, \theta)dx - T(q(\theta)) \) for \( \theta \in (\theta_0, \bar{\theta}] \). Taking the derivative with respect to \( \theta \) gives \( \tilde{U}'(\theta) = [v(q(\theta), \theta) - T'(q(\theta)) q'(\theta) + \int_{\tilde{q}}^{q(\theta)} v_2(x, \theta)dx = \int_{\tilde{q}}^{q(\theta)} v_2(x, \theta)dx \), where the second equality uses (4). Let \( U_+ = \lim_{\theta \uparrow \theta_0} \tilde{U}(\theta) \).

\footnote{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 \( \int_{q_0}^{\tilde{q}} T'(x)dx \geq \int_{q_0}^{\tilde{q}} v(x, \theta_0)dx \). Letting \( T_+ = \lim q \downarrow q_0 T(q) \), it gives \( T(q) - T_+ \geq \int_{q_0}^{\tilde{q}} v(x, \theta_0)dx \) for \( q \geq q_0 \). Evaluating the latter at \( q(\tilde{\theta}) \) gives \( T(q(\tilde{\theta})) \geq \int_{q_0}^{\tilde{q}} v(x, \theta_0)dx + T_+ \). Hence, \( T(q(\tilde{\theta})) \geq \int_{q_0}^{\tilde{q}} v(x, \theta_0)dx \) for \( \tilde{\theta} > \theta_0 \) if \( T_+ \geq 0 \). We show later that \( T_+ = 0 \).}
Integrating the previous equation gives $\tilde{U}(\theta) = \int_{\theta_0}^{\theta} \left\{ \int_{q}^{q(u)} v_2(x, u) \right\} du + U_+$. Using the definition of $\tilde{U}(\theta)$, we obtain

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

(1.6)

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

$$\max_{q(\cdot), \theta_0, U_+} \Pi = \int_{\theta_0}^{\theta} \left\{ \int_{q}^{q(\theta)} v(x, \theta) dx \right\} f(\theta) d\theta - \int_{\theta_0}^{\theta} \left\{ \int_{q}^{q(u)} v_2(x, u) dx \right\} f(\theta) d\theta - U_+ \int_{\theta_0}^{\theta} f(\theta) d\theta - U_+ [1 - F(\theta_0)] - C \left[ qF(\theta_0) + \int_{\theta_0}^{\theta} q(\theta) f(\theta) d\theta \right].$$

The second term becomes

$$\int_{\theta_0}^{\theta} \left\{ \int_{q}^{q(\theta)} v_2(x, \theta) dx \right\} d\theta - \int_{\theta_0}^{\theta} \left\{ \left[ \int_{q}^{q(\theta)} v_2(x, \theta) dx \right] F(\theta) \right\} d\theta$$

by integration by parts. After rearranging terms and noting that $\tilde{U}(\cdot)$ appears through $U_+$ only, the firm’s problem becomes

$$\max_{q(\cdot), \theta_0, U_+} \Pi = \int_{\theta_0}^{\theta} \left\{ \int_{q}^{q(\theta)} v(x, \theta) dx \right\} f(\theta) - [1 - F(\theta)] \left[ \int_{q}^{q(\theta)} v_2(x, \theta) dx \right] d\theta - U_+ [1 - F(\theta_0)] - C \left[ qF(\theta_0) + \int_{\theta_0}^{\theta} q(\theta) f(\theta) d\theta \right].$$

(1.7)

Maximization of $\Pi$ with respect to $U_+$ gives trivially $U_+ = 0$. The optimal control problem is nonstandard because $\theta_0$ 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 $\theta \in [\theta, \theta]$, $v(q, \theta) - [(1 - F(\theta))v_2(q, \theta)/f(\theta)]$ is strictly monotone or identically equal to zero in $q$.

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

\[v(q, \theta) = C'(Q) + \frac{1 - F(\theta)}{f(\theta)} v_2(q, \theta) \quad \forall \theta \in (\theta_0, \theta] \quad (1.8)\]

\[
\lim_{\theta \downarrow \theta_0} q(\theta) = q \quad (1.9)
\]

\[T'(q) = v(q, \theta) \quad \forall \theta \in (\theta_0, \theta] \quad (1.10)\]

\[
\lim_{q \downarrow q} T(q) = 0, \quad (1.11)
\]

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

Conditions (8) and (9) characterize the optimal $q(\cdot)$ and the optimal cutoff type $\theta_0$. 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 $\theta_0$ for which there is no distortion.

When the cost function is nonlinear, the publisher considers only the marginal cost of the last unit of the total quantity produced $Q$. For the highest type, its marginal 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 $\theta_0$ and $q$, 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 $[q, +\infty) \times [\theta, \overline{\theta}]$ and $f(\cdot)$ is continuously differentiable on $[\theta, \overline{\theta}]$. Moreover, $\forall \theta \in [\theta, \overline{\theta}]$ and $\forall q \in [q, +\infty)$

(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 $\theta$.

These assumptions are quite 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. 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 $\theta$ increases, while A4-(iv) states that the hazard rate of the distribution of types does not decline too rapidly as $\theta$ increases. A large class of distribution functions satisfy the latter property.

**Lemma 2:** Under $A1$, $A2$, and $A_4$-(i,ii) or under $A1$, $A2$ and $A_4$-(i,iii,iv), $q(\cdot)$ is strictly increasing and continuously differentiable on $[\theta_0, \overline{\theta}]$ with $q'(\cdot) > 0$ on $[\theta_0, \overline{\theta}]$. Moreover, $T(\cdot)$ is strictly increasing and twice continuously differentiable on $[q, \overline{q}]$ with $T'(\cdot) > C'(Q)$ on $[q, \overline{q}]$ with $T'(q) = v(q, \theta_0)$ and $T'(\overline{q}) = C'(Q)$ where $\overline{q} \equiv q(\overline{\theta})$.

---

20In 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 $\alpha(\theta)$.

21These 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 (2010) 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.
Regarding the verification of the second-order conditions, we invite the reader to consult Maskin and Riley (1984). Tirole (1988) indicates that $T''(\cdot) < 0$, i.e. the price schedule is strictly concave in $q$.

1.4 Nonparametric Identification and Estimation

1.4.1 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 (marginal) 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 $[T(\cdot), q, F(\theta_0), G^*(\cdot), Q]$.\footnote{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.}

The function $G^*(\cdot)$ denotes the truncated (marginal) distribution of businesses purchases, i.e. $G^*(\cdot) = \Pr(q \leq \cdot)/\Pr(q > q)$. The data also provide some exogenous characteristics $Z$ of businesses such as their industry heading and a trademark affiliation. A natural question is how to introduce this observed heterogeneity in the model of Section 3. Such an introduction 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 to consider that $\theta$ is a scalar aggregation of the business’ observed and unobserved heterogeneity. Let $\epsilon$ denote the latter. Formally, we make the following assumption.
Assumption B1: The business’ type satisfies $\theta = r(Z, \epsilon)$ with $\epsilon$ distributed as $F_{\epsilon|Z}(\cdot|\cdot)$ given $Z$.

In particular, this allows dependence between the business’ type $\theta$ and characteristics $Z$. We also remark that $Z$ and $\epsilon$ do not need to 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 $\theta$ only. Hence, the cutoff type $\theta_0$ is independent of $Z$ by (9). The publisher’s optimization problem is still given by (5), while Proposition 1 and Lemma 2 hold.\footnote{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 $(\theta, z)$. The derivation of this model is left for future research.}

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), (9), (10) and (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 Nesheim (2004) and Heckman, Matzkin and Nesheim (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 Nesheim (2004) establish identification of the marginal payoff and the distribution of unobserved heterogeneity 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 addition, 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 (2010) 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 $\theta$.

**Assumption B2:** The businesses’ marginal payoff function is of the form

$$v(q, \theta) = \theta v_0(q),$$

(1.12)

where $v_0(\cdot)$ satisfies $v_0(q) > 0$, and $v_0'(q) < 0$ for $q \geq q_0$ and for all $\theta \in [\theta_0, \theta_1] \subset (0, +\infty)$.\(^{24}\)

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

$$\theta v_0(q) = C'(Q) + \frac{1 - F(\theta)}{f(\theta)} v_0(q) \quad \forall q \in (q_0, q_1]$$

(1.13)

$$T'(q) = \theta v_0(q) \quad \forall q \in (q_0, q_1],$$

(1.14)

\(^{24}\)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, p156) that the price schedule is strictly concave in $q$ when A4-(iv) is strengthened to a hazard rate $\rho(\theta)$ increasing in $\theta$.  


where $\theta = q^{-1}(q)$ and $\overline{\theta} = q(\overline{\theta})$ since $q(\cdot)$ is strictly increasing by Lemma 2. The one-to-one mapping between the unobserved type $\theta$ and the observed advertising quantity $q$ is the key of our identification result. Following B2, the model structure becomes $[v_0(\cdot), r(\cdot, \cdot), F_{\epsilon|Z}(\cdot), C(\cdot)]$.

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 $F_{\epsilon|Z}(\cdot)$ 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).

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 the total amount produced. The next lemma formalizes this result.

**Lemma 3:** The cost function is not identified except for the marginal cost at the total amount produced, which satisfies $C'(Q) = T'(\overline{q})$.

This follows immediately from Lemma 2. Since we observe both the price schedule $T(\cdot)$ and $\overline{q}$, we are able to recover $C'(Q)$, i.e. the latter is identified. Identification of the cost function could be improved by using data from different markets providing different values for $Q$.

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 identification argument relies on the availability of some continuous exogenous variables that are independent of the term of unobserved heterogeneity. In our case, the only variables that are

---

25 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.

26 Their separability assumption is also different from ours as they assume $v(q, z, \epsilon) = \epsilon v_0(q, z)$, while we have $v(q, z, \epsilon) = r(z, \epsilon) v_0(q)$. 
available are discrete. Moreover, these variables are likely to be correlated with the business’
taste $\theta$ 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 $\theta$ and the base marginal payoff $v_0(\cdot)$ are unknown. The next lemma
formalizes this result.

**Lemma 4:** Consider a structure $S = [v_0(\cdot), F(\cdot), C(\cdot)] \in S$. 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, $\tilde{S} \in S$ and the
two structures $S$ and $\tilde{S}$ lead to the same set of observables $[T(\cdot), q, F(\theta_0), G^*(\cdot), Q]$, i.e. the two
structures are observationally equivalent.

Several scale normalizations can be entertained. Three natural choices are to fix $\theta$, $\theta_0$ or $\theta$.
Before discussing the appropriate choice of normalization, we establish Lemma 5, in which the
business marginal payoff $v_0(\cdot)$ and its unobserved type $\theta$ are expressed as functions of the quantity
purchased $q$ and other observables $[T(\cdot), q, F(\theta_0), G^*(q), Q]$. Based on Lemma 5, the choice of
normalization and the nonparametric identification of $[v_0(\cdot), F(\cdot)]$ are readily established.

**Lemma 5:** Let $[v_0(\cdot), F(\cdot), C(\cdot)] \in S$. Denote $\gamma \equiv C'(Q)$ and $\theta(\cdot) \equiv q^{-1}(\cdot)$. The necessary
conditions (13) and (14) are equivalent to

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

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

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

(1.16)
with $\xi(q) = 1$ and $\xi(\bar{q}) = \lim_{q \to \bar{q}} \xi(q) = \bar{\theta}/\theta_0$.

The proof of Lemma 5 exploits the one-to-one mapping between the advertising quantity $q$ and the business’ type $\theta \in [\theta_0, \bar{\theta}]$ obtained from Lemma 2. For each $q \in (q, \bar{q}]$, we observe the truncated marginal distribution $G^*(q) = \Pr(\bar{q} \leq q | \bar{q} > q) = \Pr(\bar{\theta} \leq \theta(q) | \bar{\theta} > \theta(q)) = [F(\theta) - F(\theta_0)]/[1 - F(\theta_0)]$ 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 $\theta$ in terms of observables, which are the corresponding quantity $q$, the truncated quantity distribution $G^*(\cdot)$, its density $g^*(\cdot)$ and the price schedule $T(\cdot)$ as well as $\gamma = C'(Q)$, which is identified by Lemma 3. There is a clear parallel here with auction models. Guerre, Perrigne and Vuong (2000) use the 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 $\theta$ 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 in addition to the distribution of businesses’ type $F(\cdot)$ to recover, i.e. the base marginal payoff $v_0(\cdot)$. To this end, we also exploit the relationship between the shape of the price schedule $T(\cdot)$ and the distribution of the unobserved 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 the following subsection.

In view of Lemma 5, a natural normalization is $\theta_0 = 1$. The base marginal payoff $v_0(\cdot)$ can be uniquely recovered on $(q, \bar{q}]$ from the observables $T'(\cdot)$, $T''(\cdot)$, $G^*(\cdot)$ and $\gamma = T'(\bar{q})$ and
hence at $\bar{q}$ by continuity of $v_0(\cdot)$. Similarly, the truncated marginal type distribution $F^*(\cdot) \equiv [F(\cdot) - F(\theta_0)]/[1 - F(\theta_0)]$ can be uniquely recovered on $[\theta_0, \bar{\theta}]$ 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 $[v_0(\cdot), F(\cdot), C(\cdot)] \in S$. Under B2-B3, the marginal payoff $v_0(\cdot)$ and the truncated business type distribution $F^*(\cdot)$ are identified on $[\underline{q}, \bar{q}]$ and $[\theta_0, \bar{\theta}]$, respectively. In particular, $v_0(q) = T'(q)$ and $\bar{\theta} = \lim_{q \uparrow \bar{q}} \xi(q)$.

We note that $F(\cdot)$ can be recovered from $F^*(\cdot)$ on $[\theta_0, \bar{\theta}]$ 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 $[\theta, \theta_0)$, 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 $\underline{q}$. Here again, we can make a parallel to auctions, in which a binding reserve price does not allow to identify the distribution of bidders’ private values for private values below the reserve price. See Guerre, Perrigne and Vuong (2000).

As indicated previously, alternative normalizations can be entertained. For instance, the normalization $\bar{\theta} = 1$ also allows to identify $v_0(\cdot)$ and $F^*(\cdot)$ on $[\underline{q}, \bar{q}]$ and $[\theta_0, \bar{\theta}]$, respectively. In this case, (15) and (16) give

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

$$\theta(q) = \bar{\theta} \left[1 - G^*(q)\right]^{\frac{\gamma}{\bar{\theta}(\gamma - 1)}} \exp\left\{-\gamma \int_q^{\bar{q}} \frac{T''(x)}{T'(x)^2} \log \left[1 - G^*(x)\right] dx\right\},$$
for all \( q \in (\underline{q}, \overline{q}] \). Therefore, if \( \tilde{\theta} = 1 \), 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 \( F_{\epsilon | \bar{Z}}(\cdot | \cdot) \) appearing in B1. Note that the first step identifies not only the truncated distribution \( F^*(\cdot) \) of \( \theta \) by Proposition 2 but also recovers the type \( \theta \) for each business from its purchase \( q > \underline{q} \) by (15) and (16). On the other hand, under B1, \( \theta = r(z, \epsilon) \). Since \( \theta \) 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 \( F_{\epsilon | \bar{Z}}(\cdot | \cdot) \) extends Matzkin (2003) results to a censored nonseparable model under the following assumption. Consider the partition \( \bar{Z} = (Z_1, Z_2) \). Let \( S_{Z_1} \) and \( S_{Z_2|z_1} \) be the supports of \( Z_1 \) and \( Z_2 \) given \( Z_1 = z_1 \), respectively. Let also \( S_{\epsilon, Z_2|z_1} \) be the support of \( (\epsilon, Z_2) \) given \( Z_1 = z_1 \).

**Assumption B4:** We have:

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

(ii) There exists a known value \( z_1^0 \in S_{Z_1} \) such that \( S_{Z_2|z_1^0} = S_{Z_2} \) and \( r(z_1^0, z_2, \epsilon) = \epsilon \) for all \( (\epsilon, z_2) \in S_{\epsilon, Z_2|z_1^0} \).

(iii) For every \( (z_1, z_2) \in S_{Z_1, Z_2} \), 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 \( S_{\epsilon|z_1, z_2} \).

Assumption B4-(i) is weaker than the independence between \( \epsilon \) and \( Z \). Assumption B4-(ii) follows Matzkin (2003) first normalization. In particular the first part of (ii) is a full support requirement on \( z_1^0 \). Assumption B4-(iii) underlies the use of quantiles for identification. For every \( z \in S_Z \), 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

\[
\begin{align*}
  r(z_1, z_2, e) &= F_{\theta|Z_1, Z_2}^{s-1}[F_{\epsilon|Z_1, Z_2}^s(e|z_1, z_2)|z_1, z_2] \\
  F_{\epsilon|Z_1, Z_2}(e|z_1, z_2) &= \Pr[\hat{\theta} = \theta_0|Z_1 = z_1^0, Z_2 = z_2] \\
  &\quad \quad + (1 - \Pr[\hat{\theta} = \theta_0|Z_1 = z_1^0, Z_2 = z_2])F_{\theta|Z_1, Z_2}^s(e|z_1^0, z_2),
\end{align*}
\]

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

\[
F_{\epsilon|Z_1, Z_2}^s(e|z_1, z_2) = 1 - \frac{1 - \Pr[\hat{\theta} = \theta_0|Z_1 = z_1^0, Z_2 = z_2]}{1 - \Pr[\hat{\theta} = \theta_0|Z_1 = z_1, Z_2 = z_2]}[1 - F_{\theta|Z_1, Z_2}^s(e|z_1^0, z_2)].
\]

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 \( \Pr[\hat{\theta} = \theta_0|Z_1 = z_1, Z_2 = z_2] \) 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 \( F_{\epsilon|Z_1, Z_2}^s(e|z_1, z_2) \) is the distribution of businesses types conditional on advertising purchase and characteristics \((z_1, z_2)\). Since \( \theta \) 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 \( \theta \) given \( Z \). Second, there is an adjustment due to the censoring as \( \Pr[\hat{\theta} = \theta_0|Z = (z_1^0, z_2)] \neq \Pr[\hat{\theta} = \theta_0|Z = (z_1, z_2)] \). Third, \( r(z, e) \) is identified only for \( e \geq \max[e(z), \theta_0] \). Lastly, the term \( \theta_0 \) in the truncated supports of \( r(\cdot, \cdot) \) and \( F_{\epsilon|Z}(\cdot|\cdot) \) arises because \( \theta_0 = r(z_1^0, z_2, e(z_1^0, z_2)) = e(z_1^0, z_2) \) by B4-(ii).
1.4.2 Nonparametric Estimation

In view of B2, (15) provides expressions for the marginal payoff $v_0(\cdot)$ and the business type $\theta(\cdot)$ as functions of $T'(\cdot)$ and $\xi(\cdot)$. The latter depends on $T'(\cdot)$, $T''(\cdot)$, $\gamma$ 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 proceed with 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 density of the business’ type from which we can easily estimate the corresponding density using the observed proportion of firms choosing $\overline{q}$. Because we use data from a single market, our estimation procedure is not performed conditionally upon some exogenous variables $Z$ 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 $Z$ 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 $\overline{q}$, while $q_i, i = 1, 2, ..., N^*$ denotes the quantity purchased by each of those firms. Following (16) we estimate $\xi(\cdot)$ by

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

(1.17)
where \( q_{\text{max}} = \max_{i=1,\ldots,N^*} q_i \), \( \hat{\gamma} = T'(q_{\text{max}}) \), \( \hat{G}^*(\cdot) \) is the empirical distribution

\[
\hat{G}^*(q) = \frac{1}{N^*} \sum_{i=1}^{N^*} \mathbb{1}(q_i \leq q), \quad \text{for } q \in [\underline{q}, \bar{q}],
\]

(1.18)

and \( \mathbb{1}(\cdot) \) is the indicator function. In particular, \( q_{\text{max}} \leq \bar{q} \).

As a matter of fact, for \( q \in [\underline{q}, q_{\text{max}}] \), \( \hat{\xi}(q) \) 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 \( (\underline{q}, \bar{q}) \), the integral in (17) can be written as the finite sum of integrals from \( q^j \) to \( q^{j+1} \). On each of these intervals, \( \log[1 - \hat{G}^*(\cdot)] \) is constant, while the primitive of \( -T''(\cdot)/T'(\cdot) \) is \( 1/T'(\cdot) \). Thus, for \( q \in [\underline{q}, q_{\text{max}}] \), we have

\[
\hat{\xi}(q) = \left[ 1 - \hat{G}^*(q^j) \right]^{\hat{\gamma}/\bar{q}(\cdot)} \exp \left\{ \hat{\gamma} \sum_{t=0}^{j-1} \left[ \frac{1}{T'(q^t)} - \frac{1}{T'(q^{t+1})} \right] \log(1 - \hat{G}^*(q^t)) + \hat{\gamma} \frac{1}{T'(q^j)} \log(1 - \hat{G}^*(q^j)) \right\}
\]

(1.19)

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

\[
\lim_{q \uparrow q_{\text{max}}} \hat{\xi}(q) = \exp \left\{ \hat{\gamma} \sum_{t=0}^{J-1} \left[ \frac{1}{T'(q^t)} - \frac{1}{T'(q^{t+1})} \right] \log(1 - \hat{G}^*(q^t)) \right\}, \quad (1.20)
\]

which is finite.\(^{27}\) Thus, \( \hat{\xi}(\cdot) \) is a well defined strictly positive \textit{cadlag} (continue à droite, limites à gauche) function on \([\underline{q}, \bar{q}]\) with steps at \( q^1 < q^2 < \ldots < q^{J-1} \).

\(^{27}\)When \( q \uparrow q_{\text{max}} \), then \( q \in [q^{J-1}, q^J) \). Thus \( \hat{G}^*(q) = \hat{G}^*(q^{J-1}) \leq (N^* - 1)/N^* \) and \( \lim_{q \uparrow q_{\text{max}}} [1 - \hat{G}^*(q)]^{\gamma/T'(q)} = 1 \) as \( \lim_{q \uparrow q_{\text{max}}} T'(q) = \hat{\gamma} \).
Moreover, the highest type $\bar{\theta}$ is estimated by $\theta_{\text{max}} = \hat{\xi}(q_{\text{max}})$, which is given by (20). To complete the first step, following (15) and using B2 we estimate $v_0(\cdot)$ and $\theta(\cdot)$ by

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

(1.21)

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 $\theta_{\text{max}} > \theta_0 = 1$, while $\hat{v}_0(\cdot)$ and $\hat{\theta}(\cdot)$ are strictly decreasing and increasing on $[\underline{q}, q_{\text{max}}]$, 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 density of business’ type from the pseudo sample by using the kernel estimator

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

(1.22)

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.

We make the following assumption on the data generating process.

**Assumption B5:** The observed and unobserved heterogeneity $(Z_i, \epsilon_i), i = 1, \ldots, N$, where $N$ is the total number of businesses, 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 $q_i = q(\theta_i) \mathbb{1}(\theta_i > \theta_0) + q \mathbb{1}(\theta_i \leq \theta_0)$, the observed advertising quantities $q_i$ are i.i.d.

The next lemma establishes the strong consistency of $\hat{\gamma} = T'(q_{\text{max}})$ for the marginal cost $\gamma = C'(Q)$ with a rate of convergence faster than $\sqrt{N}$. It also provides the asymptotic distribution of
These properties follow from the delta method combined with known asymptotic properties of the highest order statistic $q_{\text{max}}$ 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) $N(\hat{\gamma} - \gamma) \xrightarrow{D} -T''(\bar{q})E/\left[g^*(\bar{q})(1 - F(\theta_0))\right]$, where $E$ is standard exponential distributed, as $N \to \infty$.

Following Campo et al. (2009), $g^*(\bar{q})$ can be estimated consistently by a one-sided kernel density estimator $\hat{g}^*(\cdot)$ evaluated at $q_{\text{max}}$, while $T''(q_{\text{max}})$ and $1 - F(\theta_0)$ can be estimated as usual by $T''(q_{\text{max}})$ and $N^*/N$, respectively. Thus, $N(\hat{\gamma} - \gamma)\hat{g}^*(q_{\text{max}})/T''(q_{\text{max}}) \xrightarrow{D} -E$, which can be used for hypothesis testing.

The next proposition establishes the asymptotic properties of $\hat{v}_0(\cdot)$. Following the empirical process literature introduced in econometrics by Andrews (1994), we view $\hat{v}_0(\cdot)$ as a stochastic process defined on $[q, \bar{q}]$ and hence as a random element of the space $D[q, \bar{q}]$ of cadlag functions on $[q, \bar{q}]$. Because of the singularity of $\xi(\cdot)$ at $\bar{q}$, we consider instead the space $D[q, q_\dagger]$ with its uniform metric $\|\psi_1 - \psi_2\|_1 = \sup_{q \in [q, q_\dagger]}|\psi_1(q) - \psi_2(q)|$, where $q_\dagger \in (q, \bar{q})$.\footnote{As usual measurability issues are ignored below. This can be addressed by considering either the projection $\sigma$-field on $D[q, q_\dagger]$ 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).}

**Proposition 4:** Under $A1$, $A2$, $A4$, $B1$–$B3$, for any $q_\dagger \in (q, \bar{q})$, we have as $N \to \infty$

(i) $\|\hat{v}_0(\cdot) - v_0(\cdot)\|_1 \xrightarrow{a.s.} 0$ and $\hat{v}_0(q_{\text{max}}) \xrightarrow{a.s.} v_0(\bar{q})$,

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

$$Z(\cdot) = \left[1 - \frac{\gamma}{T'(\cdot)}\right] \frac{B_{G^*}(\cdot)}{1 - G^*(\cdot)} - \gamma \int_q^{T''(x)} \frac{B_{G^*}(x)}{1 - G^*(x)} dx,$$

(1.23)
where $B_{G^*}()$ is the $G^*$-Brownian bridge process on $[q, \bar{q}]$.\(^{29}\)

The first part establishes the uniform almost sure convergence of $\hat{v}_0(\cdot)$ on any subset $[q, q_t]$ with $q_t < \bar{q}$. It also establishes the strong consistency of $\hat{v}_0(q_{\text{max}})$ for $v_0(\bar{q})$. The second part gives the asymptotic distribution of $\hat{v}_0(\cdot)$. It is worth noting that its rate of convergence is the parametric rate $\sqrt{N}$. This remarkable result comes from that $\hat{v}_0(\cdot)$ is a functional of the empirical cdf $\hat{G}^*(\cdot)$. The appendix shows that $Z(\cdot)$ has zero mean and covariance $E[Z(q)Z(q')] = \omega(q)$ for $q \leq q \leq q' \leq q_t$, where

\[
\omega(q) = \left(1 - \frac{\gamma}{T'(q)}\right)^2 \frac{G^*(q)}{1 - G^*(q)} - 2\gamma \int_q^q \left(1 - \frac{\gamma}{T'(x)}\right) \frac{T''(x)}{T'(x)^2} \frac{G^*(x)}{1 - G^*(x)} dx. \tag{1.24}
\]

Note that the covariance $E[Z(q)Z(q')]$ 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 $q \leq q \leq q' \leq q_t$. In particular, Proposition 3-(ii) implies that

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

for every $q \in [q, q_t]$ and hence for every $q \in [q, \bar{q}]$. The asymptotic variance of $\hat{v}(q)$ vanishes at $q = q$ as $\omega(q) = 0$, which is expected since $\hat{v}_0(q) = T'(q)$, while $\omega(q)$ increases as $q$ increases to $\bar{q}$ 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 $v_0(q)$, $1 - F(\theta_0)$

\(^{29}\)The $G^*$-Brownian bridge process on $[q, \bar{q}]$ is the limit of the empirical process $\left(1/\sqrt{N}\right) \sum_i \{I(q_i \leq \cdot) - G^*(\cdot)\}$ indexed by $[q, \bar{q}]$. 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 $q \leq q \leq q' \leq \bar{q}$.
and \( \omega(q) \) by \( \hat{\omega}_0(q) \), \( N^* / N \) and \( \hat{\omega}(q) \), respectively where \( \hat{\omega}(q) \) is obtained from (24) by replacing \( \gamma \) and \( G^*(\cdot) \) by their estimates \( T'(q_{\max}) \) and \( \hat{G}^*(\cdot) \).

We now turn to the second step. We have the following corollary.

**Corollary 1:** Under A1, A2, A4, B1–B3, for any \( q^\dagger \in (q, \bar{q}) \), we have as \( N \to \infty \)

(i) \( \| \hat{\theta}(\cdot) - \theta(\cdot) \|_1 \overset{a.s.}{\to} 0 \) and \( \theta(q_{\max}) \overset{a.s.}{\to} \bar{\theta} \),

(ii) as random functions in \( D[q, q^\dagger] \), \( \sqrt{N}[\hat{\theta}(\cdot) - \theta(\cdot)] \Rightarrow \theta(\cdot)Z(\cdot) / \sqrt{1 - F(\theta_0)} \).

In particular, the pseudo businesses’ types \( \hat{\theta}_i \) obtained in the first step converge uniformly to the businesses’ types \( \theta_i \) at the rate \( \sqrt{N^*/\log \log N^*} \), which is the uniform rate of convergence of \( \sup_{q \in [q, q^\dagger]} | \hat{G}^*(q) - G^*(q) | \) 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 \( \theta_i \) does not affect the second step. Consequently, the standard kernel estimator (22), which uses the pseudo businesses’ types, possesses the standard asymptotic properties of uniform convergence and limiting distribution, namely

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

(ii) \( \sqrt{N^* h}[\hat{f}(\theta) - f(\theta)] \overset{d}{\to} N(0, f(\theta) \int K(x)^2 dx) \) for every \( \theta \in (\theta_0, \bar{\theta}) \) provided \( N^* h^5 \to 0 \),

as \( f(\cdot) \) is twice continuously differentiable and bounded away from zero on \( [\theta_0, \bar{\theta}] \). 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’

\(^{30}\)Recalling that \( N^* / N \overset{a.s.}{\to} 1 - F(\theta_0) \), Proposition 3-(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 \( q \in [q, q^\dagger] \), where \( c \) is the constant defined by \( \Pr (|Z(q)|_1 \leq c) = 1 - \alpha \) with \( 0 < \alpha < 1 \) and \( Z(\cdot) \) is the Gaussian process introduced above.
types $\theta_i$ are observed can be achieved by our two-step procedure when the $\theta_i$ are unobserved but firm’s consumption choices $q_i$ are observed.

### 1.5 Empirical Results

Using the quality-adjusted quantities derived from (C.1), the price schedule and the firm’s purchases, we apply the estimators (18) to (21), where $q_{\text{max}} = 6,230$. 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. It means that an additional line of listing or 1.85 quality-adjusted square picas costs at the margin for the publisher $\$15.34$, which can be compared to $\$100.80$ charged to businesses. Data from a single phone directory do not allow us to recover more of the cost function as shown in Section 4.1. If data from several phone directories published by the same company are collected satisfying some criteria of homogeneity such as population size, we could recover the cost function at several values of $Q$. We can then hope to identify $C(\cdot)$ partially.

The estimated marginal 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 have been obtained using the estimators (17) and (21). Figure 1.1 shows that $\hat{v}_0(\cdot)$ is strictly decreasing as assumed in B1, while Figure 1.2 shows that $\hat{\theta}(\cdot)$ is strictly increasing as predicted by the model. Using (22) with a triweight kernel and a rule-of-thumb bandwidth, the estimated (truncated) density of types $\hat{f}^*(\cdot)$ on $[1, \theta_{\text{max}}] = [1, 13.95]$ is displayed in Figure 1.3. The estimated density reveals substantial heterogeneity among businesses in their taste for advertising. We note that the density displays three modes, a first one around 1.3, a second one around 2.3 and a third one around 6. The first mode

---

$^{31}$To obtain smooth representations, we have used the integral of a kernel density estimator with a triweight kernel and a rule-of-thumb bandwidth instead of (18).
corresponds to the businesses adding a few more lines and/or a larger font to their standard listing. The second mode corresponds to the businesses choosing the smallest size of space, while the third mode corresponds to the businesses choosing the smallest size of display. The shape of the type density shows other irregularities such as little bumps at 7 and 9 corresponding to firms choosing displays between one fifth and one quarter of a page. Figure 1.4 displays \( \theta - [(1 - E(\theta))/\hat{f}(\theta)] \), which is strictly increasing thereby satisfying A4-(iv).

With the estimates of \( v_0(\cdot) \) and \( \theta(\cdot) \), we can assess the payoff or informational rent for each business using \( U(q_i, \theta_i) = U_0(\theta_i) + \theta_i \int_{q}^{\theta} v_0(x) dx - T(q_i), i = 1, \ldots, 4,671 \). The term \( U_0(\theta) = \theta \int_{0}^{\theta} v_0(x) dx \) is not identified since \( v_0(\cdot) \) is not identified on \([0,q]\). On the other hand, we can estimate the additional rent for advertising beyond \( q \) for each firm by \( \Delta \hat{U}_i \equiv \hat{U}(q_i, \hat{\theta}_i) - U_0(\hat{\theta}_i) = \hat{\theta}_i \int_{q}^{\hat{\theta}_i} \hat{v}_0(x) dx - T(q_i) \). As expected, this additional rent strictly increases with the quality-adjusted advertising size purchased. Table 1.3 gives summary statistics on these values 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 1.3 also gives this additional rent as a proportion of the gross additional rent for buying advertising \( \Delta \hat{U} + T \). We note that this proportion is quite large suggesting that informational rents left to businesses are quite substantial. Given that the term \( U_0 \) is not included, our results provide a lower bound for the informational rent left to businesses. Overall, the sum of the \( \Delta \hat{U}_i \) is $3,448,307.85, which needs to be compared to the total payment made by the businesses $6,017,632. This gives a lower bound for the overall informational rent of the order of 36%. In particular, in a world of complete information, the publisher would know each business’ type and the price that the publisher charges to each business would be equal to each business’ benefit, i.e. \( \theta \int_{q}^{\theta} v_0(x) dx \) leaving the
minimal rent $U_0(\theta)$ to the business. Thus, given the minimum quantity $q$, the proportion 61\% measures the extent of the cost of asymmetric information for the publisher.

In view of Proposition 3, we could estimate $r(\cdot, \cdot)$ and $F_{\theta|Z}(\cdot|\cdot)$. The available business’ characteristics are the business heading and the corresponding number of businesses in the heading. 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 the variation between headings.\(^{32}\) This indicates that the variability of the businesses types is mainly due to idiosyncratic factors. This result is confirmed when considering a Tobit model due to the censoring on $\theta$ as explained previously. With all the headings as dummy variables one obtains a pseudo $R^2$ of 0.1163 with only 180 heading dummies among 1193 that are significant at the 10\% level. With only the number of businesses, one obtains a pseudo $R^2$ of 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 $\$3,112,052$. 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 $\theta > \theta^L_0$, where $\theta^L_0$ is the threshold

\(^{32}\)We obtain a similar result when decomposing the total variance of the advertising quantity purchased.
level associated with a linear price and given by $\theta^L_0 v_0(q) = p$, i.e. $\theta^L_0 = p/v_0(q)$. Equivalently, given $p$ we have $q = v^{-1}_0(p/\theta)$ for $\theta > p/v_0(q)$ and $q = q$ for $\theta \leq p/v_0(q)$. Using the estimate of $v_0(\cdot)$, we can estimate the publisher’s expected profit by

$$p \int_{p/v_0(q)}^{\hat{\theta}} \hat{v}_0^{-1}(p/\theta) \hat{f}(\theta)d\theta - \hat{\gamma} \left[ q\hat{F}(p/\hat{v}_0(q)) + \int_{p/\hat{v}_0(q)}^{\hat{\theta}} \hat{v}_0^{-1}(p/\theta) \hat{f}(\theta)d\theta \right],$$

for every given $p \geq \hat{v}_0(q)$ since we identify $F(\cdot)$ and $v_0(\cdot)$ on $[1, \theta_{max}]$ and $[q, q_{max}]$, respectively.

From Table 1.3, the lowest value for $p$ that we can consider is $47.81$ per quality-adjusted pica square. The largest expected profit is $1,739,502$, which is obtained for $p = 47.81$. Thus, this value is a lower bound for the profit under an optimal linear price $p \leq 47.81$. 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, the businesses with a large $\theta$ value will buy lower quantities than under the current nonlinear price because these large buyers will not benefit for a discount for large quantities. The highest quantity bought under the linear prices is $159.2$ in comparison with $6229.9$ under the nonlinear pricing schedule. The total rent for the firms is $997,884.23$, which is also at a loss comparing to that under nonlinear prices.

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 2.4 provides summary statistics on the estimated $\theta$, the number of businesses and the proportion of businesses buying advertising for each group. It can be seen that businesses with households as primary customers in average
have higher estimated $\theta$ than other two groups, and businesses with businesses as primary customers in average have higher estimated $\theta$ than non-profit organizations. Figure 1.5 shows the first-order stochastic dominance among the distributions of the estimated $\theta$ for the three groups. Suppose that the publisher is allowed to offer different nonlinear price schedules for different groups, each optimal price schedule will satisfy the conditions in Proposition 1. We first estimate the distribution of $\hat{\theta}$ for each group, and according to the conditions in Proposition 1 and the estimated function $v_0(\cdot)$, we simulate the optimal price schedules for the three groups and the optimal purchases under these prices. Figure 1.6 displays the optimal price schedules for the three groups and the current schedule. Figure 1.7 shows the optimal purchases under the different price schedules.

1.6 Concluding Remarks

This paper studies the nonparametric identification of the nonlinear pricing model. Such a result 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 of advertising. This result is in the spirit of the recent literature on the nonparametric identification of incomplete information models such as first-price auctions though our identification result is more involved as additional functions need to be identified. Following the nonparametric identification, a nonparametric two-step estimation procedure is developed to estimate the model structure. Asymptotic properties of our estimator are established. Data on advertising in yellow pages from the utility publisher in Central Pennsylvania are analyzed. Empirical results show an important heterogeneity in business 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 structure. Recent results by Aryal, Perrigne and Vuong (2009) in the context of insurance suggest that observations on repeated purchases allow to identify the type distribution using a moment generating function argument thereby avoiding the extensive use of the first-order conditions of the principal’s optimization problem. Another extension of interest is to test adverse selection in this market. To perform a structural test of adverse selection, we would need to derive the restrictions imposed by the nonlinear pricing model under both incomplete and complete information environments. We then can test which set of restrictions are validated by the data. Another extension could study the role of competition by incorporating a second publisher in the model. In the county where we collected the data, a non-utility publisher distributes a second phone book though of less importance and distributed at a significantly lower number. See Aryal and Huang (2009) for an analysis of competition under a Stackelberg framework. Lastly, the methodology developed in this paper in combination with Perrigne and Vuong (2010) can entertain other models with adverse selection in contract theory with a broad range of applications to retailing and labor to name a few.
1.7 Appendix A: Proofs in Section 1.3

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 $\theta$, 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 $\theta_1$ and $\theta_2$ satisfy $\theta_0 < \theta_1 < \theta_2 \leq \bar{\theta}$. If $U(\theta_2, \theta_1) > U(\theta_1, \theta_1)$, we have

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

We show that $U_1(x, \theta_1) \leq 0$ for $x \in [\theta_1, \theta_2]$ hence leading to a contradiction. By definition, $U(\tilde{\theta}, \theta) = \int_0^{\tilde{\theta}} v(x, \theta) dx - T(q(\bar{\theta}))$. Thus we have $U_{12}(\tilde{\theta}, \theta) = U_{21}(\tilde{\theta}, \theta) = v_2(q(\tilde{\theta}), \theta)q'(\tilde{\theta}) > 0$. Hence, $U_1(x, \theta_1) \leq U_1(x, x) = 0$ for $x \geq \theta_1$, 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 $\theta_0 < \theta_1 < \theta_2 \leq \bar{\theta}$.

Similarly, let $\theta_1$ and $\theta_2$ satisfy $\theta_0 < \theta_2 < \theta_1 \leq \bar{\theta}$. 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 $U_1(x, \theta_1) \geq U_1(x, x) = 0$ for $x \leq \theta_1$ by a similar argument, leading to a contradiction. Thus the local second-order condition holds globally. $\Box$
**Proof of Proposition 1:** Consider the change of variables $t = \bar{\theta} - \theta$ and $t_0 = \bar{\theta} - \theta_0$ in (7) so that $\theta_0$ becomes the terminal time. Hence, using $q_0 = q$, (7) becomes

\[
\Pi = -\int_{t_0}^{t_0} \left\{ \int_{\mathcal{Q}} q(\bar{x}) \nu(x, \bar{\theta} - t) dx \right\} f(\bar{\theta} - t) - \left[ 1 - F(\bar{\theta} - t) \right] \left\{ \int_{\mathcal{Q}} q(\bar{x}) \nu_2(x, \bar{\theta} - t) dx \right\} dt
\]

\[-C \left[ qF(\bar{\theta} - t_0) - \int_{t_0}^{t_0} q(\bar{\theta} - s) f(\bar{\theta} - s) ds \right].
\]

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

\[
\max_{\bar{q}(\cdot), t_0 \in [0, \bar{\theta} - \theta]} \Pi = \int_{0}^{t_0} \left\{ \int_{\mathcal{Q}} \bar{q}(\bar{x}, t) dx \right\} \bar{f}(t) + \bar{F}(t) \left\{ \int_{\mathcal{Q}} \bar{v}_2(x, t) dx \right\} dt
\]

\[-C \left[ q \left[ 1 - \bar{F}(t_0) \right] + \int_{t_0}^{t_0} \bar{q}(s) \bar{f}(s) ds \right]. \tag{A.2}
\]

We treat $\bar{q}(t)$ as the control variable and $\int_{0}^{t} \bar{q}(s) \bar{f}(s) ds$ as the state variable. The problem (A.2) can be written as a standard free terminal time and free-end point control problem

\[
\max_{\bar{q}(\cdot), t_0 \in [0, \bar{\theta} - \theta]} \Pi = \int_{0}^{t_0} \Psi [\bar{q}(t), t] dt + K [X(t_0), t_0], \tag{A.3}
\]

where

\[
X(t) = \int_{0}^{t} \bar{q}(s) \bar{f}(s) ds \quad \forall t \in [0, \bar{\theta} - \theta]
\]

\[
\Psi [\bar{q}(t), t] = \int_{\mathcal{Q}} \bar{q}(\bar{x}, t) dx \left[ \bar{f}(t) + \bar{F}(t) \left\{ \int_{\mathcal{Q}} \bar{v}_2(x, t) dx \right\} \right] \quad \forall t \in [0, \bar{\theta} - \theta]
\]

\[
K [X(t_0), t_0] = -C \left[ q \left[ 1 - \bar{F}(t_0) \right] + X(t_0) \right].
\]
The Hamiltonian function is

$$\mathcal{H}[X(t), \overline{q}(t), \lambda(t), t] = \Psi[\overline{q}(t), t] + \lambda(t)\overline{q}(t)\overline{f}(t),$$  \hspace{1cm} (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

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

$$\lambda'(t) = -\mathcal{H}_1[X(t), \overline{q}(t), \lambda(t), t] \quad \forall t \in [0, \overline{t} - \theta_0)$$

$$\lim_{t \uparrow t_0} \lambda(t) = \lim_{t \uparrow t_0} K_1[X(t), t]$$

$$\lim_{t \uparrow t_0} \mathcal{H}[X(t), \overline{q}(t), \lambda(t), t] = -\lim_{t \uparrow t_0} K_2[X(t), t].$$  \hspace{1cm} (A.5)$$

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

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

$$\lambda'(t) = 0 \quad \forall t \in [0, \overline{t} - \theta_0)$$

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

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

$$\lambda(t) = -C'\left(q \left[1 - \overline{F}(t_0)\right] + X(t_0)\right) \quad \forall t \in [0, \overline{t} - \theta_0).$$  \hspace{1cm} (A.9)$$
Plugging (A.9) into (A.6) gives the optimal \( \bar{q}(\cdot) \)

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

Plugging the optimal \( \lambda(\cdot) \) into (A.4) and letting \( \lim_{t \to t_0} \bar{q}(t) = q \) gives the following expression for the left-hand side of (A.5)

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

By definition of \( K[X(t_0), t_0] \), the right-hand side of (A.5) is \( -C' \left( q \left[ 1 - F(t_0) \right] + X(t_0) \right) q \bar{f}(t_0) \).

After equating the two terms and rearranging, we obtain

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

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

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

Letting \( \Gamma(x, t_0) \equiv \bar{v}(x, t_0) + [F(t_0)/\bar{f}(t_0)]\bar{v}_2(x, t_0) = \bar{v}(x, t_0) - [(1 - F(\theta_0))/f(\theta_0)]\bar{v}_2(x, t_0) \), the previous equality can be written as \( \int_{\bar{q}}^{q} [\Gamma(x, t_0) - \Gamma(\bar{q}, t_0)] dx = 0 \), which implies \( \bar{q} = q \) 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 \( \bar{q} = q \) 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 \( \lim_{\theta \to 0} \int_{q^{(\theta)}} q(x, \theta) \, dx = \lim_{\theta \to 0} T(q(\theta)) \). The left-hand side is equal to zero using (9) thereby establishing (11) since \( q(\cdot) \) is strictly increasing on \((\theta, \bar{\theta}]\). □

**Proof of Lemma 2:** To simplify the exposition, we suppress the arguments of functions hereafter. Taking the total derivative of (8) with respect to \( \theta \) 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

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

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

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

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 \((\theta_0, \bar{\theta}]\). 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 \((\theta_0, \bar{\theta}]\). That \( q(\cdot) \) is strictly increasing and continuous on \([\theta_0, \bar{\theta}]\) is obvious using (9). We now show that \( q(\cdot) \) is continuously differentiable at \( \theta_0 \), i.e. that \( q'(\theta_0) \) exists and is strictly positive with \( q'(\theta_0) = \lim_{\theta \to \theta_0} q'(\theta) < \infty \). By the Mean Value Theorem, we have
\[ q'(\theta_0) \equiv \lim_{\theta \downarrow \theta_0} \frac{[q(\theta) - q(\theta_0)]}{(\theta - \theta_0)} = q'(\hat{\theta}), \text{ where } \theta_0 < \hat{\theta} < \theta. \] But, \( \lim_{\theta \downarrow \theta_0} q'(\theta) \) is equal to the right-hand side of (A.12) evaluated at \( \theta_0 \), 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
\[ \frac{1}{v_2} \frac{\partial (v_2/\rho)}{\partial \theta} - 1 = \frac{v_{22}}{v_2^2} - \left( 1 + \frac{\rho'}{\rho^2} \right) < 0. \]

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 \((\theta_0, \bar{\theta})\) from above, it follows that the IC constraint (4) can be written as \( T'(q) = v(q, \theta(q)) \forall q \in (\underline{q}, \bar{q}) \). Recall that \( v(\cdot, \cdot) \) is continuously differentiable on \([\underline{q}, +\infty) \times [\underline{\theta}, \bar{\theta}]\) by A1 while \( \theta(\cdot) \) is continuously differentiable on \([\underline{q}, \bar{q}]\) as \( \theta'(q) = 1/q'(\theta) \) with \( q'(\cdot) \) strictly positive and continuous on \([\theta_0, \bar{\theta}]\) as noted above. Thus, \( T'(\cdot) \) is continuously differentiable on \((\underline{q}, \bar{q})\), i.e. \( T(\cdot) \) is twice continuously differentiable on \((\underline{q}, \bar{q})\). 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 \((\underline{q}, \bar{q})\). Thus, (4) holds on \([\underline{q}, \bar{q}]\) showing that \( T'(\cdot) \) is continuously differentiable on \([\underline{q}, \bar{q}]\) as \( v(\cdot, \cdot) \) and \( \theta(\cdot) \) are continuously differentiable on \([\underline{q}, +\infty) \times [\underline{\theta}, \bar{\theta}]\) and \([\underline{q}, \bar{q}]\), respectively. Hence, \( T(\cdot) \) is twice continuously differentiable on \([\underline{q}, \bar{q}]\) and \( T'(q) = v(q_0, \theta_0) > 0 \).

Regarding the other assertions on \( T'(\cdot) \), combining (8) and (10) gives
\[ T'(q) = C'(Q) + \frac{1 - F'(\theta)}{f(\theta)} v_2(q, \theta) \quad \forall \theta \in (\theta_0, \bar{\theta}). \]
This establishes $T'(\cdot) > C'(Q)$ on $(\bar{q}, \bar{q})$ by A1-(iii), while $T'(\bar{q}) = C'(Q)$.

**1.8 Appendix B: Proofs in Section 1.4**

**Proof of Lemma 4:** Let $\tilde{\theta} = \alpha \theta$, which is distributed as $\tilde{F}(\cdot)$ on $[\tilde{\theta}, \tilde{\theta}] = [\alpha \theta, \alpha \bar{\theta}]$. Let $\tilde{T}(\cdot) \equiv T(\cdot), \tilde{q}(\cdot) \equiv q(\cdot/\alpha), \tilde{\theta}_0 = \alpha \theta_0$ and $\tilde{Q} \equiv \tilde{q} \tilde{F}(\tilde{\theta}_0) + \int_{\tilde{\theta}_0}^{\tilde{\theta}} \tilde{q}(u) \tilde{f}(u) du$. 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 $\tilde{q}$, $\tilde{Q} = \tilde{Q}$, and $\tilde{F}(\tilde{\theta}_0) = F(\theta_0)$. Hence, the observables $[\tilde{T}(\cdot), q, \tilde{F}(\tilde{\theta}_0), \tilde{G}^*(\cdot), \tilde{Q}]$ generated by the structure $\tilde{S}$ are the same observables $[T(\cdot), q, F(\theta_0), G^*(\cdot), Q]$ generated by the structure $S$. Lastly, we show $\tilde{S} \in S$.

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

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

for all $\theta \in (\theta_0, \bar{\theta})$. If $\tilde{Q} = Q$, the above equation holds for all $\theta = \theta/\alpha \in (\theta_0, \bar{\theta})$ 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 $\tilde{F}(\cdot), \tilde{q}(\cdot), \tilde{\theta}_0, \tilde{\theta}$ and $\tilde{f}(\cdot)$, we have

$$\tilde{Q} = q F(\alpha \theta_0/\alpha) + \int_{\alpha \theta_0}^{\alpha \bar{\theta}} q(u/\alpha) f(u/\alpha) du = q F(\theta_0) + \int_{\theta_0}^{\bar{\theta}} q(\theta) f(\theta) d\theta = Q.$$
Second, we show $\tilde{G}^*(\cdot) = G^*(\cdot)$. Namely,

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

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 $S$. Assumptions B1 and A2 are trivially satisfied. Regarding A3, we have

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

which is strictly monotone in $\tilde{q}$ or identically equal to zero for all $\theta \in (\theta_0, \tilde{\theta}]$. 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 $\tilde{\theta}$. \square

**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

\[ T'(q) = \gamma + \frac{1 - G^*(q)}{g^*(q)} \theta'(q)v_0(q) \text{ for } q \in (q, \bar{q}). \]  

(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

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

and hence for \( q \in (q, \bar{q}] \) 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

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

Integrating both sides of the above equation from \( q \) to \( \bar{q} \) gives

\[ \log \left( \frac{v_0(q)}{v_0(q)} \right) = \log \left( \frac{T'(q)}{T'(q)} \right) - \int_q^{\bar{q}} \frac{g^*(x)}{1 - G^*(x)} \left( 1 - \frac{\gamma}{T'(x)} \right) dx \text{ for } q \in (q, \bar{q}). \]  

(B.2)

Taking the exponential gives

\[ \frac{v_0(q)}{v_0(q)} = \frac{T'(q)}{T'(q)} \exp \left[ - \int_q^{\bar{q}} \frac{g^*(x)}{1 - G^*(x)} \left( 1 - \frac{\gamma}{T'(x)} \right) dx \right] \text{ for } q \in (q, \bar{q}). \]  

(B.3)
Condition (14) evaluated at \( q \) gives \( T'(q) = \theta_0 v_0(q) \). Multiplying the right-hand side of (A.15) by \( T'(q) \) and the left-hand side by \( \theta_0 v_0(q) \) gives for \( q \in (q, \bar{q}] \)

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

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

\[
= \frac{T'(q)}{\theta_0} \left[ 1 - G^*(q) \right] \exp \left[ -\gamma \left\{ \log(1 - G^*(x)) \right\}^q_q + \int_q^q \log(1 - G^*(x)) \frac{T''(x)}{T'(x)} dx \right]
\]

\[
= \frac{T'(q)}{\theta_0} \left[ 1 - G^*(q) \right]^{1 - \frac{\gamma}{T'(q)}} \exp \left\{ -\gamma \int_q^q \log \left[ 1 - G^*(x) \right] \frac{T''(x)}{T'(x)^2} dx \right\}
\],

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 \( \xi(q) = 1 \) and \( \lim_{q \to \bar{q}} \xi(q) = \bar{q}/\theta_0 \) since \( \bar{q} v_0(q) = T'(\bar{q}) \).

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

\[
\tilde{v}_0(q) = \frac{T'(q)}{\tilde{\theta}_0} \left[ 1 - G^*(q) \right]^{1 - \frac{\gamma}{T'(q)}} \exp \left\{ -\gamma \int_q^q \log \left[ 1 - G^*(x) \right] \frac{T''(x)}{T'(x)^2} dx \right\} \quad \forall q \in (q, \bar{q}]
\]

\[
\tilde{\theta}(q) = \tilde{\theta}_0 \left[ 1 - G^*(q) \right]^{\frac{\gamma}{T'(q)}} \exp \left\{ \gamma \int_q^q \log \left[ 1 - G^*(x) \right] \frac{T''(x)}{T'(x)^2} dx \right\} \quad \forall q \in (q, \bar{q}].
\]

By B2, \( \theta_0 = \tilde{\theta}_0 \) showing \( \tilde{v}_0(\cdot) = v_0(\cdot) \) and \( \theta(\cdot) = \tilde{\theta}(\cdot) \) on \([q, \bar{q}]\) and hence on \([q, \bar{q}]\) by continuity at \( q \). Thus, \( \tilde{F}^*(\cdot) = G^*(\tilde{q}(\cdot)) = G^*(q(\cdot)) = F^*(\cdot) \) on \([\theta_0, \bar{q}]\). Thus, \( v_0(\cdot) \) and \( F^*(\cdot) \) are uniquely
We begin by noting that for every \( z = (z_1, z_2) \in S_Z = S_{Z_1, Z_2} \)

\[
F^*_\epsilon|Z(e|z) = \Pr[\epsilon \leq e|Z = z, \epsilon > e(z)] = \Pr[\theta \leq r(z, e)|Z = z, \theta > 1]
\]

\[= F^*_\theta|Z(r(z, e)|z), \quad (B.4)\]

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

\[
F^*_\epsilon|Z(e|z) = \frac{F\epsilon|Z(e|z) - F\epsilon|Z(e(z)|z)}{1 - F\epsilon|Z(e(z)|z)}, \quad (B.5)
\]

for every \( e \in [e(z), +\infty) \). We note that \( F\epsilon|Z(e(z)|z) = \Pr[\epsilon \leq e(z)|Z = z] = \Pr[r(z, \epsilon) \leq 1|Z = z] = \Pr[\theta = \theta_0|Z = z] \), thereby establishing that \( F\epsilon|Z(e(z)|z) \) is identified for every \( z \in S_Z \).

We now turn to the identification of \( F\epsilon|Z(\cdot|\cdot) \). Applying (B.4) and (B.5) at \((z^0_1, z_2)\) with \( z_2 \in S_{Z_2|z^0_1} \) gives

\[
\frac{F\epsilon|Z(z^0_1, z_2|z^0_1, z_2) - F\epsilon|Z(z^0_1, z_2|z^0_1, z_2)}{1 - F\epsilon|Z(z^0_1, z_2|z^0_1, z_2)} = F^*_\theta|Z(z^0_1, z_2), \quad (B.6)
\]

for every \( e \in [e(z^0_1, z_2), +\infty) \cap S_{\epsilon|z^0_1, z_2} = [\theta_0, +\infty) \cap S_{\epsilon|z^0_1, z_2} \) since \( e(z^0_1, z_2) = \theta_0 \) as explained in the text. By B4-(i), we have \( S_{\epsilon|z^0_1, z_2} = S_{\epsilon|z_1, z_2} \) and \( F\epsilon|Z(z^0_1, z_2|z^0_1, z_2) = F\epsilon|Z(z_1, z_2|z_1, z_2) \).

Using \( S_{Z_2|z^0_1} = S_{Z_2} \) by B4-(ii), simple algebra establishes the expression for \( F\epsilon|Z(z_1, z_2|z_1, z_2) \) in Proposition 3, and hence its identification on \( \{(z_1, z_2, e) \in S_{Z_1, Z_2, \epsilon} : e \geq \theta_0\} \).
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 \( \{(z, e) \in S_{Z,\epsilon} : e \geq \max[e(z), \theta_0]\} \). Inverting (B.4) establishes the expression for \( r(z, e) \) in Proposition 3, and hence its identification on \( \{(z, e) \in S_{Z,\epsilon} : e \geq \max[e(z), \theta_0]\} \). Moreover, \( \max[e(z), \theta_0] \) as a function of \( z \in S_Z \) is identified by

\[
\max[e(z), \theta_0] = F^{-1}_{\epsilon|Z}[\Pr(\tilde{\theta} = 1|Z = z)] - \theta_0 \leq \Pr(\tilde{\theta} = 1|Z = (z^0_1, z^0_2)).
\]

This follows from \( e(z) > \theta_0 \) if and only if \( \Pr(\tilde{\theta} = 1|Z = z) > \Pr(\tilde{\theta} = 1|Z = (z^0_1, z^0_2)) \), while \( e(z) = F^{-1}_{\epsilon|Z}[\Pr(\tilde{\theta} = 1|Z = z)] \) 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) \( q_{\max} = \bar{q} + O_{a.s.}[\log \log N^*/N^*] \), and (ii) \( N^*(q_{\max} - \bar{q}) \xrightarrow{D} -\mathcal{E}/g^*(\bar{q}) \) 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 \( [q, \bar{q}] \) as shown at the beginning of the proof of Lemma 5. For instance, (i) follows from Galambos (1978) Theorem 4.3.1 and Example 4.3.2 by letting \( u_{N^*} = \bar{q} - \kappa(\log \log N^*)/N^* \) for any \( \kappa > 1 \) so that \( \sum_{N^* = 2}^\infty [1 - G^*(u_{N^*})] \exp\{-N^*[1 - G^*(u_{N^*})]\} = \sum_{N^* = 2}^\infty v_{N^*} < \infty \) as \( v_{N^*} \sim \hat{v}_{N^*} \equiv \kappa g^*(\bar{q})[\log \log N^*]/[N^*(\log N^*)^\kappa g^*(\bar{q})] \) as \( N^* \to \infty \) with \( \sum_{N^* = 2}^\infty \hat{v}_{N^*} < \infty \). Thus, \( \Pr[\bar{q} - q_{\max} \geq \kappa(\log \log N^*)/N^* \text{ i.o.}] = 0 \), i.e. \( \Pr[0 \leq (N^*/\log \log N^*) (\bar{q} - q_{\max}) \leq \kappa \text{ for } N^* \text{ sufficiently large}] = 1 \). Similarly, (ii) follows from Galambos (1978) Theorem 2.1.2 and Section 2.3.1 with \( a_n = \bar{q} \) and \( b_n = \bar{q} - G^{*-1}(1-1/N^*) \). Specifically,
since \( \lim_{t \to \infty} \frac{1-G^*(\bar{q}-1/(tx))}{1-G^*(\bar{q}-1/t)} = 1/x \) for \( x > 0 \), we obtain \((q_{\text{max}}-\bar{q})/b_n \overset{D}{\to} -\mathcal{E} \), i.e. (ii) as \( b_n \sim 1/[g^*(\bar{q})N^*] \) as \( N^* \to \infty \).

The lemma then follow from the standard delta method. Namely, \( \hat{\gamma} - \gamma = T'(q_{\text{max}}) - T'(\bar{q}) = T''(\bar{q})(q_{\text{max}} - \bar{q}) \), where \( q_{\text{max}} < \hat{q} < \bar{q} \) using a Taylor expansion and the continuous differentiability of \( T'(\cdot) \) by Lemma 2. Moreover, \( N^*/N \overset{a.s.}{\to} 1 - F(\theta_0) > 0 \). \( \square \)

To establish Proposition 3, we use two lemmas.

**Lemma 7:** Under B3, as \( N^* \to \infty \) we have

\[
(i) \left\| \log \left( \frac{1 - \hat{G}^*(\cdot)}{1 - G^*(\cdot)} \right) \right\|_{\mathcal{T}} \overset{a.s.}{\to} 0, \quad (ii) \sqrt{N^*} \log \left( \frac{1 - \hat{G}^*(\cdot)}{1 - G^*(\cdot)} \right) \Rightarrow \frac{-B_{G^*}(\cdot)}{1 - G^*(\cdot)}. \]

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

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

(i) Let \( \mathcal{C} \) be the set of continuous functions on \( [\bar{q}, q_\dagger] \) 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 \( \|\psi\|_\mathcal{T} < 1 \) 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, \( \|\hat{G}^* - G^*\|_\mathcal{T} \overset{a.s.}{\to} 0 \) by the Glivenko-Cantelli theorem as \( N^* \to \infty \). It follows from Van der Vaart (1998) Theorem 18.11-(iii) that \( \|h(\hat{G}^*) - h(G^*)\|_\mathcal{T} \overset{a.s.}{\to} 0 \), where \( h(G^*) = \log(1 - G^*) \) since \( G^* \in \mathcal{C} \) as \( \|G^*\|_\mathcal{T} \leq G^*(q_\dagger) < 1 \). Noting that \( h(\hat{G}^*) = \log(1 - \hat{G}^*) \) because \( \|\hat{G}^*\|_\mathcal{T} \leq \hat{G}^*(q_\dagger) \leq (N^*-1)/N^* < 1 \), 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 \( \|\psi\|_\mathcal{T} < \sqrt{N^*} \) and \( h(\psi)(\cdot) = 0 \) otherwise. Let \( h_0(\cdot) \) be minus the identity. In particular, if \( \|\psi_{N^*} - \psi\|_\mathcal{T} \to 0 \) with \( \psi \in \mathcal{D} \), we have \( \|\psi_{N^*}\|_\mathcal{T}/N^\alpha \to 0 \) for any \( \alpha > 0 \) because \( \|\psi\|_\mathcal{T} < \infty \).
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 \( \log(1 + x) = x - x^2/[2(1 + x)^2] \) with \( 0 < \bar{x} < |x| \) from a second-order Taylor expansion, we obtain for \( x = -\psi_{N^*}(q)/\sqrt{N^*} \) and \( 0 < |\psi_{N^*}(q)| < |\psi_{N^*}(q)|/\sqrt{N^*} \)

\[
\| h_{N^*}(\psi_{N^*}) - h_0(\psi) \|_1 = \sup_{q \in [\bar{q}, \bar{q}]} \sqrt{N^*} \log \left[ 1 - \psi_{N^*}(q)/\sqrt{N^*} \right] + \psi(q)
\]

\[
= \sup_{q \in [\bar{q}, \bar{q}]} \left| \psi(q) - \psi_{N^*}(q) - \frac{\psi_{N^*}(q)^2}{2\sqrt{N^*}(1 + \psi_{N^*}(q))^2} \right|
\]

\[
\leq \| \psi_{N^*} - \psi \|_1 + \frac{1}{2} \sup_{q \in [\bar{q}, \bar{q}]} \psi_{N^*}(q)^2/\sqrt{N^*} = 1 - \inf_{q \in [\bar{q}, \bar{q}]} |1 + \psi_{N^*}(q)|^2
\]

\[
\leq \| \psi_{N^*} - \psi \|_1 + \frac{1}{2} \| \psi_{N^*} \|_1^2/\sqrt{N^*} \]

Thus, \( \| h_{N^*}(\psi_{N^*}) - h_0(\psi) \|_1 \to 0 \) as \( \| \psi_{N^*} - \psi \|_1 \to 0 \) and \( \| \psi_{N^*} \|_1/\sqrt{N^*} \to 0 \) for \( \alpha = 1/2 \) and 1.

On the other hand, \( \sqrt{N^*} (\hat{G}^*_n - G^_n)/(1 - G^*_n) \Rightarrow B_{G^*_n}/(1 - G^*_n) \) since \( \sqrt{N^*} (\hat{G}^*_n - G^*_n) \Rightarrow B_{G^*_n} \) 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

\[
h_{N^*} \left( \sqrt{N^*} \frac{\hat{G}^*_n - G^*_n}{1 - G^*_n} \right) \Rightarrow h_0 \left( \frac{B_{G^*_n}}{1 - G^*_n} \right) = -B_{G^*_n}/(1 - G^*_n)
\]

But, as \( N^* \to \infty \), \( \| \sqrt{N^*} (\hat{G}^*_n - G^*_n)/(1 - G^*_n) \|_1 < \sqrt{N^*} \) since \( \| (\hat{G}^*_n - G^*_n)/(1 - G^*_n) \|_1 \overset{a.s.}{\to} 0 \). Thus,

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

This establishes the desired result. \( \square \)

**Lemma 8:** Under A1, A2, A4, B1–B3, for any \( q \in (\bar{q}, \bar{q}) \), we have as \( N \to \infty \)
(i) \(\|\hat{\xi}(-) - \xi(-)\| \xrightarrow{a.s.} 0\) and \(\hat{\xi}(q_{\text{max}}) \xrightarrow{a.s.} \xi(\bar{q})\),

(ii) as random functions in \(D[q, q]\), \(\sqrt{N}[\hat{\xi}(-) - \xi(-)] \Rightarrow \xi(-)Z(-)/\sqrt{1 - F(\theta_0)}\), where \(Z(-)\) is a tight Gaussian process.

**Proof of Lemma 8:** By (16) and (17), \(\xi(q) > 0\) and \(\hat{\xi}(q) > 0\) for all \(q \in [q, q_{\text{max}}]\). Thus,

\[
\log \frac{\hat{\xi}(q)}{\xi(q)} = \left[\frac{\gamma}{T'(-q)} - 1\right] \log \left[\frac{1 - \hat{G}^*(q)}{1 - G^*(q)}\right] + \gamma \int_q q' \frac{T''(x)}{T'(x)^2} \log \left[\frac{1 - \hat{G}^*(x)}{1 - G^*(x)}\right] dx
\]

\[
+ \frac{\hat{\gamma} - \gamma}{T'(q)} \log \left[1 - \hat{G}^*(q)\right] + (\hat{\gamma} - \gamma) \int_q q' \frac{T''(x)}{T'(x)^2} \log \left[1 - \hat{G}^*(x)\right] dx, \quad (B.7)
\]

for \(q \in [q, q_{\text{max}}]\). Moreover, \(q_{\dagger} < q_{\text{max}}\) almost surely as \(q_{\text{max}} \xrightarrow{a.s.} \bar{q}\) by the proof of Lemma 6.

(i) From (A.16), Lemma 6-(i) and Lemma 7-(i), it follows that \(\|\log(\hat{\xi}/\xi)\| \xrightarrow{a.s.} 0\). By the continuous mapping theorem, we obtain \(\|\hat{\xi}/\xi\| \xrightarrow{a.s.} 1\) and hence \(\|\hat{\xi} - \xi\| \xrightarrow{a.s.} 0\) as \(0 < \|\xi\|_{\dagger} < \infty\). Next, to show \(\hat{\xi}(q_{\text{max}}) \xrightarrow{a.s.} \xi(\bar{q}) > 0\), we first note that

\[
\left[\frac{\gamma}{T'(-q)} - 1\right] \log[1 - G^*(q)] = \frac{1}{T'(q)} \frac{T'(-q) - T'(-q)}{\bar{q} - q} \frac{\bar{q} - q}{1 - G^*(q)}[1 - G^*(q)] \log[1 - G^*(q)] \to 0
\]

as \(q \uparrow \bar{q}\). Hence, \([1 - G^*(q)]^{\gamma/T'(q)^{-1}} \to 1\) as \(q \uparrow \bar{q}\) implying that

\[
\lim_{q \uparrow \bar{q}} \exp \left\{\gamma \int_q q' \frac{T''(x)}{T'(x)^2} \log[1 - G^*(x)] dx\right\} = \xi(\bar{q}) = \bar{\theta} < \infty
\]

using (15), (16) and B2. Hence, for any \(\epsilon > 0\), there exist \(q_{\dagger} \in (q, \bar{q})\) such that

\[
\left|\xi(\bar{q}) - \exp \left\{\gamma \int_q q' \frac{T''(x)}{T'(x)^2} \log[1 - G^*(x)] dx\right\}\right| < \epsilon \quad (B.8)
\]
On the other hand, (20) can be written as

\[
\frac{1}{\gamma} \log [\hat{\xi}(q_{\max})] = \int_{q}^{q_{\max}} \frac{T''(x)}{T'(x)^2} \log [1 - \hat{G}(x)] \, dx + \int_{q}^{q_{\max}} \frac{T''(x)}{T'(x)^2} \log [1 - \hat{G}(x)] \, dx \tag{B.9}
\]

where the first integral converges almost surely to the same integral with \(\hat{G}(\cdot)\) replaced by \(G(\cdot)\) by Lemma 7-(i), while the second integral is bounded in absolute value by

\[
\sup_{q \in [q, q_{\max}]} \frac{|T''(q)|}{T'(q)^2} \int_{q}^{q_{\max}} \log [1 - \hat{G}(x)] \, dx \leq \sup_{q \in [q, q_{\max}]} \frac{|T''(q)|}{T'(q)^2} (q_{\max} - q) \log N^*
\]

as \(1/N^* \leq 1 - \hat{G}(x) \leq 1\) for all \(x \in [q, q_{\max}]\). To be completed

(ii) Using Lemma 6-(ii) and Lemma 7-(i), it follows from (A.16) that

\[
\sqrt{N} \log \frac{\hat{\xi}(q)}{\xi(q)} = \sqrt{\frac{N}{N^*}} \left\{ \frac{\gamma}{T'(q)} - 1 \right\} \sqrt{N^*} \log \left[ 1 - \hat{G}^*(q) \right] + \gamma \int_{q}^{q_{\max}} \frac{T''(x)}{T'(x)^2} \sqrt{N^*} \log \left[ 1 - \hat{G}^*(x) \right] \, dx + o_P(1)
\]

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

\[
\sqrt{N} \log \frac{\hat{\xi}(\cdot)}{\xi(\cdot)} \Rightarrow \frac{Z(\cdot)}{\sqrt{1 - F(\theta_0)}} \tag{B.10}
\]

by (23). From a first-order Taylor expansion we have

\[
\sqrt{N} \log \frac{\hat{\xi}(q)/\xi(q)}{\xi(q)} = \sqrt{N} \left[ \hat{\xi}(q) - \xi(q) \right] / \hat{\xi}(q),
\]

where \(\xi(q) < \hat{\xi}(q) < \hat{\xi}(q)\) for any \(q \in [q, q_{\max}]\). Thus, \(\sqrt{N} \left[ \hat{\xi}(q) - \xi(q) \right] = \hat{\xi}(q) \sqrt{N} \log \left[ \hat{\xi}(q)/\xi(q) \right] \).

The desired result follows from (A.19) and \(\|\hat{\xi} - \xi\|_{\sup} \overset{a.s.}{\rightarrow} 0\). Lastly, \(Z(\cdot)\) is a tight Gaussian process as it is linear in \(\mathcal{B}_{G^*(\cdot)}\). \(\square\)
Proof of Proposition 3: Using B2, we have \( v_0(q) = T'(q)/\xi(q) > 0 \) by (15), while \( \hat{v}_0(q) = T'(q)/\hat{\xi}(q) > 0 \) by (21) for all \( q \in [q, q_{\text{max}}) \) as \( T'(\cdot) > 0 \) by Lemma 2. Moreover, \( 0 < \|\xi\|_\dagger < \infty \). Thus, (i) follows from Lemma 8-(i), \( \hat{v}_0(q_{\text{max}}) = T'(q_{\text{max}})/\hat{\xi}(q_{\text{max}}) \) and \( v_0(\bar{q}) = T'(\bar{q})/\xi(\bar{q}) \). Regarding (ii), we have \( \sqrt{N} \log[v_0(\cdot)/v_0(\cdot)] = \sqrt{N} \log[\hat{\xi}(\cdot)/\hat{\xi}(\cdot)] \Rightarrow -Z(\cdot)/\sqrt{1-F(\theta_0)} \) on \( D[q, q_t] \).

The desired result follows by the argument at the end of the proof of Lemma 8. \( \square \)

Proof of Corollary 1: Using B2, we have \( \theta(q) = \xi(q) \) by (15), while \( \hat{\theta}(q) = \hat{\xi}(q) \) by (21) for all \( q \in [q, q_{\text{max}}) \). Moreover, \( \hat{\theta}(q_{\text{max}}) = \hat{\xi}(q_{\text{max}}) \), while \( \bar{\theta} = \theta(\bar{q}) = \xi(\bar{q}) \). Thus, the desired results follow from Lemma 8. \( \square \)

Proof of (24): We have \( E[B_{G^*}(q)B_{G^*}(q')] = G^*(q)[1-G^*(q')] \) for \( q \leq q' \leq \overline{q} \). See e.g. Van der Vaart (1998, p.266). Thus, for \( q \leq q' < \overline{q} \)

\[
E \left[ \frac{B_{G^*}(q)}{1-G^*(q)} \frac{B_{G^*}(q')}{1-G^*(q')} \right] = \frac{G^*(q)}{1-G^*(q)}.
\]
which is independent of \( q' \). Hence, from the definition of \( Z(\cdot) \) we have for \( q \leq q \leq q' < \bar{q} \)

\[
E[Z(q)Z(q')] = \left[ 1 - \gamma \right] \left[ 1 - \gamma \right] E \left[ \frac{G^*(q)}{1 - G^*(q)} \frac{G^*(q)}{1 - G^*(q')} \right] - \gamma \left( 1 - \gamma \right) \int_q^{q'} \frac{T''(x)}{T'(x)^2} E \left[ \frac{B_{G^*}(q)}{1 - G^*(q)} \frac{B_{G^*}(x)}{1 - G^*(x)} \right] dx
\]

\[
- \gamma \left( 1 - \gamma \right) \int_q^{q'} \frac{T''(x)}{T'(x)^2} E \left[ \frac{B_{G^*}(x)}{1 - G^*(x)} \frac{B_{G^*}(q')} {1 - G^*(q')} \right] dx
\]

\[
+ \gamma^2 \int_q^{q'} \left\{ \int_q^{q'} \frac{T''(x)}{T'(x)^2} \frac{T''(x')}{T'(x')^2} E \left[ \frac{B_{G^*}(x)}{1 - G^*(x)} \frac{B_{G^*}(x')}{1 - G^*(x')} \right] dx' \right\} dx
\]

\[
= \left[ 1 - \gamma \right] \left[ 1 - \gamma \right] \frac{G^*(q)}{1 - G^*(q)} \left. \left. - \gamma \left( 1 - \gamma \right) \int_q^{q'} \frac{T''(x)}{T'(x)^2} \frac{G^*(x)}{1 - G^*(x)} dx \right. \right. + \frac{G^*(q)}{1 - G^*(q)} \int_q^{q'} \frac{T''(x)}{T'(x)^2} dx
\]

\[
- \gamma \left( 1 - \gamma \right) \int_q^{q'} \frac{T''(x)}{T'(x)^2} \frac{G^*(x)}{1 - G^*(x)} dx \right. \right. + \gamma^2 \int_q^{q'} \left. \left. \int_q^{q'} \frac{T''(x)}{T'(x)^2} \frac{T''(x')}{T'(x')^2} \frac{G^*(x')}{1 - G^*(x')} dx' \right. \right. + \frac{G^*(x)}{1 - G^*(x)} \int_q^{q'} \frac{T''(x')}{T'(x')^2} dx'
\]

Using \( \int_a^b \frac{T''(x)}{T'(x)^2} dx = \left[ 1/T'(a) \right] - \left[ 1/T'(b) \right] \)

\[
\int_q^{q'} \frac{T''(x)}{T'(x)^2} \int_q^{x} \frac{T''(x')}{T'(x')^2} \frac{G^*(x')}{1 - G^*(x')} dx'dx
\]

\[
\left. \left. = \left[ 1 - \gamma \right] \left[ 1 - \gamma \right] \frac{G^*(q)}{1 - G^*(q)} \int_q^{q'} \frac{T''(x)}{T'(x)^2} \frac{G^*(x)}{1 - G^*(x)} dx \right. \right. + \int_q^{q'} \frac{T''(x)}{T'(x)^2} \frac{G^*(x)}{1 - G^*(x)} dx,
\]

which is obtained by integration by parts, the desired result follows after some algebra. \( \square \)
1.9 Appendix C: Price Schedule

As seen in Section 1.2, advertisements differ in size and quality (color and other features). It may be useful to review how quality can be integrated in models nonlinear pricing. Maskin and Riley (1984) consider a situation where the monopolist can discriminate consumers by an optimal bundling of the quantity and quality levels of the product, both taking continuous values. With a one dimensional parameter of adverse selection, the optimal bundle should line up on an increasing curve in the quantity-quality space. Data display a very imperfect correlation between the number of colors and the advertising size chosen by businesses. In particular, businesses choose different color options for a given advertising size. An alternative model would be to consider a multidimensional parameter of adverse selection. This would lead to a complex model with 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) for multidimensional screening in nonlinear pricing models.

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, the price ratio is equal to 1.5 across the 9 possible advertisement sizes. Similarly, when considering the no color and multicolor display price schedules, the ratio is equal to 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. In a similar spirit, because of technological constraints, advertisement sizes cannot be offered on a continuous scale. The various color options can then be viewed as extending the size scale.

Following this empirical evidence, we construct a quality-adjusted quantity index. We consider the price schedule for multicolor and adjust the advertising sizes for other color options. Our choice of the price schedule for multicolor is made without loss of generality. We could have chosen another price schedule though the observed size range would be narrower. The multicolor price schedule avoids this
problem. Using the multicolor price information from the Yellow Page Association, we obtain

\[ \log(T) = 4.1602 + 0.7317 \times \log(q) + 0.0062 \times (\log(q))^2, \]  

(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 almost a perfect fit.\(^{33}\) Equation (C.1) gives the price schedule \( T(\cdot) \) in the theoretical model of Section 3.

The quality-adjusted quantity \( q \) is constructed as follows. For every observed price, the quality-adjusted quantity is obtained by solving (C.1) for \( q \). As an illustration, a one-page advertising display measuring 3,020 square picas with no color and no additional feature now corresponds to a multicolor adjusted quantity of 1,470 picas. Similarly, a half-page advertising display of 1,485 square picas with one color corresponds now 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, we measure the standard listing offered at zero price as \( q \) as 1.8512/2 .

\(^{33}\) A least squares estimator is appropriate since there is no endogeneity problem in this equation as the variable \( q \) is not endogenous.
Table 1.1: Revenue Ranking by Industry Headings

<table>
<thead>
<tr>
<th>Industry heading</th>
<th>Revenue</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attorneys</td>
<td>$512,587</td>
<td>8.53%</td>
</tr>
<tr>
<td>Dentists</td>
<td>$222,324</td>
<td>3.70%</td>
</tr>
<tr>
<td>Insurance</td>
<td>$188,302</td>
<td>3.13%</td>
</tr>
<tr>
<td>Storage household &amp; commercial</td>
<td>$141,063</td>
<td>2.35%</td>
</tr>
<tr>
<td>Restaurants</td>
<td>$130,673</td>
<td>2.17%</td>
</tr>
<tr>
<td>Hotels</td>
<td>$130,655</td>
<td>2.17%</td>
</tr>
<tr>
<td>Physicians and Surgeons</td>
<td>$126,070</td>
<td>2.10%</td>
</tr>
<tr>
<td>Plumbing contractors</td>
<td>$101,840</td>
<td>1.69%</td>
</tr>
<tr>
<td>Carpet &amp; rug cleaners</td>
<td>$99,861</td>
<td>1.66%</td>
</tr>
<tr>
<td>Veterinarians</td>
<td>$91,607</td>
<td>1.52%</td>
</tr>
<tr>
<td>Listing</td>
<td>Picas of a page</td>
<td># Purchases</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>12</td>
<td>0.4%</td>
<td>2,152</td>
</tr>
<tr>
<td>18</td>
<td>0.6%</td>
<td>844</td>
</tr>
<tr>
<td>24</td>
<td>0.79%</td>
<td>223</td>
</tr>
<tr>
<td>27</td>
<td>0.89%</td>
<td>802</td>
</tr>
<tr>
<td>30</td>
<td>0.99%</td>
<td>117</td>
</tr>
<tr>
<td>36</td>
<td>1.19%</td>
<td>127</td>
</tr>
<tr>
<td>39</td>
<td>1.29%</td>
<td>133</td>
</tr>
<tr>
<td>42</td>
<td>1.39%</td>
<td>31</td>
</tr>
<tr>
<td>45</td>
<td>1.49%</td>
<td>8</td>
</tr>
<tr>
<td>46</td>
<td>1.52%</td>
<td>12</td>
</tr>
<tr>
<td>48</td>
<td>1.59%</td>
<td>9</td>
</tr>
<tr>
<td>51</td>
<td>1.69%</td>
<td>45</td>
</tr>
<tr>
<td>54</td>
<td>1.79%</td>
<td>23</td>
</tr>
<tr>
<td>57</td>
<td>1.89%</td>
<td>1</td>
</tr>
<tr>
<td>58</td>
<td>1.92%</td>
<td>1</td>
</tr>
<tr>
<td>60</td>
<td>1.99%</td>
<td>4</td>
</tr>
<tr>
<td>63</td>
<td>2.09%</td>
<td>11</td>
</tr>
<tr>
<td>66</td>
<td>2.19%</td>
<td>15</td>
</tr>
<tr>
<td>72</td>
<td>2.38%</td>
<td>10</td>
</tr>
<tr>
<td>75</td>
<td>2.48%</td>
<td>1</td>
</tr>
<tr>
<td>78</td>
<td>2.58%</td>
<td>1</td>
</tr>
<tr>
<td>81</td>
<td>2.68%</td>
<td>10</td>
</tr>
<tr>
<td>84</td>
<td>2.78%</td>
<td>3</td>
</tr>
<tr>
<td>87</td>
<td>2.88%</td>
<td>3</td>
</tr>
<tr>
<td>93</td>
<td>3.08%</td>
<td>8</td>
</tr>
<tr>
<td>96</td>
<td>3.18%</td>
<td>2</td>
</tr>
<tr>
<td>99</td>
<td>3.28%</td>
<td>2</td>
</tr>
<tr>
<td>105</td>
<td>3.48%</td>
<td>2</td>
</tr>
<tr>
<td>108</td>
<td>3.58%</td>
<td>3</td>
</tr>
<tr>
<td>111</td>
<td>3.68%</td>
<td>1</td>
</tr>
<tr>
<td>114</td>
<td>3.77%</td>
<td>1</td>
</tr>
<tr>
<td>117</td>
<td>3.87%</td>
<td>3</td>
</tr>
<tr>
<td>120</td>
<td>3.97%</td>
<td>1</td>
</tr>
<tr>
<td>132</td>
<td>4.37%</td>
<td>1</td>
</tr>
<tr>
<td>135</td>
<td>4.47%</td>
<td>1</td>
</tr>
<tr>
<td>138</td>
<td>4.57%</td>
<td>1</td>
</tr>
<tr>
<td>144</td>
<td>4.77%</td>
<td>1</td>
</tr>
<tr>
<td>159</td>
<td>5.26%</td>
<td>1</td>
</tr>
<tr>
<td>162</td>
<td>5.36%</td>
<td>1</td>
</tr>
<tr>
<td>189</td>
<td>6.26%</td>
<td>1</td>
</tr>
<tr>
<td>192</td>
<td>6.36%</td>
<td>1</td>
</tr>
<tr>
<td>198</td>
<td>6.56%</td>
<td>1</td>
</tr>
<tr>
<td>225</td>
<td>7.45%</td>
<td>1</td>
</tr>
<tr>
<td>228</td>
<td>7.55%</td>
<td>1</td>
</tr>
<tr>
<td>282</td>
<td>9.34%</td>
<td>2</td>
</tr>
<tr>
<td>336</td>
<td>11.13%</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4,623</strong></td>
<td><strong>$648,127</strong></td>
</tr>
<tr>
<td>Picas²</td>
<td>Percent of a page</td>
<td># Purchases</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>54</td>
<td>1.79%</td>
<td>806</td>
</tr>
<tr>
<td>66</td>
<td>2.19%</td>
<td>88</td>
</tr>
<tr>
<td>72</td>
<td>2.38%</td>
<td>215</td>
</tr>
<tr>
<td>78</td>
<td>2.56%</td>
<td>24</td>
</tr>
<tr>
<td>81</td>
<td>2.68%</td>
<td>5</td>
</tr>
<tr>
<td>84</td>
<td>2.78%</td>
<td>26</td>
</tr>
<tr>
<td>90</td>
<td>2.96%</td>
<td>3</td>
</tr>
<tr>
<td>96</td>
<td>3.18%</td>
<td>11</td>
</tr>
<tr>
<td>99</td>
<td>3.28%</td>
<td>5</td>
</tr>
<tr>
<td>102</td>
<td>3.38%</td>
<td>1</td>
</tr>
<tr>
<td>108</td>
<td>3.55%</td>
<td>108</td>
</tr>
<tr>
<td>111</td>
<td>3.68%</td>
<td>2</td>
</tr>
<tr>
<td>120</td>
<td>3.97%</td>
<td>17</td>
</tr>
<tr>
<td>126</td>
<td>4.17%</td>
<td>6</td>
</tr>
<tr>
<td>132</td>
<td>4.37%</td>
<td>9</td>
</tr>
<tr>
<td>135</td>
<td>4.47%</td>
<td>2</td>
</tr>
<tr>
<td>144</td>
<td>4.77%</td>
<td>122</td>
</tr>
<tr>
<td>153</td>
<td>5.07%</td>
<td>1</td>
</tr>
<tr>
<td>156</td>
<td>5.17%</td>
<td>6</td>
</tr>
<tr>
<td>162</td>
<td>5.36%</td>
<td>4</td>
</tr>
<tr>
<td>168</td>
<td>5.56%</td>
<td>3</td>
</tr>
<tr>
<td>183</td>
<td>6.06%</td>
<td>1</td>
</tr>
<tr>
<td>189</td>
<td>6.26%</td>
<td>2</td>
</tr>
<tr>
<td>198</td>
<td>6.56%</td>
<td>1</td>
</tr>
<tr>
<td>204</td>
<td>6.75%</td>
<td>1</td>
</tr>
<tr>
<td>216</td>
<td>7.15%</td>
<td>7</td>
</tr>
<tr>
<td>228</td>
<td>7.55%</td>
<td>1</td>
</tr>
<tr>
<td>234</td>
<td>7.75%</td>
<td>1</td>
</tr>
<tr>
<td>243</td>
<td>8.05%</td>
<td>3</td>
</tr>
<tr>
<td>297</td>
<td>9.83%</td>
<td>1</td>
</tr>
<tr>
<td>324</td>
<td>10.73%</td>
<td>1</td>
</tr>
<tr>
<td>351</td>
<td>11.62%</td>
<td>2</td>
</tr>
<tr>
<td>612</td>
<td>20.26%</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1,486</td>
</tr>
</tbody>
</table>
Table 1.2 (continued)

<table>
<thead>
<tr>
<th>Picas²</th>
<th>Percent of a page</th>
<th># Purchases</th>
<th>Revenue</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Display</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>174</td>
<td>5.76%</td>
<td>9</td>
<td>$18,182</td>
<td>0.30%</td>
</tr>
<tr>
<td>201</td>
<td>6.66%</td>
<td>137</td>
<td>$287,260</td>
<td>4.77%</td>
</tr>
<tr>
<td>208</td>
<td>6.90%</td>
<td>11</td>
<td>$27,192</td>
<td>0.45%</td>
</tr>
<tr>
<td>213</td>
<td>7.05%</td>
<td>11</td>
<td>$21,559</td>
<td>0.36%</td>
</tr>
<tr>
<td>220</td>
<td>7.29%</td>
<td>2</td>
<td>$5,443</td>
<td>0.09%</td>
</tr>
<tr>
<td>225</td>
<td>7.45%</td>
<td>6</td>
<td>$12,650</td>
<td>0.21%</td>
</tr>
<tr>
<td>235</td>
<td>7.79%</td>
<td>126</td>
<td>$320,074</td>
<td>5.32%</td>
</tr>
<tr>
<td>237</td>
<td>7.85%</td>
<td>2</td>
<td>$3,881</td>
<td>0.06%</td>
</tr>
<tr>
<td>246</td>
<td>8.15%</td>
<td>1</td>
<td>$2,180</td>
<td>0.04%</td>
</tr>
<tr>
<td>247</td>
<td>8.19%</td>
<td>14</td>
<td>$36,982</td>
<td>0.61%</td>
</tr>
<tr>
<td>259</td>
<td>8.58%</td>
<td>2</td>
<td>$5,645</td>
<td>0.09%</td>
</tr>
<tr>
<td>262</td>
<td>8.68%</td>
<td>1</td>
<td>$2,281</td>
<td>0.04%</td>
</tr>
<tr>
<td>271</td>
<td>8.98%</td>
<td>2</td>
<td>$5,381</td>
<td>0.09%</td>
</tr>
<tr>
<td>289</td>
<td>9.58%</td>
<td>2</td>
<td>$7,184</td>
<td>0.12%</td>
</tr>
<tr>
<td>301</td>
<td>9.98%</td>
<td>1</td>
<td>$4,082</td>
<td>0.07%</td>
</tr>
<tr>
<td>355</td>
<td>11.76%</td>
<td>126</td>
<td>$320,074</td>
<td>5.32%</td>
</tr>
<tr>
<td>382</td>
<td>12.66%</td>
<td>135</td>
<td>$603,636</td>
<td>10.03%</td>
</tr>
<tr>
<td>391</td>
<td>12.96%</td>
<td>2</td>
<td>$9,715</td>
<td>0.16%</td>
</tr>
<tr>
<td>394</td>
<td>13.05%</td>
<td>19</td>
<td>$93,626</td>
<td>1.56%</td>
</tr>
<tr>
<td>406</td>
<td>13.45%</td>
<td>5</td>
<td>$21,500</td>
<td>0.36%</td>
</tr>
<tr>
<td>409</td>
<td>13.55%</td>
<td>2</td>
<td>$9,565</td>
<td>0.16%</td>
</tr>
<tr>
<td>421</td>
<td>13.95%</td>
<td>1</td>
<td>$3,541</td>
<td>0.06%</td>
</tr>
<tr>
<td>436</td>
<td>14.45%</td>
<td>3</td>
<td>$18,616</td>
<td>0.31%</td>
</tr>
<tr>
<td>517</td>
<td>17.13%</td>
<td>1</td>
<td>$6,743</td>
<td>0.11%</td>
</tr>
<tr>
<td>537</td>
<td>17.76%</td>
<td>2</td>
<td>$8,947</td>
<td>0.15%</td>
</tr>
<tr>
<td>544</td>
<td>18.02%</td>
<td>1</td>
<td>$5,860</td>
<td>0.10%</td>
</tr>
<tr>
<td>549</td>
<td>18.16%</td>
<td>3</td>
<td>$17,062</td>
<td>0.28%</td>
</tr>
<tr>
<td>562</td>
<td>18.62%</td>
<td>1</td>
<td>$7,523</td>
<td>0.13%</td>
</tr>
<tr>
<td>564</td>
<td>18.66%</td>
<td>70</td>
<td>$472,910</td>
<td>7.86%</td>
</tr>
<tr>
<td>576</td>
<td>19.06%</td>
<td>9</td>
<td>$64,538</td>
<td>1.07%</td>
</tr>
<tr>
<td>582</td>
<td>19.25%</td>
<td>1</td>
<td>$7,321</td>
<td>0.12%</td>
</tr>
<tr>
<td>588</td>
<td>19.45%</td>
<td>1</td>
<td>$7,346</td>
<td>0.12%</td>
</tr>
<tr>
<td>600</td>
<td>19.85%</td>
<td>2</td>
<td>$9,653</td>
<td>0.16%</td>
</tr>
<tr>
<td>621</td>
<td>20.55%</td>
<td>1</td>
<td>$8,190</td>
<td>0.14%</td>
</tr>
<tr>
<td>663</td>
<td>21.94%</td>
<td>1</td>
<td>$9,161</td>
<td>0.15%</td>
</tr>
<tr>
<td>735</td>
<td>24.34%</td>
<td>1</td>
<td>$5,872</td>
<td>0.10%</td>
</tr>
<tr>
<td>762</td>
<td>25.23%</td>
<td>24</td>
<td>$221,752</td>
<td>3.69%</td>
</tr>
<tr>
<td>774</td>
<td>25.63%</td>
<td>6</td>
<td>$50,512</td>
<td>0.84%</td>
</tr>
<tr>
<td>783</td>
<td>25.93%</td>
<td>1</td>
<td>$11,214</td>
<td>0.19%</td>
</tr>
<tr>
<td>786</td>
<td>26.03%</td>
<td>1</td>
<td>$10,458</td>
<td>0.17%</td>
</tr>
<tr>
<td>798</td>
<td>26.42%</td>
<td>1</td>
<td>$6,174</td>
<td>0.10%</td>
</tr>
<tr>
<td>978</td>
<td>32.38%</td>
<td>1</td>
<td>$13,385</td>
<td>0.22%</td>
</tr>
</tbody>
</table>
Table 1.2 (continued)

<table>
<thead>
<tr>
<th>Picas²</th>
<th>Percent of a page</th>
<th># Purchases</th>
<th>Revenue</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Display</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1137</td>
<td>37.65%</td>
<td>6</td>
<td>$82,877</td>
<td>1.38%</td>
</tr>
<tr>
<td>1485</td>
<td>49.17%</td>
<td>2</td>
<td>$35,280</td>
<td>0.59%</td>
</tr>
<tr>
<td>1512</td>
<td>50.07%</td>
<td>47</td>
<td>$759,594</td>
<td>12.62%</td>
</tr>
<tr>
<td>1521</td>
<td>50.36%</td>
<td>1</td>
<td>$15,990</td>
<td>0.27%</td>
</tr>
<tr>
<td>1524</td>
<td>50.46%</td>
<td>4</td>
<td>$56,248</td>
<td>0.93%</td>
</tr>
<tr>
<td>1536</td>
<td>50.86%</td>
<td>2</td>
<td>$36,061</td>
<td>0.60%</td>
</tr>
<tr>
<td>1566</td>
<td>51.85%</td>
<td>1</td>
<td>$10,748</td>
<td>0.18%</td>
</tr>
<tr>
<td>1593</td>
<td>52.75%</td>
<td>1</td>
<td>$19,039</td>
<td>0.32%</td>
</tr>
<tr>
<td>1623</td>
<td>53.74%</td>
<td>1</td>
<td>$19,090</td>
<td>0.32%</td>
</tr>
<tr>
<td>1794</td>
<td>59.40%</td>
<td>1</td>
<td>$21,017</td>
<td>0.35%</td>
</tr>
<tr>
<td>3020</td>
<td>99.98%</td>
<td>1</td>
<td>$32,395</td>
<td>0.54%</td>
</tr>
<tr>
<td>3047</td>
<td>100.88%</td>
<td>8</td>
<td>$261,298</td>
<td>4.34%</td>
</tr>
<tr>
<td>3071</td>
<td>101.67%</td>
<td>1</td>
<td>$32,597</td>
<td>0.54%</td>
</tr>
<tr>
<td>3074</td>
<td>101.77%</td>
<td>2</td>
<td>$62,737</td>
<td>1.04%</td>
</tr>
<tr>
<td>3083</td>
<td>102.07%</td>
<td>1</td>
<td>$33,076</td>
<td>0.55%</td>
</tr>
<tr>
<td>3119</td>
<td>103.26%</td>
<td>1</td>
<td>$33,378</td>
<td>0.55%</td>
</tr>
<tr>
<td>3182</td>
<td>105.35%</td>
<td>1</td>
<td>$34,733</td>
<td>0.58%</td>
</tr>
<tr>
<td>3275</td>
<td>108.43%</td>
<td>1</td>
<td>$37,667</td>
<td>0.63%</td>
</tr>
<tr>
<td>3311</td>
<td>109.62%</td>
<td>1</td>
<td>$34,008</td>
<td>0.57%</td>
</tr>
<tr>
<td>6066</td>
<td>200.86%</td>
<td>1</td>
<td>$60,380</td>
<td>1.00%</td>
</tr>
<tr>
<td>6147</td>
<td>203.54%</td>
<td>1</td>
<td>$61,478</td>
<td>1.02%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>714</td>
<td>$4,236,973</td>
<td>70.41%</td>
</tr>
<tr>
<td>Grand Total</td>
<td></td>
<td>6,823</td>
<td>$6,017,632</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Table 1.3: Summary Statistics of the Estimates

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>1,288.30</td>
<td>100.80</td>
<td>61,478.40</td>
<td>3,289.85</td>
</tr>
<tr>
<td>$q$</td>
<td>74.61</td>
<td>1.85</td>
<td>6,229.90</td>
<td>277.53</td>
</tr>
<tr>
<td>$\hat{v}_0$</td>
<td>16.07</td>
<td>0.59</td>
<td>47.81</td>
<td>10.14</td>
</tr>
<tr>
<td>$\theta$</td>
<td>2.41</td>
<td>1.04</td>
<td>13.95</td>
<td>1.69</td>
</tr>
<tr>
<td>$\Delta \hat{U}$</td>
<td>38.23</td>
<td>-59.34</td>
<td>29,863.17</td>
<td>2,278.24</td>
</tr>
<tr>
<td>$\Delta \hat{U}/(\Delta \hat{U} + T)$</td>
<td>0.01</td>
<td>-1.431</td>
<td>0.467</td>
<td>0.445</td>
</tr>
</tbody>
</table>
Table 1.4: Examples of Estimated Firms’ Utility for No Color

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

Table 1.5: Summary Statistics in the Third Degree Discrimination

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>Buying</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 $\hat{\theta}$</td>
<td>4,439</td>
<td>70.96%</td>
<td>2.49</td>
<td>1.04</td>
<td>13.95</td>
<td>1.80</td>
</tr>
<tr>
<td>Group 2 $\hat{\theta}$</td>
<td>1,258</td>
<td>67.71%</td>
<td>2.30</td>
<td>1.04</td>
<td>12.36</td>
<td>1.50</td>
</tr>
<tr>
<td>Group 3 $\hat{\theta}$</td>
<td>526</td>
<td>50.00%</td>
<td>1.04</td>
<td>1.10</td>
<td>8.37</td>
<td>1.09</td>
</tr>
</tbody>
</table>
Fig. 1.1: Function $\hat{v}_0(q)$

![Graph of $\hat{v}_0(q)$](image1)

Fig. 1.2: Function $\hat{\theta}(q)$

![Graph of $\hat{\theta}(q)$](image2)
Fig. 1.3: Estimated Untruncated Density of Type
Fig. 1.4: Estimated $\theta - \frac{1-F(\theta)}{f(\theta)}$

Fig. 1.5: Optimal Price Schedules under Third Degree Discrimination
Fig. 1.6: Optimal Price Schedules under Third Degree Discrimination

Fig. 1.7: Optimal Purchases under Third Degree Discrimination
Chapter 2

Competition and Nonlinear Pricing in Yellow Pages

2.1 Introduction

The optimal way to design a selling mechanism by a seller who faces consumers with unobserved and heterogeneous valuation for her good(s) has been an important subject of study in economics for few decades now. Analyzed under the heading of optimal screening problem, the problem is well understood for a single seller environment when the unobserved heterogeneity is summarized by one dimensional parameter. The question of interested for us is to analyze and empirically quantify the effect of competition on the selling mechanism and their subsequent welfare implication. In their seminal paper, Musa and Rosen (1978) formalize the idea that when facing consumers with unobserved preferences, it is profitable for the seller to offer different varieties of goods at different prices, often leading to nonlinear price schedules. Some other important contribution to this problem include Spence (1977), Maskin and Riley (1984) and Wilson (1993). The fact that we do observe such pricing scheme corroborates the importance of this research for empirical analysis. For instance, telephone companies offer a variety of tariffs that are differentiated by distance and time of use. The advertising rates in newspapers and magazines are based on the size, the placement of the advertisement etc. One significant feature of these tariffs is that the average price paid per unit depends on the size of the good purchased and the tariffs are not strictly proportional to the quantity and hence they are known as the nonlinear pricing or second degree price discrimination.1 Under nonlinear pricing, a menu of options are offered from which consumers can select their choice depending on their preferences. Therefore, allowing for unobserved consumer heterogeneity is important to explain the observed nonlinear price schedules observed in the data. However, there are

1When products differ in terms of quality attributes, a similar interpretation applies, where increment in transaction represents higher quality.
many economic examples where multiple sellers (principals) compete by choosing a nonlinear tariff. It has been well recognized that because a monopoly seller always distorts quantity, socially suboptimal quantities is produced. Moreover, price discrimination can have adverse distributional effects and it can promote allocative inefficiency. Therefore, it is important to know the effect of competition on the distortion, and how the gains (if any) are shared amongst different types of consumers. For such a welfare consideration an empirical study of markets with competition is essential. In this paper we analyze the effect of competition on pricing strategies and welfare in a market with duopoly sellers. To that end, we propose a structural framework based on model that generalizes Ivaldi and Martimort (1994) and estimate the model using data for yellow-page advertisements.

With competition sellers have an added incentives to differentiate their product and make it more attractive to consumers, as differentiation results in competitive advantage. Some important papers that analyze competition between sellers include Oren, Smith and Wilson (1983) who consider a oligopoly market with homogenous goods, and some recent papers analyze markets with differentiated products such as Ivaldi and Martimort (1994), Stole (1995), Armstrong and Vickers (2001), Rochet and Stole (2003), Stole (2007) and Yang and Ye (2008).

Because nonlinear pricing is ubiquitous in various markets and the theory behind it is well understood, it has resulted in an increase in empirical literature. Some early empirical study in this area include Lott and Roberts (1991) who provided an alternative cost based explanation for commonly observed second-degree price discrimination. Others account for cost variation in explaining the nonlinear price schedules, such as Shepard (1991) who quantifies the evidence of price discrimination in self versus full-serve gas station; Clerides (2002) studies hard versus soft cover books and Verboven (2002) studies gasoline and diesel cars. In a recent paper with monopoly setting, Perrigne and Vuong (2009) study a market for yellow page advertising and show that the with utility linear in types both the base and marginal utility function and the distribution of types can be nonparametrically identified. They estimate the model using data from yellow pages advertising and measure the cost of asymmetric information. However, the focus of these papers have been only on a single seller markets. Given the prevalence of multiple seller markets,
empirical investigation of multiple seller markets seems very desirable. There are some recent empirical studies of such markets. This empirical literature can be broadly categorized into three different types: (i) those using a reduced form analysis, such as Borenstein (1991) who shows that fall in competition has resulted in decrease in price margins between leaded and unleaded gas; Borenstein and Rose (1994) show that price dispersion in US airline market increases with competition and Busse and Rysman (2005) study advertising prices in yellow page industry and show that competition results in higher price curvature, i.e. higher quantity discount, (ii) those using random utility discrete choice models to analyze the demand side while taking the supply side as exogenous to estimate structural parameters, such as Leslie (2004) who analyzes tickets sales for a Broadway plays; McManus (2006) who studies an oligopolistic market of specialty coffee and Cohen (2006) who estimates consumer’s preferences over different sizes and brands of paper towels and (iii) those that endogenize the observed price schedules, such as Crawford and Shum (2006) who investigate the magnitude of quality distortion in cable television and Ivaldi and Martimort (1994) who study a duopoly competition model where two principals supply two differentiated goods (electricity and oil products) to diary farms with private valuation for each good. The model solves for the optimal price schedules by aggregating the two-dimensional consumer heterogeneity into a one-dimensional random variable. Miravete and Röller (2004) apply this model to analyze the early U.S. cellular industry. The estimated structural parameters are used to evaluate the effect of competition, the policy change in cellular license awarding and welfare under alternative pricing rules. Because their data do not contain individual customers’ demand, the ability to identify the demand parameters is limited and their policy conclusions are contingent on the validity of those estimates.

This paper presents a structural model of competition in yellow page advertising with incomplete information about the willingness to pay for the two advertisements. Allowing for heterogeneity in preferences for advertisements is important because informative effects of advertising can vary with the advertisers. We collected the data from two sources: the Yellow Page Association and the two phone books. The association provided us the detailed price schedules offered by the publishers and the books provided us with information about the choices made by all the business units; for more detail see section
2. The data suggests that the utility publisher (Verizon) enters first and is the market leader while the non-utility publisher (Ogden) is the follower. Therefore, we consider a Stackelberg duopoly model of competition. Our model allows the firms to advertise with both yellow pages, which is consistent with the observed data and the payoff function for each advertiser is indexed by a two dimensional vector of willingness to pay. A minimum of two dimensional vector is needed because with only one the two advertisements bought must be perfectly correlated but we observe very little correlation between the advertisement bought from them, see Fig. (??). With competition and two dimensional preference parameter, designing optimal nonlinear price schedules becomes a difficult problem. The problem gets accentuated because pooling at equilibrium is unavoidable because of fewer screening instruments than types, making the problem intractable. In the theoretical literature, characterization of optimal price schedules are characterized only under some parametric specification. We solve the multidimensional screening problem with competition by using two pronged approach. First, we focus only on quadratic pricing strategy by the leader and seek to derive the best response of the follower. Second, we use an aggregation method which combines the effect of competition -cheaper advertisement with the one book makes the other less attractive but the degree varies with the willingness to pay- and the multidimensionality into one; for original treatment see Ivaldi and Martimort (1994). An effect of this dimension reduction is that there will be some pooling at the equilibrium, which is consistent with the literature on multidimensional screening, see Rochet and Choène (1998). But because we observe that the follower also offers quadratic price schedules, we explore the condition in the primitive of the model that ensures that the best response to quadratic price schedule is also quadratic. Following the industry norm, both publishers offer an exogenous standard listing for free. Although that quantity is exogenous, the price

---

2With only one dimension, the ranking of a consumer will be the same for both sellers. Then a scatter plot of the advertisement pair would fall on an increasing straight line, in the X-Y plane where coordinates correspond to choice made from each publisher, see also Ivaldi and Martimort (1994).

3For instance, Yang and Ye (2008) analyze a duopoly model with vertical and horizontal differences in preferences where agents can buy from only one seller assuming that horizontal and vertical preferences are independent and uniformly distributed.

4For a robustness check of our results we explore alternative methods that do not rely on the aggregation technique. See Aryal and Huang (2009).
schedules are optimally chosen to exclude some low valuation firms.\textsuperscript{5} The linearity of inverse hazard rate of the “aggregated” type is necessary and sufficient for the optimal price schedules to be quadratic.\textsuperscript{6} It is straightforward to show that a random variable with a linear inverse hazard rate function must be Burr Type XII. This means that the marginal distributions of the two aggregate variables (one for each publisher) belong to the class of Burr Type XII distributions. Another important feature that our model and the econometrics specification incorporate is that both observed price schedules are quadratic in the quantity.

Once we solve for the optimal nonlinear price schedules, we exploit the mapping between the (unobserved) structural parameters of the model and the equilibrium outcomes (observed advertisement choice and payment) to identify the former. The optimal price schedules define the one to one mapping between unobserved types and the choice of advertisement, akin to the bidding strategy in auctions; see Guerre, Perrigne and Vuong (2000). We identify the structure of the model, which includes the utility function and the cost function and the marginal distribution of the “aggregate” variables. Since the data contains information on the choice of all firms in the market, we use the observed pair of advertisements to identify the joint distribution of the preference parameters conditional on the advertisement choices being larger than the free listing. The estimation method is straightforward and uses method of moments to estimate the parameters of the model as well as a kernel estimator to estimate the joint distribution. We analyze the data on advertisements in two yellow pages in Central Pennsylvania, which we gathered from the Yellow Page Association and the two directories. The data is very unique because we have information on all the price schedules offered by both principals and we also have data on the advertisements chosen by each business unit in the market and the amount paid. Such individual consumption data is rarely observed in other markets, such as in telecommunications. Our empirical results show a significant heterogeneity in firms’ willingness to pay for the two advertisements. The estimated model is used to assess

\textsuperscript{5}Armstrong (1996) shows that in multidimensional screening, excluding some demand is optimal.

\textsuperscript{6}Note that this implication is consistent with the fact that, in general, economic theory seldom has implications regarding parametric structures of the distributions but the solution concept used in the economic models typically imply some shape restriction, such as linearity, monotonicity etc. Here, the shape restriction imply linearity of the inverse hazard rate function.
the welfare loss due to (i) asymmetric information, (ii) a merger between the two publishers and (iii) the withdrawal of the non-utility publisher from the market.

The Paper is organized as follows: Section 2 describes the data. Section 3 presents the model, while Section 4 discusses the identification and the estimation of econometric model. Section 5 presents the estimation results and the various counterfactuals. Section 6 concludes. All the omitted proofs are collected in the appendix.

2.2 Nonlinear Pricing in Yellow Pages

In this section we provide the essential features of the data on yellow page advertisements that we collected for Central Pennsylvania (State College and Bellefonte).

2.2.1 The Yellow Page Industry

The data consists of two parts: (i) the price schedules and detailed advertisement options offered by the two publishers and (ii) is the advertisements chosen by each firm (business unit) both for the year 2006-2007.\footnote{An entry in a directory is classified as a business unit if the corresponding phone number is registered as a business phone.} We collected the first part of the data from the Yellow Page Association, which is an industry trading group. The second part of the data (on demand for different advertising categories) is directly constructed from the two directories. The two companies Verizon (the utility publisher) and Ogden Directories Inc. (the non-utility publisher) publish two different yellow page directories. Busse and Rysman (2005) document that the Yellow Pages industry (in the United States) is characterized by competition between asymmetric publishers, as in our case. They document that the data for yellow page advertisements display a nonlinear pricing pattern and find that with competition the publishers offer more quantity discount.

Verizon has been operating in this market for more than two decades while Ogden entered the market in 2000, following a court ruling in 1992 stating copyright protection does not apply to the yellow pages
advertising.\textsuperscript{8} Verizon's directory is slightly bigger with three columns per page while Ogden's directory has only two columns per page. The two also differ in terms of the number of copies distributed: More than 215,400 copies for Verizon while only 73,000 for Ogden. Moreover, the paper quality of Verizon directory is better than that of Ogden's. Such a difference in the distribution channel and the quality of the directories can be the cause of a significant source product differentiation and cost differences. This also suggests that advertising with Verizon might be more effective than Ogden, which is corroborated by the data. For instance, advertisements with Verizon generate a revenue of nearly 6 million dollars, while Ogden generates less than 1 million dollars.

\textbf{Price Schedules}

In terms of size, the Yellow Pages Industry uses three general categories: listing, space listing and display. Listing refers to the name, address and phone number of a firm under appropriate industry heading. A listing is typically a line or two in a column, with possibly different fonts. Both Verizon and Ogden provide the standard listing—smallest font required to list a business name with its phone number and address—for free. Verizon also provides three different fonts with the option to add multiple extra lines to their listing while Ogden provides only two fonts without the possibility of adding extra lines. The space listing allocates a space within the column under the corresponding heading, while a display advertisement provides a listing under the heading and allocates another space somewhere else, which can cover up to two pages. Verizon provides five available sizes within the space listing category, and nine sizes within the display category. Ogden has the same number of sizes within the space listing, while seven sizes within the display category. For the firms who choose display advertisement, Verizon also offers a listing with a particular font for free.

The different size options are measured in picas, which is the unit commonly used in the publishing industry. One pica corresponds approximately to 1/6 of an inch. For example a standard listing in the Verizon's directory is 12 picas, and a full page is 3,020 picas, while for Ogden they are 9 and 1,845

\textsuperscript{8}Similar entries were observed in market in the local telephone service after the 1996 Telecommunications Act in New York.
picas, respectively. In addition to the sizes, advertisers can choose from various colors. Verizon offers five color categories, namely no color, one color, white background, white background plus one color, multiple colors including photos while Ogden offers no color, one color, white background plus one color and multiple colors including photos. These color options are not available for all the sizes. For instance, the multiple colors option is available only with the display advertisement. Table 1 lists a partial price schedule for both directories in terms of advertisement size and color. Three important features emerge from this table. First, color accounts for an important difference in the advertising price for any given size for both directories. For instance, a full-page display advertisement with no color costs $18,510 increasing up to $32,395 with multiple colors in the Verizon directory. Similarly, a full-page display advertisement with no color is $6,324 while the same size advertisement with multiple colors is $9,435 in the Ogden directory. Second, the Verizon’s price is significantly higher than that of Ogden’s across all the comparable advertising options. For instance, a half-page no color display advertisement is $10,093 in Verizon’s directory and only $3,372 in Ogden’s directory. A full-page no color display advertisement is $18,510 in the Verizon directory and $6,324 in the Ogden directory. However, the difference between the two price schedules is not uniform. For example, in the no color category, across the comparable sizes, the Verizon price is on an average 17% higher than Ogden in the listings category, 18% higher in the space listings, and 130% higher in the display category. In another word, the two price schedules are not perfectly correlated. This suggests that the competition is stronger at the lower end of the preferences than at the higher end. This could be because the preferences for the advertisement with Verizon is concentrated at the higher end as compared to that of Ogden. One can then expect to find that the marginal distribution of preferences for advertisement with Verizon is more skewed on the top than that with Ogden. Lastly, the fact that the options provided by both the publishers are comparable while the prices charged vary significantly suggests that the competition directly affects only the price and not the product choices.

Third, for a given color category, we observe that the price per square pica decreases as the advertising size increases in both directories. For instance, in the Verizon directory the unit price

---

9The model we present closely matches this feature of the data.
for a double-page, a full-page, and a half-page display advertisements with no color are $5.68, $6.13 and $6.90 per square picas, respectively. In the Ogden directory, the unit price for a full-page, half-page and a third-page display advertisement with no color is $3.43, $3.68, $3.72 per square picas, respectively. This suggests the presence of some discount for larger quantity as predicted by the curvature of the optimal price schedules. Besides color, there are other qualitative features, namely guide, anchor listing and trade marks etc. For more on the characteristics of the options available with Verizon see Perrigne and Vuong (2009).

The essential features of the observed demand is presented in Table 2, such as the number of purchases and the generated revenue for the three general sizes. For both directories, more than 70% of the revenue accrues from selling display advertisements, which also indicates significant heterogeneity in demand. About 66% of the firms purchase listing in the Verizon’s directory and 14% purchase display ads, while around 94% of firms purchase listing and 3.8% of them buy display in the Ogden directory. Of all the firms, 54 percent purchase advertising from the utility publisher only, 12 percent from both, 2 percent from the non-utility publisher only, and the rest have the standard listings. Summary statistics on the demand pattern is presented in Table 3. It can be seen that some of the popular headings repeat themselves among the four demand choices that can be interpreted as an evidence of heterogeneity of demand even within industries. The average prices paid in each directory by the firms purchasing from both directories are higher than those who purchase from only one directory, which may indicate a higher evaluation of advertising among this group. A similar pattern is observed with respect to advertisement sizes. One question that may arise is whether firms purchasing from both directories have reached the cap in the Verizon one. If this is true, we would observe that all these firms purchase the top category in Verizon’s directory. However, among these firms, roughly 30% purchase listings, 30% purchase space listings, and 30% purchase display ads from Verizon. In the Ogden’s directory this composition is 53%, 21% and 26%, respectively. Finally, the correlation between the ad sizes is 0.32 and the correlation between their price is 0.37. The overall correlation between ad size for all firms is 0.25, between purchasing price is 0.29.

\[10\] This pattern of consumption implies that the environment must be modeled as common agency game and not exclusive agency game, see the section on model for detail.
The lack of strong correlation between purchases from the two directories justifies our choice of a two-dimensional taste parameter for firms (see the model). If indeed the unobserved agents’ heterogeneity is a scalar then (for any price schedule) the chosen bundles will lie on some one dimensional curve. This would then imply strong correlation between the quantities purchased because of the monotonicity in consumptions regarding to the adverse selection parameter. Similar observation is made in Ivaldi and Martimort (1994) regarding their observed demand pattern.

Quality-Adjusted Quantity

In all the theoretical model of nonlinear pricing the quantity is treated as continuous and of single dimension, yet the observed price schedules and product-menu include qualitative features. In this section, we explain why the data suggests that we can use a single dimensional continuous quantity to approximate the observed advertisement options. This method is proposed in Perrigne and Vuong (2009) where we transform the quantity data with quality differences into a single dimensional ‘quality-adjusted quantity.’

From the previous subsection we note that the advertisements differ in their size and other qualitative aspects. In the classic theoretical literature on nonlinear pricing, such as Musa and Rosen (1978) and Maskin and Riley (1984), a monopolist can discriminate consumers by optimal bundling of quality and quantity, both taking continuous values. Maskin and Riley (1984) shows that the optimal bundle of quantity and quality should lie on a unique curve in the quantity-quality space and the optimal quantity allocation should increase with quality along this curve. In the data, we observe that each publisher offers various (finite) qualities for each size of advertisement. Unlike the prediction of Maskin and Riley (1984), we do not observe perfect correlation between the quality and quantity consumption. One solution to this problem would be to index each firm with different taste for quantity and quality. Perrigne and Vuong (2009) argue convincingly that this method is undesirable because it not only leads to a complex multidimensional screening problem which is not only difficult to solve but also is unnecessary because the data suggests that quality is not used to screen advertisers. Instead, the technological constraints (on

\[11\] In other words it is restrictive to claim that knowledge of preference for one good determines exactly the taste for other.
publishing) limit the advertisement sizes to be discrete and leave some scope of further discrimination. Allowing for catalog of various quantity-quality pairs fills up the holes in the size scale. As a result, the number of different options to choose from increases and can be roughly assumed to be continuous. There are two properties of the price-quantity data that favor such an argument. First, there is some curvature in the price schedule suggesting that discounts are offered for large advertisements while no such discounts are observed in the quality dimension. Second, we also observe that the ratio of the (marginal) prices for two different qualities remain the same across sizes, for more on Verizon see Perrigne and Vuong (2009). Thus the data supports the argument that various quality dimensions are not used by the publishers to discriminate the firms, thereby corroborating the choice as a good approximation of the true data generating process.

Then we construct a quality-adjusted quantity index by considering the price schedule for multi colors as the continuous price schedule then adjusting the advertising size for other color options in view of the price schedule for the multi colors. If we let \( q_i \) be the multi color size sold by publisher \( i \), then the sizes for all other colors will be assigned lower values by using a fitted nonlinear function. For example, one page in one color will correspond to smaller quantity but with multi colors. The fitted quadratic functions (explained in the model below) are

\[
\hat{T}_1(q_{j1}) = 1512 + 11.27q_{j1} - 0.00027q_{j1}^2,
\]

\[
\hat{T}_2(q_{j2}) = 103 + 6.25q_{j2} - 0.00066q_{j2}^2,
\]

where \( T_{ji} \) is the price in dollars for publisher \( i \), and \( q_{ji} \) is the advertisement size measured in square picas purchased by firm \( j \). All the coefficients are estimated using an ordinary least squares estimator. The \( R^2 \) of such regression is above 0.99 for both regressions and all the coefficients are significant at the 1% level. The quality-adjusted quantities are constructed by plugging, on these curves all the observed prices for other quantities and we then plug on these curve all the observed prices for other qualities to solve
for the quality-adjusted quantities. One key observation is that the tariff functions are both quadratic, which plays a vital role in our entire analysis.

2.3 The Model

In this section we extend Ivaldi and Martimort (1994) duopoly competition in nonlinear pricing model to that of Stackelberg duopoly with optimal exclusion. We characterize optimal nonlinear pricing in which one publisher (first entrant) chooses its price schedule, and the second publisher (second entrant) chooses its own price schedule after observing the choice of his competitor. Firms are heterogeneous in terms of their valuations for advertisements and are indexed by a pair \((\theta_1, \theta_2)\) of willingness to pay for both the advertisements. These valuations are unknown to the publishers and each publisher discriminates these firms by offering a nonlinear tariff. When a \((\theta_1, \theta_2)\) firm chooses \((q_1, q_2)\) its utility/gain is

\[
U(q_1, q_2; \theta_1, \theta_2) = \theta_1 q_1 - \frac{b_1}{2} q_1^2 + \theta_2 q_2 - \frac{b_2}{2} q_2^2 + cq_1 q_2 - T_1(q_1) - T_2(q_2) \tag{C.1}
\]

with \(b_i > 0, i = 1, 2\). For the utility function to be concave, we assume that \(b_1 b_2 - c^2 > 0\). Since the gross marginal utility of \(q_1\), i.e. \(MU_i(q_1, q_2; \theta_1, \theta_2) = q_i - b_i q_i + c q_j\), everything else being the same, higher \(\theta_i\) implies higher marginal utility and vice versa. We assume that \((\theta_1, \theta_2) \sim F(\cdot, \cdot)\) on \([\theta_1, \bar{\theta}_1] \times [\theta_2, \bar{\theta}_2]\) with a joint density \(f(\cdot, \cdot) > 0\). If \(c > 0\) \((c < 0)\) then \((q_1, q_2)\) are complements (substitutes). When they are complements, marginal utility of \(q_i\) increases with \(q_j\), and hence the demand for \(q_i\) should be an increasing function of \(q_j\). Because we do not observe such positive dependence in the data, we focus only on \(c < 0\). The cost function \(C(\cdot)\) for publisher \(i\) is assumed to be linear

\[
C_i(q_i) = K_i + m_i q_i, \tag{C.2}
\]

with \(K_i \geq 0\) and \(m_i > 0\).

In view of the data, we consider Stackelberg-Nash equilibrium: given the menu (product range and price schedule) chosen by P1, P2’s menu maximizes its expected total profit, and conditional on the fact
that P2 always best responds to P1’s choice, P1’s choice maximizes its expected total profit. One key feature of the market is that it is a norm in the industry to offer \((q_{10}, q_{20})\) for free. This quantity is known as the basic listing and corresponds to the smallest space required to print the name of the business and hence every firm has its name and phone number listed in both the books. This option can be treated as outside option for a firm, whose valuation is determined by its type.

For any choice of \(T_1(\cdot)\) and \(T_2(\cdot)\) a \((\theta_1, \theta_2)\) firm’s behavior is determined by the following two first order conditions:

\[
(\theta_1 - b_1 q_1 + c q_2 - T'_1(q_1))(q_1 - q_{10}) = 0; \quad (\theta_2 - b_2 q_2 + c q_1 - T'_2(q_1))(q_2 - q_{20}) = 0.
\]

Implementability of any nonlinear tariff must take into account these equations. However, without any restriction on \(T_1(\cdot)\) and \(T_2(\cdot)\) characterizing nonlinear pricing at the equilibrium is difficult. Therefore, following Ivaldi and Martimort (1992), we make a simplifying assumption that P1 offers quadratic price schedule:

Assumption 1 The leading firm chooses a quadratic tariff

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

where \(\gamma_1 > 0, \alpha_1 > 0\) and \(\beta_1 < 0\).

We do not restrict \(T_2(\cdot)\) to be quadratic, but in view of the quadratic price schedule observed in the data, it is desirable that \(T_2(\cdot)\) be quadratic too. However, it is not always the case that the best response to a quadratic nonlinear tariff is also quadratic. We find a condition, which we explain later, on the primitive of the model that is necessary and sufficient for the best response \(T_2(q_2)\) to be quadratic. The usefulness of this condition is two folds: first it allows to formulate a model that matches the data, and second it also simplifies our identification and estimation procedure. Using this condition, one can show

\[12\] For detailed analysis see Aryal (2009).
that offering quadratic price schedule by P1 is also optimal conditional on the fact that P2’s best reply
is quadratic.

2.3.1 The Demand Pattern

This section characterizes the type space that generates the patterns of demand observed in the data,
viz. \((q_{10}, q_{20}), (q_1, q_{20}), (q_{10}, q_2)\) and \((q_1, q_2)\) with \(q_1 > q_{10}\) and \(q_2 > q_{20}\) corresponding to the subsets of
the space \([\theta_1, \bar{\theta}_1] \times [\theta_2, \bar{\theta}_2]\), respectively as \(C_0, C_1, C_2\) and \(C_b\) ; see Figure (2.1).

Case 1:

The set \(C_0\) is defined from the optimality condition \(\frac{\partial U(q_{10}, q_{20}, \theta_1, \theta_2)}{\partial q_i} \leq 0\) for \(i = 1, 2\). Using (C.3) they
can be simplified to \(\theta_1 - b_1 q_{10} + c q_{20} \leq \alpha_1 + \beta_1 q_{10}\) and \(\theta_2 - b_2 q_{20} + c q_{10} \leq T'_2(q_{20})\).\(^{13}\) We denote by
\((\theta_1^*, \theta_2^*)\) the type of firms for which \((q_{10}, q_{20})\) is the first best choice and are

\[
\theta_1^* = \alpha_1 + (b_1 + \beta_1) q_{10} - c q_{20} \quad \text{(C.4)}
\]

\[
\theta_2^* = T'_2(q_{20}) + b_2 q_{20} - c q_{10}. \quad \text{(C.5)}
\]

Since the marginal utility is increasing in type, any firm with type \((\theta_1, \theta_2) \ll (\theta_1^*, \theta_2^*)\) demands \((q_{10}, q_{20})\).

Case 2:

\(C_1\) contains types satisfying \(\theta_1 - b_1 q_1 + c q_{20} = \alpha_1 + \beta_1 q_1\) and \(\theta_2 - b_2 q_{20} + c q_{10} \leq T'_2(q_{20})\). The first
equality gives \(q_1 = \frac{\theta_1 - \alpha_1 + c q_{20}}{b_1 + \beta_1}\), which together with the second inequality determines the threshold type
\(\theta_2^{**}\) for whom \(q_2 > q_{20}\) is the first best and for all \(\theta_2 \leq \theta_2^{**}\) it is optimal to demand \(q_{20}\). Since the marginal
utility from \(q_2\) depends on the choice of \(q_1\), this threshold type is a function of \(\theta_1\) and is

\[
\theta_2^{**} = \left( b_2 - \frac{c^2}{b_1 + \beta_1} \right) q_{20} + T'_2(q_{20}) + \frac{c \alpha_1}{b_1 + \beta_1} - \frac{c}{b_1 + \beta_1} \theta_1. \quad \text{(C.6)}
\]

\(^{13}\)We assume that \(T_i(\cdot)\) is right differentiable at \(q_{i0}\) for \(i = 1, 2\).
Case 3:

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

$$\theta_1^{**} = \left(b_1 + \beta_1 - \frac{c^2}{b_2}\right)q_{10} + \alpha_1 + \frac{c}{b_2}T'_2(q_2) - \frac{c}{b_2}\theta_2.$$  \hfill (C.7)

Case 4:

$C_b$ is determined by

$$\theta_1 - b_1q_1 + cq_2 = \alpha_1 + \beta_1q_1, \hfill (C.8)$$

$$\theta_2 - b_2q_2 + cq_1 = T'_2(q_2), \hfill (C.9)$$

The set where these equalities hold is the set of all type-pairs that do not belong to any of the previous sets.

2.3.2 The Follower’s Problem

The follower observes $T_1(\cdot)$ given by (C.3) and responds by proposing its own tariff. Suppose $P_2$ observes a firm choosing some $q_2$, then in equilibrium, $P_2$ can infer the choice of $q_1$ by this firm using (C.8), namely

$$q_1 = \frac{\theta_1 - \alpha_1 + cq_2}{b_1 + \beta_1}, \hfill (C.10)$$

if $\theta_1 > \theta^*$ and $q_1 = q_{10}$, otherwise. Using (C.10) in (C.9) gives the necessary condition that determines the choice of $q_2$, namely

$$\theta_2 + \frac{c\theta_1}{b_1 + \beta_1} = \frac{c\alpha_1}{b_1 + \beta_1} + \left(\frac{b_2 - \frac{c^2}{b_2}}{b_1 + \beta_1}\right)q_2 + T'_2(q_2). \hfill (C.11)$$
Let \( z_2 = \theta_2 + \frac{c\theta_1}{b_1 + \beta_1} \), then the optimal \( q_2 \) is determined by (C.11) with \( z_2 \) replacing the left hand side.

For P2, \( z_2 \) is unobserved and is exogenously determined, and hence can be treated as the (unobserved) preference of a firm for \( q_2 \). As \( z_2 \) captures the firm’s taste for both products and its reaction to the competitor’s contract, knowing \( z_2 \) is equivalent to knowing \((\theta_1, \theta_2)\) for P2 because any contract that depends on both \( \theta_1 \) and \( \theta_2 \) can be transformed into a payoff equivalent contract that depends only on \( z_2 \); see Ivaldi and Martimort (1992) for a formal treatment. Hence \( z_2 \) is a sufficient statistic for \((\theta_1, \theta_2)\).  

Then because \( z_2 \) is one dimensional, the characterization of optimal nonlinear pricing can then be formulated as a single principal selling to agents with one dimensional private information: \( z_2 \sim G_2(\cdot) \) with the density \( g_2(z_2) = \int_{\theta_2}^{\theta_2^*} f \left( \theta_1, z_2 - \frac{c\theta_1}{b_2 + \beta_2} \right) d\theta_1 \) on \([z_2, \infty)\]. This method of aggregation is a standard solution technique in the multidimensional screening literature.

Let \( z_0^2 \) be the threshold type below which firms choose \( q_{20} \). It is determined by (5) and (6) (cases 1 and 2) and is given by

\[
\begin{align*}
z_0^2 &= \begin{cases} 
\theta_2^* + \frac{c\theta_1}{b_1 + \beta_1}, & \theta_1 \leq \theta_1^* \\
\theta_2^{**} + \frac{c\theta_1}{b_1 + \beta_1}, & \theta_1 \geq \theta_1^* 
\end{cases}
\end{align*}
\]

allowing us to interpret this principal-agent problem as one with optimal exclusion. The P2’s optimization problem can be written as

\[
\max_{T_2(\cdot), q_2(\cdot), z_0^2} E\Pi_2 = \int_{z_2^0}^{\tau_2} \left( T_2(q_2(z_2)) - m_2q_2(z_2) \right) g_2(z_2) dz_2 - K_2 - m_2q_{20}G_2(z_2^0), \tag{C.12}
\]

subject to the appropriate IC (truth-telling) and IR (participation) constraints. Because the net surplus generated by a trade between any firm and P2 depends also on \( q_1 \), we disaggregate the total surplus for any firm into two parts to isolate the net addition that is controlled only by P2, irrespective of \( q_1 \). Let

\[\text{14}\] Observe that for a fixed \( \theta_2 \) and \( \beta_1 \), higher \( \theta_1 \) implies lower \( z_2 \), which means firms that value \( q_1 \) more are considered lower types by P2 and vice versa.

\[\text{15}\] See Aryal and Huang (2009) for a different but more general method of solving this problem.
Let \( W_2(\theta_1, z_2) \) be the indirect utility for a firm with type \((\theta_1, z_2)\).

\[
W_2(\theta_1, z_2) = \max_{q_1 \geq q_{10}, q_2 \geq q_{20}} \left[ u(q_1, q_2; \theta_1, z_2 - \frac{c\theta_1}{b_1 + \beta_1}) - T_1(q_1) - T_2(q_2) \right].
\]

Let \( w_2(\theta_1, z_2) \) be the indirect utility if the same firm demands optimal amount of \( q_1 \) while consuming only \( q_{20} \)

\[
w_2(\theta_1, z_2) = \max_{q_1 \geq q_{10}} \left[ u(q_1, q_{20}; \theta_1, z_2 - \frac{c\theta_1}{b_1 + \beta_1}) - T_1(q_1) \right]
= -\gamma_1 + \left( \frac{\theta_1 - \alpha_1}{2(b_1 + \beta_1)} \right)^2 + \left( z_2 - \frac{c\alpha_1}{b_1 + \beta_1} \right) q_{20} + \left( \frac{\gamma^2}{2(b_1 + \beta_1)} - \frac{b_2}{2} \right) q_{20}^2,
\]

where the second equality follows from (C.10). Note that \( w_2(\cdot, \cdot) \) is the part of indirect utility that accrues solely from the trade between a firm and P1 and next we disaggregate \( W(\theta_1, z_2) \) as a sum of \( w_2(\theta_1, z_2) \) and a some (yet) unknown function depending only on \((q_2, z_2)\). Since

\[
W_2(\theta_1, z_2) = \max_{q_1 \geq q_{10}, q_2 \geq q_{20}} \left[ \theta_1 q_1 - \frac{b_1}{2} q_1^2 + \left( z_2 - \frac{c\theta_1}{b_1 + \beta_1} \right) q_2 - \frac{b_2}{2} q_2^2 + cq_1 q_2 - T_1(q_1) - T_2(q_2) \right]
= \left( z_2 - \frac{c\theta_1}{b_1 + \beta_1} \right) q_{20} - \frac{b_2}{2} q_{20}^2 + cq_1 q_{20} - \left( z_2 - \frac{c\theta_1}{b_1 + \beta_1} \right) q_{20} + \frac{b_2}{2} q_{20}^2 - cq_1 q_{20}
\]

where the third equality follows from the definition of \( w_2(\theta_1, z_2) \) and (C.10). We have

\[
s_2(q_2, z_2) = \max_{q_2 \geq q_{20}} \left\{ \left( z_2 - \frac{c\alpha_1 - c^2 q_2}{b_1 + \beta_1} \right) (q_2 - q_{20}) - \frac{b_2}{2} (q_2^2 - q_{20}^2) - T_2(q_2) \right\}.
\]
Thus $s_2(q_2(z_2), z_2) = s_2(z_2)$ is the net addition to the indirect utility (net surplus) from consuming the optimal $q_2 > q_{20}$. From (C.13) we get

$$T(z_2) = \left( z_2 - \frac{c_0 - c^2 q_2}{b_1 + \beta_1} \right) (q_2 - q_{20}) - \frac{b_2}{2} (q_2^2 - q_{20}^2) - s_2(z_2). \quad (C.14)$$

The incentive compatibility constraint requires that $s_2(q_2(z_2); z_2) \geq s_2(q_2(\tilde{z}_2); z_2), \forall z_2, \tilde{z}_2 \in [\underline{z}_2, \overline{z}_2]$, and it can be shown that $s_2(\cdot)$ is continuous, convex and satisfies the envelope condition

$$s'_2(z_2) = q_2(z_2) - q_{20} \quad \forall z_2 \in (\underline{z}_2, \overline{z}_2]. \quad (C.15)$$

Equations (C.14) and (C.15) show that $P_2$ can be assumed to choose the firm’s surplus (residual rent) $s_2(z_2)$. If $s_2(z_2)$ is implementable by some price schedule $T_2(q_2)$ and the utility function is convex in type then $s_2(z_2)$ must be convex. Rochet (1987) showed that in the quasi linear utility, the convexity of $s_2(z_2)$ is also sufficient for implementability, where $s_2^+ = \lim_{z_2 \downarrow z_2^{-}} s_2(z_2)$:

**Lemma 1:** (Rochet (1987)) The global IC constraint is satisfied if and only if the following conditions hold:

(a) (Envelope Condition): $s_2(z_2) = \int_{z_2}^{z_2^+} (q_2(t) - q_{20})dt + s^+, \forall z_2 \in [\underline{z}_2, \overline{z}_2]$. 

(b) $s_2(\cdot)$ is convex or equivalently $q_2(z_2)$ is increasing in $z_2$.  

Suppressing the dependence on type, the participation (IR) constraint is given by

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

---

\textsuperscript{16}Note that if $q_2 = q_{20}$, there should be no addition, which is confirmed because $s_2(q_{20}, z_2) = 0$.

\textsuperscript{17}An advantage of using dual approach is that it not only makes finding optimal allocation rule easier (using Euler-Lagrange equation) but because implementability requires that the choice be convex, which then guarantees that the quantity allocation rule is continuous in the agent’s type - a corollary of Envelope theorem.

\textsuperscript{18}For $q_2$ to be global optimal, we must have $-b_2 + \frac{c^2}{b_1 + \beta_1} - T_2''(q_2) < 0$. From the FOC we get

$$1 = 2 \left( b_2 - \frac{c^2}{b_1 + \beta_1} + T_2''(q_2) \right) \frac{dq_2}{dz_2},$$

and hence $\frac{dq_2}{dz_2} > 0 \iff -b_2 + \frac{c^2}{b_1 + \beta_1} - T_2''(q_2) < 0$. Hence the strictly increasing allocation rule is necessary and sufficient condition for global optimal.
for all \((\theta_1, z_2)\).\(^{19}\) Since, by definition \(w_2 \geq 0\), the IR constraint becomes \(s_2(z_2) \geq 0\).

Taking into account (13)-(15) and Lemma 1, the P2’s problem can be rewritten as

\[
\max_{q(\cdot), z_2^0, s_2^+} E \Pi_2 = \int_{z_2^0}^{z_2} \left\{ z_2 - \frac{c \alpha_1 - c^2 q_2(z_2)}{\beta_1 + b_1} \right\} (q_2(z_2) - q_{20}) - \frac{b_2}{2} (q_2^2(z_2) - q_{20}^2) - m_2 q_2(z_2) - s_2^+ \\
-(q_2(z_2) - q_{20}) \frac{1 - G_2(z_2)}{g_2(z_2)} \right\} g_2(z_2) dz_2 - K_2 - m_2 q_{20} G_2(z_2^0),
\]

\[(C.16)\]

\[s.t \quad q_2'(\cdot) > 0 \quad s_2(\cdot) \geq 0.\]

\[(C.17)\]

To solve the above problem, we consider the relaxed problem where we drop the constraints and check ex post that they are indeed satisfied. Since \(s_2(\cdot)\) is increasing and the optimal allocation rule must be increasing in \(z_2\), ensuring \(s_2(z_2^0) = 0\) is sufficient for (IR) constraint to be satisfied for all \(z_2 \in (z_2^0, z_2^\bar{z})\). It is immediate to see that \(s_2^+ = 0\) is optimal. Existence of a unique solution for this optimization problem is guaranteed by a theorem that relies on Rochet and Choène (1998). The proof can be found in Aryal and Huang (2009).\(^{20}\) **Theorem 1:** Under our maintained assumption on preferences and cost and assuming \(g_1(\cdot)\) has full support, i.e. there exists \(\epsilon > 0\) such that \(g_1(z_1) \geq \epsilon\) for all \(z_1 \in [\bar{z}, \bar{z}]\), there exists a unique solution to the problem (C.12). The Optimal allocation rule \(q_2(\cdot)\) is characterized by point-wise maximization of the expected profit function and is formalized below.

**Proposition 1:** Under the assumption that the reverse hazard rate of \(G_2(\cdot)\) is decreasing on \((z_2^0, z_2^\bar{z})\) and \(b_2 > \frac{2c^2}{\beta_1 + \beta_2}\), and let \(H_2(z_2) \equiv \frac{1 - G_2(z_2)}{g_2(z_2)}\), we have

---

\(^{19}\)We are implicitly assuming that the utility of not participating in any of the two contracts (outside utility) is independent of type and is normalized to 0. See Jullien (2000) for more on adverse selection models with type dependent outside options.

\(^{20}\)The note also shows how one can use Rochet and Choène (1998) to solve this bi-dimensional screening model without using the aggregation method used here.
I. The optimal function $q(\cdot)$ and optimal $z_2^0$ as

$$q_2(z_2) = \begin{cases} \frac{z_2 - H_2(z_2) - m_2 - \frac{c^2 q_{20} + c \alpha_1}{b_1 + \beta_1}}{b_2 - \frac{c^2}{b_1 + \beta_1}}, & \forall z_2 \in (z_2^0, \bar{z}_2] \\ q_{20}, & \forall z_2 \in [\bar{z}_2, z_2^0] \end{cases}$$ (C.18)

$$z_2^0 - H_2(z_2^0) = (b_2 - \frac{c^2}{b_1 + \beta_1})q_{20} + m_2 + \frac{c \alpha_1}{b_1 + \beta_1}.$$ (C.19)

II. $T_2(q)$ must satisfy (C.14) such that the corresponding Ramsey rule for the price schedule is

$$\frac{T'_2(q_2(z_2)) - m_2}{T'_2(q_2(z_2))} = \frac{1 - G_2(z_2)}{g_2(z_2)} \frac{1}{\partial q_2}.$$ (C.20)

We can use (C.20) to analyze the effect of competition on the markup. The markup is smaller if either the distribution of $z_2$ is skewed toward the lower end, i.e. $(1 - G(z_2))$ is smaller or if $\frac{\partial s_2(q_2(z_2))}{\partial q_2}$ is higher. For instance a higher absolute $\beta_1$ (i.e. higher discount) implies a higher marginal rent with respect to $q_2$ which in turn implies decreasing markup.

Furthermore, because the empirical implementation of the model relies on the curvature of the tariff function, we are interested in characterizing the form of $T_2(q_2)$. Specifically, we are interested in finding (sufficient) conditions on the primitive of the model such that the best response by P2 is quadratic. To this end, from (C.18), we get

$$h_2(z_2) \equiv z_2 - H_2(z_2) = q_2 \left( b_2 - \frac{2c^2}{b_1 + \beta_1} \right) + m_2 + \frac{c^2 q_{20} + c \alpha_1}{b_1 + \beta_1} \equiv \tau_2(q_2) \quad \forall z_2 \in (z_2^0, \bar{z}_2].$$
Because $H_2(z_2)$ is decreasing in $z_2$, we get $z_2 = h_2^{-1}[\tau_2(q_2)]$. Substituting this into the pricing rule in (C.14), we get

$$T_2(q_2) = \left( h_2^{-1}(\tau_2(q_2)) - \frac{c\alpha_1 - c^2 q_2}{b_1 + \beta_1} \right) (q_2 - q_{20}) - \frac{b_2}{2} (q_2^2 - q_{20}^2) - \int_{q_{20}}^{q_2} (t - q_{20}) \frac{1}{h_2'(h_2^{-1}(\tau_2(t)))} \tau_2'(t) dt$$

$$= \left( h_2^{-1}(\tau(q_2)) - \frac{c\alpha_1 - c^2 q_2}{b_1 + \beta_1} \right) (q_2 - q_{20}) - \frac{b_2}{2} (q_2^2 - q_{20}^2)$$

$$- [h_2^{-1}(\tau_2(t))(t - q_{20})]_{q_{20}}^{q_2} + \int_{q_{20}}^{q_2} h_2^{-1}(\tau_2(t)) dt$$

$$= \int_{q_{20}}^{q_2} h_2^{-1}(\tau_2(t)) dt - \frac{c\alpha_1 - c^2 q_2}{b_1 + \beta_1} (q_2 - q_{20}) - \frac{b_2}{2} (q_2^2 - q_{20}^2).$$

(C.21)

Therefore $T_2(q_2)$ is quadratic if and only if the first term in the right hand side of (C.21) is quadratic which is equivalent to the integrand being linear.

**Lemma 2:** The P2’s best response to a quadratic pricing rule of P1 is also quadratic if and only if $h_2^{-1}(\tau_2(t)))$ is linear in $t$.

In the following lemma we show that $h_2^{-1}(\tau(t))$ is linear if and only if the distribution function of $z_2$ is $G_2(z_2) = 1 - (1 - c_2 + \xi z_2^2)^{\rho_2}$ with $\rho_2 > 0$. Suppressing the index, the distribution is completely characterized by $\varsigma, \xi, \rho$, such that $G(\bar{z}) = 1$ and $G(z) \geq 0$.

**Lemma 3:** Let $G(x)$ be the distribution function of a random variable $x$ with a non vanishing density $g(x)$, on $[x, \bar{x}]$. Then $H(x) = \frac{1 - G(x)}{g(x)}$ is linear in $x$ if and only if $G(x) = 1 - \left( \frac{x - \bar{x}}{\bar{x} - x} \right)^\rho$, for $\rho > 0$.\(^{21}\)

Henceforth, $z_2$ Burr Type XII defined on $[\bar{z}_2, \bar{z}]$ with distribution function

$$G_2(z_2) = 1 - \left( \frac{\bar{z}_2 - z_2}{\bar{z}_2 - \bar{z}} \right)^{\rho_2} \rho_2 > 0.$$  

(C.22)

\(^{21}\)In the statistical literature, this type of distribution function is known as Burr Type XII or log-logistic distribution. Ivaldi and Martimort (1992) assume that the marginal distribution of $z_2$ is Burr Type XII without providing a justification for it.
When we estimate the model, we shall exploit this model prediction. Such a model feature has also been used by Miravete and Roller (2004). It follows that $H_2(z_2) = \frac{1}{\rho_2} \bar{z}_2 - \frac{1}{\rho_2} z_2$ and $h^{-1}(\cdot) = \frac{1}{1+\rho_2} \bar{z}_2 + \frac{\rho_2}{1+\rho_2}$.

We write the optimal $T_2(q_2)$ in a quadratic form, namely, $T_2(q_2) = \gamma_2 + \alpha_2 q_2 + \beta_2 q_2^2$. Let $\zeta_2 = \frac{1}{1+\rho_2} \bar{z}_2$ and $l_2 = \frac{\rho_2}{1+\rho_2}$ then:

**Proposition 2:** The coefficients of $P2$’s optimal pricing rule $T_2(q_2)$ are given by

$$
\alpha_2 = \zeta_2 + l_2 m_2 + (l_2 - 1) \left( \frac{c^2 q_2 + c \alpha_1}{b_1 + \beta_1} \right) \quad \text{(C.23)}
$$

$$
\beta_2 = \left( b_2 - \frac{2c^2}{b_1 + \beta_1} \right) (l_2 - 1). \quad \text{(C.24)}
$$

The proof is by substituting the expression of $h^{-1}(\cdot)$ into (C.21).

### 2.3.3 The leader’s problem

In this subsection we characterize the (optimal) nonlinear pricing for the leader (P1). In the previous section we have shown that if $z_2$ is Burr Type XII and $T_1(\cdot)$ is quadratic then the optimal tariff $T_2(\cdot)$ is quadratic. Following the backward induction argument, we want to characterize the optimal $T_1(q_1)$ given the continuation strategy (best response path) of P2. P1 chooses an optimal allocation rule $q_1(\cdot)$ and a corresponding price schedule $T_1(\cdot)$, which are linked via the indirect utility function as in (14) for P2. The optimal allocation rule $q_1(\cdot)$ can be determined irrespective of the shape of $T_1(\cdot)$. This suggests that the necessary and sufficient condition for $T_1(\cdot)$ to be quadratic is exactly the same as in Lemma 3, i.e. linearity of inverse hazard rate function. However, conditional on a quadratic form, the choice of three parameters that characterize the price schedule is not as direct as in Proposition 2. The problem arises because, unlike $z_2$ which was exogenous for P2 as it depends only on $\beta_1$, a similar sufficient statistic

---

**22**This suggests that the competition i.e. the presence of a competing seller affects the quantity provision or choice, only through the price schedule, which in turn is affected by competition. Similar phenomenon was observed by Borenstein and Rose (1994) who find higher price dispersion in the US airline industry with higher competition and by Busse and Rysman (2005) who show that a stronger competition is associated with larger curvature i.e. higher price discounts.
$z_1$ (defined below) for P1 is a function of $\beta_2$, which in turn is a function of $\beta_1$. Hence, $z_1$ is no longer exogenous for P1. In view of this observation we solve the problem as follows: First, conditional on $T_1(\cdot)$, we determine the optimal allocation rule $q_1(\cdot)$ following the same steps as for $q_2(\cdot)$. We then determine the parameters of $T_1(\cdot)$, (endogeneizing $z_1$) and $T_2(\cdot)$, following the method used by Wilson (1993).

Recall that firms with type $\theta_1 \geq \theta_1^{**}$ choose $q_1 > q_{10}$ and while those with $\theta_1 \leq \theta_1^{**}$ choose $q_{10}$. Using (9) and $T_2(\cdot)$, we express the optimal $q_2$ as a function of $q_1$ as

$$q_2(q_1; \theta_1, \theta_2) = \frac{\theta_2 - \alpha_2 + c q_1}{b_2 + \beta_2},$$

(C.25)

such that $q_2(q_1; \theta_1, \theta_2) = q_{20}$ for $\theta_2 \leq \theta_2^{**}$. We can rewrite the firm’s optimal demand function (8) as a function of $q_1$ only and is given by $\theta_1 + \frac{c b_2}{b_2 + \beta_2} - b_1 q_1 + c \left[ \frac{c q_1 - \alpha_2}{b_2 + \beta_2} \right] q_1 = \alpha_1 + \beta_1 q_1$. Let $z_1 = \theta_1 + \frac{c b_2}{b_2 + \beta_2}$, then the optimal $q_1$ as a function of $z_1$ satisfies

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

(C.26)

Let $W_1(z_1, \theta_2)$ be the indirect utility function and $w_1(z_1, \theta_2)$ the indirect utility when $q_1 = q_{10}$ while the optimal $q_2$ is given by (23). Similar to the derivation of (C.13) we can express $W_1(\cdot, \cdot)$ as a sum of $w_1(\cdot, \cdot)$ and the extra (residual) utility from consuming any extra $q_1$ in excess of $q_{10}$. In particular

$$W_1(z_1, \theta_2) = w_2(z_1, \theta_2) + s_1(q_1, z_1),$$

where $s_1(q_1, z_1) = \max_{q_1 \geq q_{10}} \left( z_1 + \frac{c^2 q_1 - \alpha_2}{b_2 + \beta_2} \right) (q_1 - q_{10}) - b_1 \left( \frac{q_1^2 - q_{10}^2}{2} \right) - \gamma_1 - \alpha_1 q_1 - \frac{\gamma_1}{2} q_1^2$. The function $s_1(q_1(z_1), z_1) \equiv s_1(z_1)$ is the relevant rent function that a $z_1$ receives when it chooses $q_1$ and pays $T_1(q_1)$.

From the definition of $s_1(\cdot)$ we can express the tariff as

$$T_1(q_1) = \left( z_1 + \frac{c^2 q_1 - \alpha_2}{b_2 + \beta_2} \right) (q_1 - q_{10}) - b_1 \left( \frac{q_1^2 - q_{10}^2}{2} \right) - \int_{q_{10}}^{q_1} (w(t) - q_{10}) dt - s_1^+,$$

(C.27)
which allows us to write the maximization problem of P1 as

\[
\max_{q_1(\cdot), z_1, s_1^0} E \Pi_1 = \int_{z_1^0}^{z_1} \left[ \left( z_1 + \frac{c^2 q_1 - c (\zeta_2 + l_2 m_2 + \frac{c \alpha_1 (l_2 - 1)}{b_1 + \beta_1})}{b_2 + \frac{c^2 (l_2 - 1)}{b_1 + \beta_1}} \right) (q_1 - q_{10}) - \frac{b_1}{2} (q_1^2 - q_{10}^2) \right] dz_1 - (q_1 - q_{10}) H_1(z_1) - K_1 - m_1 q_1 \int g_1(z_1) d z_1 - m_1 G_1(z_1^0) q_{10} \right]
\]

\[s.t \quad q_1(\cdot) > 0, s_1(\cdot) \geq 0.\]  

(C.28)

When using cases (2) and (3) \(z_1^0\), is given by\(^{23}\)

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

The optimal quantity allocation rule for every type \(z_1\), given \(\alpha_1, \beta_1\), can be determined by the usual point-wise maximization with respect to \(q_1\). This gives us the optimal quantity allocation rule as a function of type \(z_1\) and opponent’s pricing rule:

**Proposition 3** The optimal quantity allocation rule (contract) is given by

\[
q_1(z_1) = \begin{cases} 
z_1 - H_1(z_1) - m_1 - \frac{c \alpha_2 + c^2 q_{10}}{b_2 + \beta_2} & \forall z_1 \in (z_1^0, z_1] 
q_{10} & \forall z_1 \in [z_1, z_1^0] \n\end{cases}
\]

(C.30)

and the optimal price schedule satisfies

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

To verify ex post, that the quantity allocation is increasing in type, \(z_1\), check that the sufficient condition is that \(H_1(z_1)\) be non increasing in \(z_2\), which is satisfied in our case. The optimal quantity is

\(^{23}\)At \(\theta_2^*\), we get \(z_1^0 = \alpha_1 + (b_1 + \beta_1) q_{10} + \frac{c \alpha_2 - c^2 q_{10}}{b_2 + \beta_2} \).
determined for a particular price schedule. Now, we are left to determine the price schedule. Note that in the optimization of P1 the boundaries, of the integration in profit function, are also a function of the price schedule. If we maximize $E\Pi_1$ with respect to $\alpha_1$ and $\beta_1$, the calculation becomes cumbersome and extremely lengthy. To reduce the difficulty, we adopt a slightly different substitution method originally due to Wilson (1993). If the pricing rule $T_1(q_1)$ is well behaved then the quantity allocation $z_1 \to q_1(z_1)$ is unique and regular (sufficiently many times differentiable): $q_1 = q_1(z_1) \Leftrightarrow \frac{\partial T_1(q_1)}{\partial q_1} = \Psi(q_1, z_1)$. Then we can make a change of variable in (27) to rewrite the profit function as

$$E\Pi_1 = \mathcal{E}_{q_1}(T_1(q_1) - m_1 q_1) g_1(T_1(q_1)) T_1^\prime(q_1) dq_1 - m_1 G_1(z_1^0) q_{10} - K_1,$$

where $G_1(z_1^0) = \Pr(\text{selling } q_{10})$. The optimal price is then determined by choosing $\alpha_1$ and $\beta_1$. The advantage of using this method is that we no longer have to keep track of the dependence on $z_1$ explicitly as long as we use optimal quantity allocation rule (28). In what follows we shall make this calculation explicit and determine the optimal pricing rule. The P1 optimal behavior is characterized by choosing $q_1(\cdot)$ such that

$$q_1(z_1) = \arg \max_{q_1} s_1(q_1; z_1).$$

The first order condition with respect to $q_1$: $\frac{ds_1(z_1)}{dq_1} = 0$ gives $z_1 = \Psi(q_1)$ where

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

and hence $\frac{dz_1}{dq_1} = \Psi'(q_1) dq_1 = \left\{ \left( \frac{2c^2}{b_2 + \beta_2} - b_1 \right) - T_1''(q_1) \right\}$. Then the expected profit can be rewritten by (using the fact that $q_1(\cdot)$ is increasing) as

$$E\Pi_1 = \int_{q_{10}}^{\bar{q}_1} (T_1(q_1) - C_1(q_1)) g_1(\Psi(t)) \Psi'(t) dt - m_1 G_1(z_1^0) q_{10} - K_1,$$

$$= \int_{\bar{q}_1}^{\bar{q}_1} (T_1(t) - C_1(t)) G_1'(\Psi(t)) dt - m_1 G_1(z_1^0) q_{10} - K_1.$$
Then integrating by parts, the first term in the right hand side becomes $\int_{q_{10}}^{q_1} (T_1'(t) - m_1)(1 - G_1(\Psi(t))) \, dt - m_1 q_{10} - K_1$ which uses $T_1(q_{10}) = 0$, $C_1(q_{10}) = m_1 q_{10}$ and $G_1(\Psi(\bar{q})) = 1$. The term in the second parenthesis is $1 - G_1(\Psi(t)) = \left( \frac{z_1 - \Psi(t)}{z_1 - z_1^*} \right)^{\beta_1}$. Then using the form of $T_1(q_1)$ we get

$$E\Pi_1 = \int_{q_{10}}^{q_1} (\alpha_1 + \beta_1 t - m_1) \left( \frac{z_1 - \Psi(t)}{z_1 - z_1^*} \right)^{\beta_1} \, dt - m_1 q_{10} - K_1. \tag{C.31}$$

Now the objective is to choose $\alpha_1$ and $\beta_1$ that maximize the expected profit. We have already found the optimal quantity allocation rule above, which shall be used to determine the rest of the parameters of the contract. Roughly speaking, we have 3 first order conditions and we need to determine 3 parameters, $\{q_1, \alpha_1, \beta_1\}$. The first one is already found as a function of the latter two. Since remaining parameters are the same for all qualities purchased they are independent of types $z_1$, and hence can be determined from point-wise maximization of $\Pi_1$. It is important to observe that not only $q_1$ but also $z_1$, $z_1^*$ and $\Psi(t)$ are functions of $\alpha_1$ and $\beta_1$. Since we never use the explicit form of these coefficients in our identification or estimation the structural equation characterizing these coefficients of the optimal pricing rule are collected in the appendix in Proposition (3.4).

To complete the characterization, we have to determine the fixed price to purchase advertisements that are greater than $q_{i0}$, $i = 1, 2$. Each principal will choose $\gamma_i$ so that the lowest type of firm’s participation constraint is binding. Any extra utility resulting from interaction between the two advertisements is extracted by the principal. To that end, we first determine the lowest type for each publisher and choose $\gamma_i$ such that the total utility is zero for that type. Doing that gives optimal $\gamma_1$ and $\gamma_2$ as shown in Proposition (3.5) in the appendix.
2.4 Identification and Estimation

2.4.1 Identification

In this section we study the identification problem of the model, which concerns the possibility of drawing inferences from the observed data on advertisement bought and the prices paid to the theoretical structure outlined above. Failure to identify the model structure implies that the data lacks sufficient information to distinguish between alternative structures. Let, $(Q, T_1(\cdot), T_2(\cdot))$ be the random vector representing the demand for advertisement bought and the transfer made whose joint distribution function belongs to the set $\Theta$.\footnote{The observables include the menus of advertisements offered by each publisher, the corresponding prices for each option $(T_1(\cdot), T_2(\cdot))$, the minimum quantities given for free $(q_{10}, q_{20})$, every firm’s advertising purchases from both publishers $(q_{1j}, q_{2j})$, where $j$ indexes the firms.} A structure $M$ is then a set of hypothesis which implies a unique distribution $\Psi(\cdot)$ which is observed.\footnote{Since we observe the advertisement chosen by each advertiser, $\Psi(\cdot)$ can be determined uniquely from the data and is an estimation problem, as discussed in the next subsection.} The set of all structures that are a priori possible is called the model and is denoted by $M := \{\mathcal{F}, X\}$, the family of joint distribution of types $F(\cdot, \cdot)$ and the set of other parameters $[m_1, m_2, K_1, K_2, b_1, b_2, c]$. By definition, there is a unique distribution function of $(Q, T_1(\cdot), T_2(\cdot))$ associated with each structure. The identification problem is to show that there is a unique inverse association between $M$ and $\Theta$.\footnote{In our definition of $\Theta$ we have ignored the tariffs} Two structures in $M$ are said to be observationally equivalent if they imply the same probability distribution for the observable random variables $(Q, T_1(\cdot), T_2(\cdot))$. So, we say that the structure $M \in M$ is identified if there is no other structure in $M$ which is observationally equivalent. As of now, the sets $M$ and $\Theta$ are too general, and in order to address the identification problem we need to be more specific about them. To this end, we begin with some minimal conditions that each of the sets should satisfy.

Definition: Let $M$ be the model satisfying

1. Let $\mathcal{F}$ be the set of joint distributions defined on a compact support $[\underline{\theta}_1, \overline{\theta}_1] \times [\underline{\theta}_2, \overline{\theta}_2]$ such that

   (a) Every $F \in \mathcal{F}$ is absolutely continuous with $f(\cdot, \cdot) > 0$ as the joint density.
(b) \( f_i(\cdot) = \int f(\theta_i, \theta_j) d\theta_j \) is continuously differentiable.

2. \( X \) satisfies

(a) \( b_1 b_2 - c^2 > 0. \)

(b) \( b_i + \beta_i > 0, \ i = 1, 2 \)

(c) \( (b_1 + \beta_1)(b_2 + \beta_2) - 2c^2 > 0. \)

These inequalities are sufficient conditions for the utility function to be concave, and the firms’ optimization problem to be convex. Because we focus on the joint distribution of \( G(\cdot, \cdot) \) of \((z_1, z_2)\) instead of \( F(\cdot, \cdot) \) we also need to make assumptions on the former. **Definition** Let \( \mathcal{G} \) be the set of joint distributions \( G(\cdot, \cdot) \) satisfying

1. \( G(\cdot, \cdot) \) is a joint distribution defined on a compact support \([z_1, \bar{z}_1] \times [z_2, \bar{z}_2] \).

2. The joint density \( g(\cdot, \cdot) \) is continuously differentiable in the support and \( g(\cdot, \cdot) > 0 \) on its support.

3. There exists \( \rho_i > 0 \) such that \( g_i(\cdot) \) for \( i = 1, 2 \) takes the form:

\[
g_i(t) = \begin{cases} \frac{\rho_i}{(z_i - \bar{z}_i)} \left( \frac{z_i - t}{z_i - \bar{z}_i} \right)^{\rho_i - 1}, & \text{if } t \in [z_i, \bar{z}_i] \\ 0, & \text{otherwise.} \end{cases}
\]

Thus, we shall consider the model \( \mathcal{M}' = [\mathcal{G}, X] \in \mathcal{M}' \).

**Identification of Cost Parameters**

Note that the fixed cost parameters \( K_1 \) and \( K_2 \) do not enter any of the equilibrium outcome of the model they are not identified.\(^{27}\) We know that for \( q_2(\cdot) \) to be truthfully implementable, it has to be strictly increasing in \( z_2 \). Let \( \bar{q}_2 \) be the highest quantity sold by Ogden. Then by definition we know

\(^{27}\)A similar result is found in Perrigne and Vuong (2009).
\( \bar{q}_2 = q_2(\hat{z}_2) \). Then in the part two of the proposition, we evaluate (20) at \( \bar{z}_2 \). Since \( T'_2(q_2(\bar{z}_2)) = \alpha_2 + \beta_2 \bar{q}_2 \) and \( 1 - G_2(\bar{z}_2) = 0 \) gives

\[
m_2 = \alpha_2 + \beta_2 \bar{q}_2. \tag{C.32}
\]

Each tariff function \( T_i(q_i) \) is observed and is completely characterized by \( \{ \gamma_i, \alpha_i, \beta_i \} \) for \( i = 1, 2 \), the latter are known. Hence, (C.32) identifies the marginal cost \( m_2 \).\(^{28}\) Let, \( z_i(\cdot) = q_i^{-1}(\cdot) \) for \( i = 1, 2 \), which given the monotonicity of \( q_i(\cdot) \) is well defined. Let \( \Psi(\cdot, \cdot) \) be the joint distribution of \( (q_1, q_2) \) conditional on \( q_i > q_{i0}, i = 1, 2 \), which is identified from the data using \( Q \). We can define the conditional marginal distribution \( \Psi_i(q_i) = \int \Psi(q_i, q_j) dq_j, i, j \in \{1, 2\}, i \neq j \). We note the following relationship

\[
\Psi_i(q_i) = \Pr[ q_i \leq q_i | q_i > q_{i0} ] = \Pr[ z_i \leq z_i(q_i) | z_i > z_i(q_{i0}) ] = \frac{G_i(z_i) - G_i(z_{i0})}{1 - G_i(z_{i0})}
\]

and \( \psi_i(q) = \partial \Psi_i(q) / \partial q_i = (g_i(z_i) z_i'(q))/ (1 - G_i(z_{i0})) \), the conditional density of \( q_i \). This allows us to express the inverse hazard rate of \( z_i \) in terms of observable as

\[
\frac{1 - G_i(z_i)}{g_i(z_i)} = \frac{1 - \Psi_i(q)}{\psi_i(q)} z_i'(q), \quad i = 1, 2. \tag{C.33}
\]

Using the optimal \( q_i(\cdot) \) in \( T'_i(q_i) \) for \( i = 1 \) and \( i = 2 \) gives

\[
T'_i(q_i) = z_i - \frac{ca_j}{b_j + \beta_j} - z_i + \frac{1 - G_i(z_i)}{g_i(z_i)} + m_i - \frac{c^2 q_{i0}}{b_j + \beta_j} + \frac{c^2 q_{i0}}{b_j + \beta_j} - \frac{ca_j}{b_j + \beta_j} = m_i + \frac{1 - \Psi_i(q)}{\psi_i(q_i)} z_i'(q_i). \tag{C.34}
\]

\(^{28}\)Unlike \( m_2 \), the identification of \( m_1 \) is not straightforward owing to the complicated expression for \( \alpha_1 \) and \( \beta_1 \).
Differentiating $T_i(\cdot)$ with respect to $q_i$ in (28) and (14) for $i = 1, 2$ respectively and solving for $z'_i(q_i)$ gives

$$z'_i(q_i) = T''_i(q_i) + b_i - \frac{2c^2}{b_j + \beta_j},$$

which can be used in (C.34) leading to

$$T'_i(q_i) = m_i + \frac{1 - \Psi_i(q_i)}{\psi_i(q_i)} \left( T''_i(q_i) + b_i - \frac{2c^2}{b_j + \beta_j} \right), \quad \forall q_i \in [q_{i0}, \bar{q}_i],$$

(C.35)

which is very important for identification. Evaluating (C.35) for $i = 1$ at $q_1 = q_1(\tau_1)$ gives

$$m_1 + \frac{1 - \Psi_1(\bar{q}_1)}{\psi_1(\bar{q}_1)} \left( \beta_1 + b_1 - \frac{2c^2}{b_2 + \beta_2} \right) = m_1 = \alpha_1 + \beta_1 \bar{q}_1$$

(C.36)

since $\Psi_1(\bar{q}_1) = 1$.\(^{29}\)

**Identification of $b_1, b_2$ and $c$**

Making use of the optimality condition for the highest type, $(\tau_1, \tau_2)$, the coefficients of the utility function are identified. The parameters $(b_1, b_2)$ are the indices of the degree of product differentiation between the two advertisements. For instance, if $b_1 = b_2 = 0$ then the firm will only care about the consumption level of all varieties ($q_1 + q_2$) because then the utility is linear in $q_i$ for any fixed $q_j$. With $b_i > 0$, the marginal utility from $q_i$ falls and the firm will give increasing importance to the combination of the pair $(q_1, q_2)$. In other words, keeping everything else the same, a firm with higher $b_i$ would consume higher $q_j$. In this respect, firms of type $\tau_i$, $i = 1, 2$ are those who consume the largest of $q_i$ and lowest of $q_j$, hence $b_i$ must be very small to rationalize this behavior. Indeed, note that $\tau_i$ or $(\bar{\tau}_i, \bar{\theta}_j)$ is also the type that is the most valued by the publisher $P_i$ while the corresponding $z_j$ is at most $z^0_j$ and hence is the lowest type for the publisher $P_j$; this type demands $(\bar{q}_i, q_{j0})$. Then the optimal allocation rule $q_i(\cdot)$ implies that for this type, marginal utility from consuming the pair $(q_i, q_{j0}), j \neq i$ must be equal to the

\(^{29}\)Note that evaluation (C.35) for $i = 2$ at $q_2$ gives $m_2 = \alpha_2 + \beta_2 \bar{q}_2$, which is the same as in (C.32).
marginal cost at that pair. This implies that for both \(i = 1, 2\) and \(j \neq i\)

\[
\bar{\theta}_i - b_i \bar{q}_i + cq_{j0} = \alpha_i + \beta_i \bar{q}_i,
\]

leading to

\[
c = \frac{\alpha_i + (b_i + \beta_i) \bar{q}_i - \bar{\theta}_i}{q_{j0}}. \tag{C.37}
\]

This implies that \(c\) is (over) identified conditional as we can use either \(i = 1\) or \(i = 2\) to identify \(c\). To show that \(b_1\) and \(b_2\) are identified we begin with any pair \((q_1, q_2)\) such that \(q_i \neq \bar{q}_i\) and rewrite (C.35) as

\[
\alpha_i + \beta_i q_i = m_i + \frac{1 - \Psi_i(q_i)}{\psi_i(q_i)} \left( \beta_i + b_i - \frac{2c^2}{b_j + \beta_j} \right), \tag{C.38}
\]

for \(i = 1, 2\) and for \(i = 1\) get

\[
b_1 + \beta_1 = \frac{\alpha_1 + \beta_1 q_1 - m_1}{\frac{1 - \Psi_1(q_1)}{\psi_1(q_1)}} + \frac{2c^2}{b_2 + \beta_2}. \tag{C.39}
\]

Now for \(i = 1\) and another \(\tilde{q}_1 \neq q_1\), (C.38) for \(i = 1\) gives

\[
\frac{2c^2}{b_2 + \beta_2} = \frac{\alpha_1 + \beta_1 \tilde{q}_1 - m_1}{\frac{1 - \Psi_1(q_1)}{\psi_1(q_1)}} - (b_1 + \beta_1),
\]

which with (C.39) gives

\[
b_1 = \frac{1}{2} \left( \frac{\alpha_1 + \beta_1 q_1 - m_1}{\frac{1 - \Psi_1(q_1)}{\psi_1(q_1)}} + \frac{\alpha_1 + \beta_1 \tilde{q}_1 - m_1}{\frac{1 - \Psi_1(q_1)}{\psi_1(q_1)}} \right) - \beta_1, \tag{C.40}
\]

thus identifying \(b_1\). Then, (C.40) can be used to replace \((b_1 + \beta_1)\) in (C.38) for \(i = 2\) to get

\[30\text{From (C.37) it is clear that to rationalization of larger } \overline{q}_i \text{ requires smaller } b_i, \text{ everything else the same.} \]
\[ b_2 = \frac{\alpha_1 + \beta_1 q_2 - m_2}{1 - \Psi_2(q_2)} - \beta_2 + \frac{4c^2}{\alpha_1 + \beta_1 q_2 - m_2} \cdot \frac{\alpha_2 + \beta_2 \tilde{q}_2 - m_2}{1 - \Psi_2(q_2)} - \frac{\alpha_1 + \beta_1 q_2 - m_2}{1 - \Psi_2(q_2)}. \]  

(C.41)

Now, for another \( \tilde{q}_2 \), (C.38) for \( i = 2 \) gives

\[
\frac{4c^2}{2(b_1 + \beta_1)} = \frac{\alpha_2 + \beta_2 \tilde{q}_2 - m_2}{1 - \Psi_2(q_2)} - (b_2 + \beta_2)
\]

\[
4c^2 = \left( \frac{\alpha_1 + \beta_1 q_1 - m_1}{1 - \Psi_1(q_1)} + \frac{\alpha_1 + \beta_1 \tilde{q}_1 - m_1}{1 - \Psi_1(q_1)} \right) \left( \frac{\alpha_2 + \beta_2 \tilde{q}_2 - m_2}{1 - \Psi_2(q_2)} - (b_2 + \beta_2) \right),
\]

where the second equality follows from (C.40), which can then be used in (C.41) to solve for

\[
b_2 = \frac{\alpha_1 + \beta_1 q_2 - m_2}{1 - \Psi_2(q_2)} - \beta_2 + \left( \frac{\alpha_2 + \beta_2 \tilde{q}_2 - m_2}{1 - \Psi_2(q_2)} - (b_2 + \beta_2) \right)
\]

\[= \frac{1}{2} \left( \frac{\alpha_1 + \beta_1 q_2 - m_2}{1 - \Psi_2(q_2)} + \frac{\alpha_2 + \beta_2 \tilde{q}_2 - m_2}{1 - \Psi_2(q_2)} \right) - \beta_2,
\]

(C.42)

thus identifying \( b_2 \). With \( b_1 \) and \( b_2 \) identified, \( c \) is identified too and is given as the negative root of

\[ c = \pm \frac{1}{2} \left[ \left( \frac{\alpha_1 + \beta_1 q_1 - m_1}{1 - \Psi_1(q_1)} + \frac{\alpha_1 + \beta_1 \tilde{q}_1 - m_1}{1 - \Psi_1(q_1)} \right) \left( \frac{\alpha_2 + \beta_2 \tilde{q}_2 - m_2}{1 - \Psi_2(q_2)} - (b_2 + \beta_2) \right) \right]. \]

(C.43)

**Identification of \( G_i(\cdot) \)**

In the model, we exploit the shape of the observed price schedules, to define the parametric class in which the marginal distribution of \( z_i \) belongs. To recall, we show that the optimal price schedules are quadratic if and only if \( G_i(\cdot) \) is a Burr Type XII. Therefore, \( G_i(\cdot) \) is characterized by three parameters \( \{\tilde{z}_i, \pi_i, \rho_i\} \), which we identify below. For each \( z_i, \rho_i, i = 1, 2 \), is a measure of shape (skewness/kurtosis),
i.e. it measures how high the type firms are concentrated relative to the low type firms. For instance when $\rho_i = 1$, $z_i$ has a uniform distribution. When $\rho_i > 1$, $z_i$ is more concentrated at low values. When $0 < \rho_i < 1$, $z_i$ is more concentrated at high values. From (C.24) we get

$$\rho_2 = -\left(1 + \frac{1}{\beta_2} \left(b_2 - \frac{2c^2}{b_1 + \beta_1}\right)\right)$$

and hence $\rho_2$ is identified.

Since $\beta_2$ is the curvature of $T_2(\cdot)$, the concavity of the price schedules helps to recover $\rho_2$. Let $r_i = \Pr(q_i = q_{i0})$ the fraction of firms choosing $q_{i0}$ only for $i = 1, 2$, then

$$r_i = \Pr(z_i \leq z_i^0) = 1 - \left(\frac{z_i - z_i^0}{z_i - z_i^0}\right)^{\rho_i}.$$  \hspace{1cm} (C.44)

Recall that we have the following relationships

$$z_i^0 = \frac{z_i}{1 + \rho_i} + \frac{\rho_i}{1 + \rho_i} \left(\left(b_i - \frac{c^2}{b_j + \beta_j}\right)q_{i0} + m_i + \frac{ca_j}{b_j + \beta_j}\right)$$  \hspace{1cm} (C.45)

$$z_i^0 = \alpha_i + (b_i + \beta_i)q_{i0} + \frac{ca_j - c^2q_{i0}}{b_j + \beta_j},$$  \hspace{1cm} (C.46)

for $i, j \in \{1, 2\}$ and $j \neq i$, where (C.45) is determined by the optimal contract while (C.46) is determined from the demand side. Specifically $z_i^0$ is identified as the value for which the optimal consumption is exactly $q_{i0}$, i.e. $q_i(z_i)\big|_{z_i = z_i^0} = q_{i0}, i = 1, 2$. From (C.46) we get

$$z_i^0 - \alpha_i - \beta_iq_{i0} = \left(b_i - \frac{c^2}{b_j + \beta_j}\right)$$

and using it in (C.45) gives

$$z_i - z_i^0 = \rho_i(\alpha_i + \beta_iq_{i0} - m_i),$$  \hspace{1cm} (C.47)
and from (C.44) we get
\[(z_i - z_i) = \frac{\rho_i(\alpha_i - \beta_iq_{i0} - m_i)}{(1 + r_i)} \frac{1}{\rho_i}, i = 1, 2. \tag{C.48}\]

Now, to identify \(z_i\) we use the property of the optimal allocation rule, that \(q_i(z_i) = \bar{q}_i\). Thus, evaluating \(q_i(\cdot)\) at \(z_i\) in the equations in (18) and (30) for \(i = 2, 1\) respectively we get
\[z_i = \bar{q}_i\left(\frac{b_i - 2c^2}{b_j + \beta_j} + m_i + \frac{c^2q_{i0} + \alpha_j}{b_j + \beta_j}\right). \tag{C.49}\]

It is now easy to see that \(z_i\) for both \(i = 1, 2\) are identified using (C.48) and (C.49) and is
\[z_i = \bar{q}_i\left(\frac{b_i - 2c^2}{b_j + \beta_j} + m_i + \frac{c^2q_{i0} + \alpha_j}{b_j + \beta_j} - \frac{\rho_i(\alpha_i - \beta_iq_{i0} - m_i)}{(1 + r_i)} \frac{1}{\rho_i}\right). \tag{C.50}\]

Therefore the support of both \(G_1(\cdot)\) and \(G_2(\cdot)\) are identified. Using (C.46) for \(i = 1\) in (C.45) and we can solve for \(\rho_1\) to get
\[\rho_1 = \frac{z_i - \left(\alpha_1 + (b_1 + \beta_1)q_{i0} + \frac{c(\alpha_2 - c^2q_{i0})}{b_2 + \beta_2}\right)}{(\alpha_1 + \beta_1q_{i0} - m_1)}, \tag{C.51}\]

thus identifying \(\rho_1\).

### 2.4.2 Estimation

Our estimation is based on the equilibrium strategies of Section 2. More specifically, we observe \((q_1, q_2)\); \(j = 1, 2, .., 6328\). We assume that these purchases are the outcomes of the model equilibrium in (17) and (30). We define our econometric model accordingly as

\[q_{ij}(z_{ij}) = \begin{cases} \frac{1 + \rho_i}{b_i - \frac{\beta_i}{\alpha_i + \beta_i}} z_{ij} - \frac{1}{b_i - \frac{\beta_i}{\alpha_i + \beta_i}} \left(\frac{1}{\rho_i} z_i + m_i + \frac{c^2q_{i0} - \alpha_i}{b_j + \beta_j}\right), & \forall z_{ij} \in (z_i^0, \bar{z}_i] \\ q_{i0}, & \forall z_{ij} \in [\bar{z}_i, z_i^0] \end{cases} \tag{C.52}\]
where $i = 1, 2, j = 1, 2, ..., 6328$. The pair $(z_1, z_2)$ is the source of randomness in the econometric model. Besides the above two optimal purchase equations, we have six structural equations defining the optimal price schedules, namely $\alpha_i, \beta_i, \gamma_i, i = 1, 2$ defined in propositions (3.4) and (3.5). These six equations give additional restrictions on the structural parameters.

We assume that every firm $j$ draws $(\theta^1_j, \theta^2_j)$ independently from $F(\cdot, \cdot)$. Given the tariffs choice of the two publishers, every $(\theta^1_j, \theta^2_j)$ determines a pair $(z^1_j, z^2_j)$, distributed with $G(\cdot, \cdot)$. The estimation procedure takes several steps. In the first step, the quantity sold by each publisher is separately used to estimate the nonparametrically inverse hazard rate $(1 - \Psi_i(\cdot)) / (\psi_i(\cdot))$ for $i = 1, 2$ using standard kernel estimator. In the second step, using the estimated inverse hazard rate and (C.52) in addition to the constraints we use Generalized Method of Moments to estimate all the parameters of the model. Once these parameters are estimated, we use a quantile transformation of (C.52) to estimate the joint distribution of $G(\cdot, \cdot)$, where the estimation method is close to nonparametric estimation of copula functions.

**Estimation of Parameters**

The parameters that we are interested in estimating is $\{b_1, b_2, c, m_1, m_2, \rho_1, \rho_2, \tilde{z}_1, \tilde{z}_2, \tilde{z}_1, \tilde{z}_2\}$. Let $N_i$ be the number of firms buying $q_i > q^0_i, i = 1, 2$. Thus we can estimate $\tilde{q}_i$ by $\max_{j=1,2,...,N_i} q_j$ for $i = 1, 2$. The natural estimator of $m_i$ is $\hat{m}_i = \hat{\alpha}_i + \hat{\beta}_i \hat{q}_i$, $i = 1, 2$. Since the estimated parameters should be such that $\tilde{z}_i$ must satisfy (C.52) at $\tilde{q}_i$ for $i = 1, 2$, we restrict $\hat{\tilde{z}}_i$ (the estimator for $\tilde{z}_i$) to be given by (C.52) evaluated at $\tilde{q}_i$, and other remaining parameters. Then we are left with 7 parameters to estimate. Let $\Lambda \subset \mathbb{R}^t$ be the compact set of parameters of the structural model. We use $\lambda \in \Lambda$ to denote a generic element of this set, where $\lambda_0$ is the true parameter. We estimate the parameters using a generalized methods of moment (GMM). We match eight moments, $E(q^k_i) = t_k(\lambda_0), k = 1, 2, 3, i = 1, 2$ and the last two using $r_i$ (probability of a firm choosing $q^0_i$). The estimator is obtained by replacing the population moments and solving for $\hat{\lambda}$, i.e $E(q^k_i - t_k(\lambda)) = 0$. Let $H^1_1(\gamma) \in \mathbb{R}^4$ and $H^2_2(\gamma) \in \mathbb{R}^4$ be the two sets of moments corresponding to $q_1$ and $q_2$, respectively. We assume that the population moment conditions
are satisfied uniquely at $\gamma_0$, i.e.

$$E[H^*(\gamma_0)] = E \begin{bmatrix} H_1^*(\gamma_0) \\ H_2^*(\gamma_0) \end{bmatrix} = 0.$$  

From the econometric model (C.52) we have $q_{ij} = A_i z_{ij} - B_i$. Therefore for $i = 1, 2$ we can use the following conditional moment conditions:

$$E(q_i | q_i > q_{i0}) - A_i E(z_i | z_i > z_{i1}^0) + B_i = 0$$

$$E(q_i^2 | q_i > q_{i0}) - A_i^2 E(z_i^2 | z_i > z_{i1}^0) - 2A_i B_i E(z_i | z_i > z_{i1}^0) + B_i^2 = 0$$

$$E(q_i^3 | q_i > q_{i0}) - A_i^3 E(z_i^3 | z_i > z_{i1}^0) + 3A_i^2 B_i E(z_i^2 | z_i > z_{i1}^0) - 3A_i B_i^2 E(z_i | z_i > z_{i1}^0) + B_i^3 = 0$$

$$E(r_i - Pr(z_{ij} \leq z_{i1}^0)) = 0,$$

where

$$E(z_i | z_i > z_{i1}^0) = z_{i1}^0 + \frac{\pi_i - z_{i1}^0}{\rho_i + 1}$$

$$E(z_i^2 | z_i > z_{i1}^0) = (z_{i1}^0)^2 + \frac{2(\pi_i - z_{i1}^0)}{\rho_i + 1} - \frac{2(\pi_i - z_{i1}^0)^2}{\rho_i + 2}$$

$$E(z_i^3 | z_i > z_{i1}^0) = (z_{i1}^0)^3 + \frac{3(\pi_i - z_{i1}^0)}{\rho_i + 1} - \frac{6(\pi_i - z_{i1}^0)^2}{\rho_i + 2} + \frac{3(\pi_i - z_{i1}^0)^3}{\rho_i + 3}$$

$$Pr(z_i \leq z_{i1}^0) = 1 - \left(\frac{\pi_i - z_{i1}^0}{\pi_i - z_i}\right)^{\alpha_i}$$

### Estimating the Joint Distribution

Our final objective is to estimate the joint distribution $G(\cdot, \cdot)$. We exploit the fact that we observe a pair of $(q_{1j}, q_{2j})$ for each firm $j$. It might be informative to recall the data generating process: Every firm $j$ draws $(\theta_{1j}, \theta_{2j})$ from $F(\cdot, \cdot)$, which given $(T_1(\cdot), T_2(\cdot))$ determines $(z_1, z_2)$ and hence the choice $(q_1, q_2)$. Thus unless $\theta_1$ is independent of $\theta_2$, the observed quantities $q_1$ and $q_2$ are not independent.

We use the two allocation rules given in (C.52) to estimate a pseudo pair $(\hat{z}_{1j}, \hat{z}_{2j}), j = 1, 2, \ldots, N$ by
inverting (C.52) that corresponds to each observed pair \((q_{1j}, q_{2j})\). We then use a nonparametric kernel estimator to estimate the joint distribution. We ignore the censoring at \(z_i^0, i = 1, 2\) (for the discussion). The joint densities must have both the marginal densities as Burr Type XII. Instead of implementing a constrained Kernel estimator, we use a quantile transformation to estimate the joint distribution. The idea is very simple, and is as follows: For every pair \((\hat{z}_{1j}, \hat{z}_{2j})\) as estimated above, we determine a quantile pair \((\hat{\alpha}_{1j}, \hat{\alpha}_{2j}) := (\hat{G}_1(\hat{z}_{1j}), \hat{G}_2(\hat{z}_{2j}))\). These quantile are uniformly distributed on \([0, 1]\). Then, we use Kernel estimator on these pairs \((\hat{\alpha}_{1j}, \hat{\alpha}_{2j}), j = 1, \ldots, N\) to estimate their joint distribution. The joint distribution \(G(\cdot, \cdot)\) can be recovered using

\[
G(z_1, z_2) = C(G_1(z_1), G_2(z_2)),
\]

where \(C(\cdot, \cdot)\) is the joint distribution of the transformed data, namely a bivariate distribution whose marginal distributions are uniform. We use the kernel method to estimate \(C(\cdot, \cdot)\) nonparametrically. More specifically, we use the new bivariate kernel estimator that corrects the boundary bias proposed by Gijbels and Meilniczuk (1990). See also Chen and Huang (2007). We need to correct for the boundary bias because for \((\alpha_1, \alpha_2)\) close to the boundary the uncorrected kernel estimator puts substantial mass outside the unit square, and hence the estimate is not consistent in the boundary. To correct this problem, Gijbels and Meilniczuk (1990) propose a correction that is based on mirror image modification method. In our two dimensional case, the method consists in reflecting each data point with respect to all edges and corners of the unit square and building a kernel estimate based on the enlarged data set. The additional data then put the mass of the kernel back into the unit square.

Let \(K(\cdot)\) be a symmetric kernel on \([-1, 1]\) and \(\tilde{G}_{\alpha,h}(x) = \int_{-\infty}^{x} K(t) dt\) be the distribution of \(K(\cdot)\). We use product Triweight kernel of type \(K(x_1, x_2) = \prod_{i=1}^{2} K_i(x_i)\) where

\[
K_i(x_i) = \begin{cases} 
\frac{35}{32}(1 - x_i^2)^3, & |x_i| \leq 1 \\
0, & \text{otherwise.}
\end{cases}
\]
Then an estimator of the joint density \( g(z_1, z_2) \) is given by

\[
\hat{g}(z_1, z_2) = \frac{\partial^2 C(G_1(z_1), G_2(z_2))}{\partial z_1 \partial z_2} = \frac{1}{nh^2} \sum_{j=1}^{n} \sum_{k=1}^{9} K_1 \left( \frac{\hat{G}_1(z_1) - R_{1jk}}{h_n} \right) K_2 \left( \frac{\hat{G}_2(z_2) - R_{2jk}}{h_n} \right) \hat{g}_1(z_1) \hat{g}_2(z_2),
\]

where \( \{R_{1jk}, R_{2jk}, j = 1, \ldots, n \} = \{(\pm R_{1j}, \pm R_{2j}), (\pm R_{1j}, 2 - R_{2j}), (2 - R_{1j}, \pm R_{2j}), (2 - R_{1j}, 2 - R_{2j}) \} \) with \( R_{ij} = G_i(z_{ij}) \) and \( \hat{g}_i(\cdot) \) is the estimated marginal density of \( z_i, i = 1, 2 \).

Because the observed demand is censored, the estimator has to be adapted accordingly. There are four subsets of support that are relevant for us. Let \( S_0, S_1, S_2 \) and \( S_b \) be the support of \( G(\cdot, \cdot) \) corresponding to the four demand patterns, respectively. Let \( r_0 = \Pr[z_1 \leq z_0^1, z_2 \leq z_0^2] \) then we can estimate the following

\[
\hat{g}^1(z_1, z_2) = \frac{g(z_1, z_2)}{G_2(z_0^2) - r_0} \quad \text{on} \quad S_1; \quad \hat{g}^2(z_1, z_2) = \frac{g(z_1, z_2)}{G_1(z_0^1) - r_0} \quad \text{on} \quad S_2
\]

\[
\hat{g}^b(z_1, z_2) = \frac{g(z_1, z_2)}{1 - r_0 - r_1 - r_2} \quad \text{on} \quad S_b,
\]

using the method outlined above.

### 2.5 Empirical Results

The estimates of the parameters are provided in Table (2.4). These estimates lead to the following utility function

\[
\hat{U}(q_1, q_2; \theta_1, \theta_2) = \theta_1 q_1 + \theta_2 q_2 - 5.15 q_1^2 - 0.65 q_2^2 - 0.01 q_1 q_2,
\]

suggesting that the two advertisements are neutral goods as \( \hat{c} \) is almost zero. This is consistent with the data. With complementarity we would have observed a strong dependence between the two consumption patterns, and with strong substitutability we would have not observed any advertisements with both. See also Figure (??). Our estimates imply that \( z_0^1 = 978.51 \) and \( z_0^2 = 298.83 \) are the threshold types below which the publishers sells the standard listing. Under imperfect information there is always a tradeoff between not excluding lower types at the cost of providing higher informational rents to the higher types.
and excluding the lower types at the cost of lower revenue. The magnitude of each cost depends on
the likelihood of different types. Optimal exclusion expresses this tradeoffs and determines the marginal
type that equates the two. For instance if it is more probable that any firm is of lower than higher type
(left-skewed), then the threshold type should be closer to the lower type. In our specification $\hat{\rho}_i$, measures
the skewness. Since it is greater than 1, the marginal distribution of $z_i$ is concentrated around the lower
value. This implies that $z_i^0$ must be closer to $\bar{z}_i$. As a matter of fact the difference between the two is 9.67
for Verizon and 31.05 for Ogden. The difference is relatively larger for Ogden but this is because $\rho_2 < \rho_1$.
Thus high skewness also explains why we observe only 925 and 259 firms choosing the Display option (the
largest possible size) from Verizon and Ogden, respectively.$^{31}$ Furthermore, because a large proportion
of firms advertise with only one publisher (more than the free listing), the data suggests that the joint
density of $(z_1, z_2)$ must be highly concentrated at the lower end of $(z_1, z_2)$. Our estimates corroborates
this reasoning. To given an idea about the skewness, the kernel estimation of the joint density of $(z_1, z_2)$
for the subpopulation who choose more than standard listing from both publishers is presented in Figure
(2.3). In this respect the model seems to fit the data well.

Because buying advertisements is a business-to-business activity, the demand depends on its usefulness
in creating more demand. Suppose there is a single firm in a market (say a doctor) then clearly he/she
does not need to buy advertisement, but this might change if there are more than one doctor in the
market. Therefore, we might want to ascertain if the willingness-to-pay for advertisement is affected by
the level of competition. Since we obtain the pairs $(\hat{\theta}_{1j}, \hat{\theta}_{2j})$, we can address this question by running a
simple OLS regression on some measure of level of competition. We use the number of firms in the same
subheading, the average quantity of advertising in that industry, the standard deviation of the quantity
of advertising, whether or not the firms have national brands (or trademark) and whether for those the
directories provided with guide option as our measure of competition. We recall that guide provides
additional advertising space by listing specialities. It covers Attorneys, Dentists, Physicians Insurance
companies, etc. The regression results are presented in Table (2.5), where the standard errors are reported

$^{31}$It is important to interpret the larger magnitude of $\rho_i$ relative to $(\hat{x}_i - \bar{x}_i)$ which is large.
in parenthesis and the (** and (*) denote estimates that are significant at 5% and 10% confident level, respectively. As we can see the number of firms in the same industry, the average size of ads consumed in the industry and being in certain industry all have positive effects on firm’s willingness to pay for advertising but only the average size is significant. Nonetheless, the square of the number of competitor is negative and significant, suggesting that effect of competition decreases with the level of competitors. An interesting implication of the regression is that whether or not a firm has national presence affects $\hat{\theta}_1$ negatively but $\hat{\theta}_2$ positively. Furthermore, it is also interesting to note that the marginal cost of printing a pixel for Verizon is substantially lower than that of Ogden. Even then, the price charged by Verizon is much higher across comparable categories, see Table 1. This could be because Verizon enjoys a higher brand effect which is not captured by the model. Our model does not explain either the demand pattern of firms with national brand or the brand effect of the publisher, because the demand side is captured by reduced form parameters. Explicitly modeling the demand side is important but beyond the scope of this research.

In Table (2.7) we provide the summary statistics of the recovered types. In the part that is labeled (A) we have the entire sample while (B) corresponds to the sub sample that corresponds to those who choose advertisements strictly greater than the standard listing for both the publishers. The estimated correlation coefficients between $\hat{\theta}_1$ and $\hat{\theta}_2$ for each of the two cases are 0.26 and 0.38, respectively. The fact that it is higher for case (B) is not surprising because if only those firms with higher willingness-to-pay for both the advertisements would buy strictly more than standard listing. Therefore, conditional on this sub sample, we do expect the two parameters to have higher correlation than when we consider the entire sample.

**Cost of Asymmetric Information**

We know that with asymmetric information, the quantity allocation is distorted from the social optimum. Because under incomplete information the sellers equate private marginal benefit instead of social marginal benefits to marginal costs, the quantity allocation is distorted below the optimum.
Suppose $T_2 = D_2(q_1^*, q_2; \theta)$ be the residual demand for Ogden when Verizon sells $q_1^*$, then the profit function for Ogden is $\int_{q_20}^{q_2} D_2(q_1^*, y) dy - K_2 - m_2q_2$. Thus the best response is to choose $q_2^*$ such that $D(q_1^*, q_2^*) = m_2$, which equates the marginal benefit of $q_2^*$ to the marginal social cost of producing $q_2^*$, which determines the allocation rule of Ogden. We can compute the optimal allocation for Verizon along the best response strategy of Ogden for every $q_1^*$. Given the quasi-linear utility, we find that Verizon gains $2,651,052,914$ while Ogden gains $48,330,062$ and the firms will lose $2,699,115,638$. The resulting net social welfare gain is in the order of $267,337$. One would expect that under full information, the seller will extract all consumer surplus, but because of $(q_{10}, q_{20})$ is provided for free, the consumer’s indirect utility under complete information will not be zero but be equal to its valuation for $(q_{10}, q_{20})$, which is increasing in type. See Table (2.7), which presents the quantity pair under incomplete information, under full information and the corresponding difference in utility. As predicted by the theory, since the quantity allocation is not distorted for the highest type, the difference in the quantity under the two informational regime decreases with the allocation under incomplete information.

**Effect of Merger**

One important question in empirical industrial organization is to assess the effect of competition on social welfare. To assess such an effect, we evaluate a counterfactual scenario where the two publishers merge to be a single publishing entity selling two directories. By determining the new nonlinear tariff we can evaluate the informational rent for each directory and its choices for the two advertisements. By comparing with the current data, we can ascertain how the gains from competition is distributed across the various types. Furthermore, in recent years the demand for Yellow Page industry has been eroded due to the advent of other forms of directories, such as web-based search engines (e.g., Google and yahoo). This erosion of traditional demand is not particular to Yellow Pages but also affects other forms of paper-based advertising, such as newspapers. One can expect that unprofitable publishing companies will exit the market or will merge. The optimal nonlinear pricing problem then becomes a multiproduct
nonlinear pricing with multidimensional screening. We use the results from Rochet and Chonie(1998) to characterize the optimal tariff, which is presented in Appendix A. (In Progress.)

**Effect of Exit**

With our estimates, we can also assess the welfare implication when one of the two publishers decides to exit from the market, say it is Ogden that exists. Clearly, those who have relatively higher preferences for advertising with Ogden will lose the most. Nonetheless, it could also adversely affect others too because Verizon being the only seller might substantially increase its prices for advertisements. Furthermore, comparing the profit under merger where Verizon sells two advertisements with the profit with only one advertisement choice, we can make some inference on the market structure and availability of variety of goods. When only one advertisement is available, because firms have two dimensional preferences, there are two possibilities to consider: (i) we impose \( q_2 = 0 \) and the utility is \( u(q_1; \theta_1) = \theta_1 - b_1/2q_1^2 \) and (ii) we treat \( q_1 = q_2 = q \) and the utility is \( u(q; \theta_1, \theta_2) = (\theta_1 + \theta_2)q - (b_1/2 + b_2/2 - c)q^2 \). The optimal pricing mechanism for case (i) is the usual monopoly nonlinear pricing with allocation rule and price function

\[
q_1(\theta_1) = \frac{1}{b_1} \left( \theta_1 - \frac{1 - F(\theta_1)}{f(\theta_1)} - m_1 \right),
\]

\[
T_1(q) = \theta q(\theta) - \frac{b_1}{2} q(\theta)^2 - \int_{\theta_0}^{\theta} (q(t) - q_{10}) dt,
\]

and \( \theta_0 \) solves \( \theta_0 + F(\theta_0)/f(\theta_0) = m_1 + 1 + b_1/2q_{10} \). Since the solution for case (ii) is a multidimensional screening problem, the method is very involved and is therefore provided in Appendix B.

### 2.6 Concluding Remarks

This paper contributes to the empirical literature that studies the effect of competition in a market with incomplete information. Using the insights from the theory of principal-agent with one principal to formulate a model of Stackelberg duopoly competition, we propose a tractable structural framework in which the pricing strategies in an Oligopoly market with incomplete information can be characterized and estimated. We address the question of effect of competition on the selling mechanism and subsequent effect on social welfare by fitting the model to a data on yellow page advertisements sold by two (utility and
non-utility) publishers in Central Pennsylvania. Our model and econometrics specification incorporates all the features of the data. Because the offered advertisement options have quality dimension, we construct a quality-adjusted advertisement size to incorporate the effect of quality.

The demand side is characterized by a utility function indexed by two dimensional vector of preference. This leads to a problem of multidimensional screening, which we solve by using two strategies: (i) we focus on quadratic pricing strategies which makes the competition more tractable and (ii) we use a method of aggregation by using one dimensional aggregating variable. This variable acts as “sufficient statistic” for both competition and two-dimensional private information. For the observed quadratic price schedules to be outcome of equilibrium, we show that it is necessary and sufficient that the inverse hazard rate of this “sufficient statistic” is linear, which implies that the distribution is Burr Type XII. The joint distribution of fundamental parameters pertaining to the demand and supply of the market and characterization of the distribution of the private taste of the firms (consumers) are obtained by estimating our structural model.

We use method of moments to estimate the parameters that characterize utility and costs and use a quantile transformation of the pseudo private information recovered from structural equations to estimate the joint distribution using a local linear kernel estimator that corrects for boundary bias. Our estimates suggest that the two advertisements are close to being neutral goods and that there is a significant heterogeneity amongst the firms for the two advertisements. Moreover, we find that the distribution for valuation for the utility publisher is more skewed towards lower valuation than for the non-utility publisher.

On the methodological ground, although our model seems to explain the data pretty well and can recover the joint distribution of the preference parameters, it does not allow for a general utility function and a general pricing behavior. Extending the model to some general class of utility function even within a quasi linear. This also points to the growing recognition in the empirical analysis of the markets for advancement in multidimensional screening models with competition. For instance, heterogeneity in risk and risk preferences is important in insurance markets. The structural analysis of such market

Over the recent years, advertising markets are going through a shift from traditional paper-based avenues to internet-based ones. If we want to study the effect of competition it is also important to understand the demand for advertisement. We used a parametric utility function in our model for tractability and a reduced form measure of the willingness-to-pay, which took the demand side as a black box. Our empirical results suggest that the willingness-to-pay depends on the competition as well as on the brand effect. Furthermore, in view of findings by Rysman (2004), advertisement can be viewed as a two-sided market. Hence, structural analysis of nonlinear pricing in two-sided market seems as an important generalization of our model, and is left for future research.
2.7 Appendix A: Proofs

This appendix collects the proofs of Lemmas, Propositions and Theorems in the paper.

Proof of Proposition 1:

In the first part of the proof we shall determine the optimal quantity allocation rule while in the second part we characterize the optimal exclusion, i.e. the lowest type served by the follower. The first step determines that the expected profit function is concave in $q_2$ and once the optimal quantity allocation is determined we must also confirm that the virtual profit function (the integrand in $\Pi_2$) is superovulate in $(q_2, z_2)$. The first condition guarantees existence and uniqueness of the allocation rule, while the second condition guarantees implementability; see Stole (2008). Let the integrand be $I$, then

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

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

Therefore the profit function is concave if $b_2 > \frac{2c^2}{b_1 + \beta_1}$ since from our assumption $g_2(z_2) > 0$. We can check that this condition is satisfied ex post. This allows us to use point wise maximization to determine the optimal $q_2$ and $z_2^0$. i.e. $\frac{\partial^2 I}{\partial q_2^2} = 0$ and hence

$$q_2(z_2) = \frac{z_2 - m_2 - \frac{c_2 q_{20} + c\alpha_1}{b_2 + \beta_1} - \frac{1 - G_2(z_2)}{g_2(z_2)}}{b_2 - \frac{2c^2}{b_1 + \beta_1}}.$$

Now, we have to check that $I$ is supermodular in $(q_2, \theta)$. Differentiating $\frac{\partial I}{\partial q_2}$ with respect to $z_2$ we get

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

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

$$= \left( 1 - H_2'(z_2) \right) - \left( b_2 - \frac{2c^2}{b_1 + \beta_1} \right) q_2'(\cdot) g(z_2) = 0,$$
where the second equality follows from substituting the first order condition for the optimal $q_2$ and the second equality follows by substituting the value of $q'_2(z_2)$. Hence, the supermodularity condition, 
\[ \frac{\partial^2 I}{\partial q_2 \partial z_2} \geq 0 \] is satisfied.

**Calculation of Optimal $z^0_2$**

The optimal $z^0_2$ is determined by the Euler’s method of differentiating the expected profit with respect to $z^0_2$:

\[ -\left( z^0_2 - H_2(z^0_2) - m_2 - \frac{c_2 - c^2 q_2(z^0_2)}{b_1 + \beta_1} \right) \left( q(z^0_2) - q_{20} \right) + \frac{b_2}{2} \left( q^2(z^0_2) - q_{20}^2 \right) = 0. \]

We determine the optimal $z^0_2$ by first showing that the first order necessary condition shown above is satisfied if and only if $q_2(z^0_2) = q_{20}$. This will not only confirm our definition of $z^0_2$ but will also simplify the above expression making it considerably simpler to find the value of $z^0_2$. The “IF” part is trivially true. Now for the “ONLY IF” part, we begin by rewriting the above equation, by factoring out $(q_2(z^0_2) - q_{20})$, as

\[ -\left( z^0_2 - H_2(z^0_2) - m_2 - \frac{c_2 - c^2 q_2(z^0_2)}{b_1 + \beta_1} \right) + \frac{b_2}{2} \left( q_2(z^0_2) + q_{20} \right) = 0. \]

From the optimal contract, at $z^0_2$, we get $q_2(z^0_2) \left( 2 - \frac{2c_2}{b_1 + \beta_1} \right) = z^0_2 - H_2(z^0_2) - m_2 - \frac{c^2 q_{20} - c_1}{b_1 + \beta_1}$, which with the first order condition gives

\[ \left( c^2 \left( \frac{b_2}{b_1 + \beta_1} - \frac{2}{2} \right) \right) q_2(z^0_2) = \left( c^2 \left( \frac{b_2}{b_1 + \beta_1} - \frac{2}{2} \right) \right) q_{20} \Rightarrow q_2(z^0_2) = q_{20}. \]

To determine $z^0_2$, we start with the optimal allocation rule $q_1(\cdot)$ evaluated at $z^0_2$, and use $z^0_2 - H_2(z^0_2) = h_2(z^0_2)$ to express $z^0_2$ as

\[ z^0_2 = h^{-1}_2 \left( \left( b_2 - \frac{c^2}{b_1 + \beta_1} \right) q_{20} + m_2 + \frac{c_1}{b_1 + \beta_1} \right) \]

\[ = \frac{\pi_2}{1 + \rho_2} + \frac{\rho_2}{1 + \rho_2} \left( \left( b_2 - \frac{c^2}{b_1 + \beta_1} \right) q_{20} + m_2 + \frac{c_1}{b_1 + \beta_1} \right), \]
where the last equality uses the fact that \( h^{-1}_2(\cdot) \) is linear and substitute for \( l_2 = \frac{\pi_2}{1+\rho_2} \) and \( \zeta_2 = \frac{\rho_2}{1+\rho_2} \). A point to note here is we could have derived optimal \( z_2^0 \) by solving for that particular \( z_2^0 \) which would have ensured that \( q_2(z_2^0) = q_{20} \), instead of the method chosen above. However, we want to point out that it is inconsequential which way we determine \( z_2^0 \) because solving for \( z_2^0 \) which gives \( q_2(z_2^0) = q_{20} \) is the same as above. This should not be surprising given that it is a necessary and sufficient condition for optimality of \( z_2^0 \).

**Ramsey Formula**

From (15)

\[
T'_2(q_2) = z_2 - \frac{c_1}{b_1 + \beta_1} - (b_2 - \frac{2c_2}{b_1 + \beta_1})q_2 + \frac{c_2}{b_1 + \beta_1}q_{20}. \tag{A.1}
\]

Substituting for the optimal quantity allocation \( q_2(z_2) \) and after some simplification gives us

\[
\frac{T'_2(q_2(z_2)) - m_2}{T'_2(q_2(z_2))} = \frac{1 - G_2(z_2)}{g_2(z_2)} \frac{1}{\frac{\partial s_2(q_2(z_2))}{\partial q_2}}
\]

\[
\]

**Proof of Lemma 3:**

We want to show that \( \frac{1 - G(z)}{g(z)} \) is linear iff \( G(z) = 1 - [1 - (\varsigma + \xi z)]^\rho \) with \( \rho > 0 \).

If part is obvious. Now, suppose that \( \frac{1 - G(z)}{g(z)} = A - Bz \), \( B > 0 \), then

\[
\frac{g(z)}{1 - G(z)} = \frac{1}{A - Bz} \Rightarrow \frac{g(z)}{1 - G(z)} = -\frac{1}{A - Bz} \Rightarrow \frac{d \ln(1 - G(z))}{dz} = \frac{1}{B} \frac{d \ln(A - Bz)}{dz}.
\]

Integrating both sides allows us to write

\[
\int \frac{d \ln(1 - G(w))}{dw} dw = \int \frac{d \ln(A - Bw)}{dw} dw \Rightarrow \ln(1 - G(z)) = \frac{1}{B} \ln \left( \frac{A - Bz}{A - Bz_1} \right) \Rightarrow G(z) = 1 - \left( \frac{A - Bz}{A - Bz} \right)^\frac{1}{\rho}
\]
It is easy to check that \( G(z) = 0 \) and \( 0 \leq G(\cdot) \leq 1 \). However, to complete the argument we have to make sure that \( G(\bar{z}) = 1 \), which implies that \( A = B\bar{z} \). Therefore,

\[
G(z) = 1 - \left( \frac{B\bar{z} - Bz}{B\bar{z} - Bz} \right)^\frac{1}{2} = 1 - \left( 1 - \frac{z - \bar{z}}{\bar{z} - \bar{z}} \right)^\frac{1}{2} = 1 - \left( \frac{z - \bar{z}}{\bar{z} - \bar{z}} \right)^\frac{1}{2} = 1 - (1 - \{\zeta + \xi\})^0,
\]

where \( \zeta = \frac{\bar{z}}{\bar{z}} \) and \( \xi = \frac{1}{\bar{z}} \).

\[ \Box \]

**Proof of Proposition 2**

With this linearity assumption, we can rewrite the pricing function as

\[
T_2(q_2) = \zeta_2(q_2 - q_{20}) + (b_2 - \frac{2c^2}{b_1 + \beta_1})l_2 \left( \frac{q_2^2 - q_{20}}{2} \right) + l_2 m_2 (q_2 - q_{20}) + l_2 \left( \frac{c^2 q_{20} + c\alpha_1}{b_1 + \beta_1} \right) (q_2 - q_{20}) - \frac{b_2}{2} (q_2 - q_{20})^2
\]

\[
= \left( \zeta_2 + l_2 m_2 + \frac{l_2 (c^2 q_{20} + c\alpha_1)}{b_1 + \beta_1} - \frac{(c\alpha_1 - c^2 q_2)}{b_1 + \beta_1} \right) (q_2 - q_{20}) + \frac{l_2}{2} (b_2 - \frac{2c^2}{b_1 + \beta_1}) (q_2 - q_{20})^2 - \frac{b_2}{2} (q_2 - q_{20})^2
\]

\[
= \left( \zeta_2 + l_2 m_2 + \frac{\beta_2}{b_1 + \beta_1} (l_2 - 1) + \frac{c\alpha_1 (l_2 - 1)}{b_1 + \beta_1} \right) q_2 - \left( \zeta_2 + l_2 m_2 + \frac{l_2 c^2 q_{20}}{b_1 + \beta_1} + \frac{c\alpha_1 (l_2 - 1)}{b_1 + \beta_1} \right) q_{20}
\]

\[
+ \frac{1}{2} \left[ \left( b_2 - \frac{2c^2}{b_1 + \beta_1} \right) l_2 - \left( b_2 - \frac{2c^2}{b_1 + \beta_1} \right) \right] q_2^2 - \frac{1}{2} \left( b_2 - \frac{2c^2}{b_1 + \beta_1} \right) q_{20}^2
\]

(A.2)

Equating the coefficients corresponding to \( q_1 \) and \( q_2^2 \) we get

\[
\gamma_2 = - (\zeta_2 + l_2 m_2 + \frac{l_2 c^2 q_{20}}{b_1 + \beta_1} + \frac{c\alpha_1 (l_2 - 1)}{b_1 + \beta_1}) q_{20} - \frac{1}{2} \left( b_2 - \frac{2c^2}{b_1 + \beta_1} \right) q_{20}^2,
\]

\[
\alpha_2 = (\zeta_2 + l_2 m_2 + \frac{l_2 c^2 q_{20}}{b_1 + \beta_1} + \frac{c\alpha_1 (l_2 - 1)}{b_1 + \beta_1}),
\]

\[
\beta_2 = \left( b_2 - \frac{2c^2}{b_1 + \beta_1} \right) (l_2 - 1)
\]

(A.3)

where the constant term \( \gamma_2 \) will be used to determine the optimal fixed price to participate in the contract, \( \gamma_2 \).

\[ \Box \]

**Proof of Proposition 3:**
Showing that the expected profit of P1 is concave in $q_1$ and the virtual profit is super modular in $(q_1, z_1)$ follow the same line of reasoning as was done for P2 and hence excluded. The point-wise maximization of $\Pi_1$ with respect to $q_2$ gives

$$q_1(z_1) = \frac{z_1 - H_1(z_1) - m_1 - \frac{c_2^2 q_{10}}{b_2 + \beta_2}}{b_1 - \frac{2 c^2}{b_2 + \beta_2}} \quad \forall z_1 \in (z_0, \bar{z}_1], \quad (A.4)$$

and $q_1(z_1) = q_{10}$ for all $z_1 \in [\bar{z}_1, \bar{z}_1]$. Note that this first order condition is necessary as well as sufficient for profit maximization.

\[ \square \]

**Proposition 4** The coefficients of P1’s optimal tariff function $\alpha_1, \beta_1$ are determined by (A.5) and (A.6), respectively.

**Proof of Proposition 4:**

Here, we do not show the derivation of $\alpha_1$ and $\beta_1$ due to the lack of space. Both are determined by first order conditions and are tedious and long, hence we decided not to present the derivation here and derive everything in the supplementary companion paper Aryal and Huang (2009), section 1.

Using the FOC we can solve for $\alpha_1$, and get

$$\alpha_1 = \frac{A\Psi_1 (1 + \rho_1)(m_1 - \beta_1 q_{10}) + (A - \Psi_1 \beta_1)(\bar{z}_1 - A q_{10} + M_1)}{\{\Psi_1 A(1 + \rho_1) + M_2(\Psi_1 \beta_1 - A)\}} \quad (A.5)$$

where

$$\Psi(t) = \left( \frac{2 c_2^2 (b_1 + \beta_1)(b_2 (b_1 + \beta_1) - 2c^2) (l_2 - 1)}{b_2 l_2 (l_2 + \beta_1) - 2c^2 (l_2 - 1)} \right) t - \left( \frac{(c_2^2 + c_2^2 q_{20}) (b_1 + \beta_1)}{b_2 l_2 (b_1 + \beta_1) - 2c^2 (l_2 - 1)} + \alpha_1 \right) = At - B;$$

$$B = \frac{c(\zeta_2 + l_2 m_2)(b_1 + \beta_1) + c^3 q_{20} (l_2 - 1) + c^2 q_{20} (b_1 + \beta_1)}{b_2 l_2 (b_1 + \beta_1) - 2c^2 (l_2 - 1)} + \alpha_1 \frac{b_2 l_2 (b_1 + \beta_1) - c^2 (l_2 - 1)}{b_2 l_2 (b_1 + \beta_1) - 2c^2 (l_2 - 1)}$$

$$= M_1 + M_2 \alpha_1.$$
and

\[ \Psi_{\alpha_1} = \frac{\partial \Psi(t)}{\partial \alpha_1} = \frac{c^2(l_2 - 1) - b_2l_2(b_1 + \beta_1)}{(b_2l_2(b_1 + \beta_1) - 2c^2(l_2 - 1))}. \]

To determine optimal \( \beta_1 \) we optimize with respect to \( \beta_1 \) and it can be shown that the optimal \( \beta_1 \) solves

\[ (\tau'_1 - D)(\alpha_1 - m_1) - q_{10}(J(\alpha_1 - m_1) - (\tau'_1 - D)\beta_1) \]
\[ + \left\{ \frac{\tau_1 + B}{A} - \frac{J(\alpha_1 - m_1) - (\tau'_1 - D)\beta_1}{A} \right\} \]
\[ \times \frac{\nu_2}{(1 + \rho_1)} - J\beta_1 q_{10} + \left( \frac{\Delta \tau'_1 \rho_1 \beta_1}{\tau_1 - \tau'_1} + 2J\beta_1 - 1 \right) \frac{\nu_2}{A^2(2 + \rho_1)} = 0, \tag{A.6} \]

where

\[ \Psi(t)_{\beta_1} = \frac{\partial \Psi(t)}{\partial \beta_1} = \frac{2c^2l_2 - b_2l_2(b_1 + \beta_1) - 2c^2l_2^2b_2(b_1 + \beta_1) - b_2^2l_2^2(b_1 + \beta_1)^2}{b_2l_2(b_1 + \beta_1) - 2c^2(l_2 - 1)} t \]
\[ + \frac{(c(\zeta_2 + l_2m_2) + c^2q_{20})(b_1 + \beta_1)b_2l_2 - b_2l_2c(\zeta_1 + c^2q_{20})(l_2 - 1) - c(\zeta_2 + l_2m_2) - c^2q_{20}}{b_2l_2(b_1 + \beta_1) - 2c^2(l_2 - 1)} \]
\[ = Jt + D. \]

\( \square \)

**Proposition 5** In equilibrium the fixed price of purchasing advertisements from each publisher is \( \gamma_1 \) and \( \gamma_2 \), which are respectively given by (??) and (??).

**Proof of Proposition 5**

We shall just provide the equations that determine \( \gamma_1 \) and \( \gamma_2 \), and leave the detail of the proof (and
\[
\gamma_1 = \frac{(\theta_1^* - \alpha_1)^2(b_2 + \beta_2)}{b_1 + \beta_1} + \frac{c(\theta_1^* - \alpha_1)(\theta_2 - \alpha_2)(3 - (b_1 + \beta_1))}{(b_1 + \beta_1)(b_2 + \beta_2) - c^2} + \frac{(\theta_2 - \alpha_2)^2(b_1 + \beta_1)}{(b_1 + \beta_1)(b_2 + \beta_2) - c^2} - \frac{(\theta_1^* - \alpha_1)^2}{2(b_1 + \beta_1)} \\
- \frac{c^2(\theta_1^* - \alpha_1)^2}{(b_1 + \beta_1)(b_2 + \beta_2) - c^2} \times \left(1 - \frac{1}{b_1 + \beta_1}\right) + \frac{1}{2} \left(\frac{\theta_2 - \alpha_2)(b_1 + \beta_1) + c(\theta_1^* - \alpha_1)}{(b_1 + \beta_1)(b_2 + \beta_2) - c^2}\right)^2 (1 - (b_2 + \beta_2)) \\
- \theta_1^* q_{10} + \frac{b_1}{2} q_{10}^2 + \frac{(\theta_2 - \alpha_2 + cq_{10})^2}{2(b_2 + \beta_2)} \\
(A.7)
\]

\[
\gamma_2 = \left\{\frac{3(\theta_1^* - \alpha_1)(b_2 + \beta_2)(b_1 + \beta_1) + c(b_1 + \beta_1)(2(\theta_2^* - \alpha_2) + (b_2 + b_2)q_{20} - c^2(\theta_1 - \alpha_1) - c^3 q_{20})}{2(b_1 + \beta_1)(b_2 + \beta_2) - 2c^2} \right\} \\
\left[\frac{(\theta_2^* - \alpha_2)}{b_1 + \beta_1}(b_2 + \beta_2) + \frac{c^2(\theta_1^* - \alpha_1)(b_2 + \beta_2) + c^2(\theta_2^* - \alpha_2)}{(b_1 + \beta_1)(b_2 + \beta_2)(b_2 + \beta_2) - c^2} - \frac{cq_{20}}{b_1 + \beta_1}\right] \\
+ (\theta_2^* - \alpha_2) \left[\frac{\theta_2^* - \alpha_2}{b_2 + \beta_2} + \frac{c(\theta_1^* - \alpha_1)(b_2 + \beta_2) + c^2(\theta_2^* - \alpha_2)}{(b_1 + \beta_1)(b_2 + \beta_2)(b_2 + \beta_2) - c^2}\right] - \theta_2^* q_{20} - \frac{b_2 + \beta_2}{2} \left[\frac{\theta_2^* - \alpha_2}{b_2 + \beta_2}\right] \\
+ \frac{c(\theta_1^* - \alpha_1)(b_2 + \beta_2) + c^2(\theta_2^* - \alpha_2)}{(b_2 + \beta_2)(b_1 + \beta_1)(b_2 + \beta_2) - c^2} \right\}^2 + \frac{b_2}{2} q_{20}^2 + \frac{c(\theta_1^* - \alpha_1)(b_2 + \beta_2) + c^2(\theta_2^* - \alpha_2)}{(b_2 + \beta_2)(b_1 + \beta_1)(b_2 + \beta_2) - c^2} \\
\left[\frac{(\theta_1 - \alpha_1)(b_2 + \beta_2) + c(\theta_2^* - \alpha_2)}{(b_1 + \beta_1)(b_2 + \beta_2) - c^2}\right] - \frac{c(\theta_1 - \alpha_1 + cq_{20})q_{20}}{b_1 + \beta_1} + (\zeta_2 + m_2)q_{20} \\
+ \frac{1}{2} \left(\frac{b_2(2l_2 - 1) - \frac{2c^2(l_2 - 1)}{b_1 + \beta_1}}{b_1 + \beta_1}\right) q_{20}^2 \\
(A.8)
\]
2.8 Appendix B: Merger

In order to determine the change in welfare when two publishers merge, it is necessary to characterize optimal nonlinear pricing when the duopolist merge to be one monopolist. This leads to optimal multiproduct nonlinear pricing, which we solve using Rochet and Choné (1998).

The monopoly sells \((q_1, q_2)\), and charges \(T(q_1, q_2)\), which can now depend on the pair chosen. The cost of publishing a pair \((q_1, q_2)\) is given by \(C(q_1, q_2) = K_1 + K_2 + m_1q_1 + m_2q_2\). The type-\(\theta\) firm’s (agent’s) utility if she consumes the bundle \(q = (q_1, q_2)\) is \(U(q; \theta)\) and is the same as before.\(^{32}\)

Faced with the price schedule \(T(q)\), a \(\theta\)− firm obtains a surplus of

\[
s(\theta) = \max_{q \geq q_0} U(q; \theta) - T(q).
\]

The tariff function need not be separable in \(q_1\) and \(q_2\) but must be such that \(T(q_0) = 0\), \(T(q_i, q_{j0}) = T_i(q_j)\), \(i, j \in \{1, 2\}, i \neq j\).\(^{33}\) Since the indirect utility function \(s(\cdot)\) is implementable by some tariff function \(T(q(\cdot))\) and the utility function is convex in \(\theta\), \(s(\cdot)\) is also convex and continuous and it also satisfies the envelope condition: \(\nabla s(\theta) = q\). This implies that the objective of the publisher can be viewed as choosing appropriate surplus function \(s(\theta)\) such that it maximizes the expected profit

\[
\begin{align*}
\text{EΠ}(q, T) &= \max_{q(\cdot), T(\cdot)} \int_{\Theta} \{T(q(\theta)) - C(q(\theta))\} f(\theta) d\theta \\
\text{EΠ}(s) &= \max_{s(\cdot)} \int_{\Theta} \{\theta \cdot \nabla s(\theta) - b/2 \cdot (\nabla s(\theta))^2 + cs_1(\theta)s_2(\theta) - C(\nabla s(\theta)) - s(\theta)\} f(\theta) d\theta,
\end{align*}
\]

such that \(s(\theta)\) is convex in \(\Theta\) and \(s(\theta) \geq s_0(\theta) = U(q_0; \theta)\). If \(\theta\) were only of single dimension, we could ignore (relax) the convexity restriction and maximize with respect to \(s(\cdot)\) and verify ex-post that the

\(^{32}\)Whenever we use \(\theta\) without qualification, we mean the vector \((\theta_1, \theta_2)\) and similarly \(q\) without qualification stands for a vector \((q_1, q_2)\).

\(^{33}\)When the two types are independent i.e. \(\theta_1 \perp \theta_2\) and \(c = 0\) then it is optimal for the monopoly to offer \(T(q_1, q_2) = T_1(q_1) + T_2(q_2)\), where each \(T_i(q_i)\) is a nonlinear tariff.
function is convex and satisfies the participation constraints. Two important properties of multiproduct nonlinear pricing has been established in the literature, which are: (i) The publisher (seller) will always find it profitable to exclude some positive mass of firms (denoted as $\Theta_0$) who are all bunched at outside option $q_0$, see Armstrong (1996) and (ii) The solution of the relaxed problem is generically not convex and hence there is “bunching of second type” where firms outside $\Theta_0$ are bunched to receive same quantity even though their taste parameters differ, see Rochet and Chone (1998).

Optimal nonlinear pricing is then characterized by exploiting the following simple idea: we begin with the functional $E\Pi(s) = \int\int_{\Theta} \phi(\theta_1, \theta_2, s, \nabla s(\theta))f(\theta)d\theta_1d\theta_2,$ where $\phi(\cdot)$ is the integrand of the expected profit, and if $s(\cdot)$ is the optimal surplus function then for any well behaved admissible function (defined below) $p(\theta)$, $E\Pi(s + p) - E\Pi(s)$ must be always non-positive. Then optimal $s$ must satisfy the following Euler’s equation

$$\alpha(\theta) = -\frac{\partial \phi f(\theta)}{\partial s} + \sum_{i=1}^{2} \frac{\partial}{\partial \theta_i} \left( \frac{\partial \phi}{\partial \nabla_i s(\theta)} \right) = 0.$$ 

Then to characterize the optimal solution, we follow Theorem 2’ in Rochet and Chone (1998) by finding $s(\cdot)$ such that $\Theta$ is partitioned into three regions (see Figure (2.2)):

1. The exclusion region $\Theta_0$, on which $s(\theta) = s_0(\theta)$;

2. The bunching region $\Theta_B$, where $s(\cdot)$ only depends on $\tau = \theta_1 + \theta_2$;

3. The non-bunching region $\Theta_1$, where $s(\cdot)$ is convex and firms are perfectly screened.

Thus, the problem is to first characterize the three sets, and determine the allocation and tariff for each set, and we follow Proposition 8 and 9 in Rochet and Chone (1998) closely to determine optimal $s(\cdot)$.

Since $\frac{\partial \phi}{\partial s} = 1$ we have $\alpha(\theta) = f(\theta) + \sum_{i=1}^{2} \frac{\partial}{\partial \theta_i} (\theta_i - b_i \nabla_i s(\theta) + c \nabla_j s(\theta) - m_i) f(\theta)$ and $\beta(\theta) = -\nu(\theta) \cdot n(\theta)$.

1) On $\Theta_0$, $q^*(\theta) = (q_{10}, q_{20})$ and let $\tau_0 = \theta_1 + \theta_2$ separate $\Theta_0$ from $\Theta_B$, then $\tau_0$ is determined by $\int_{\Theta_0} \alpha(\theta)d\theta + \int_{\partial \Theta_0} \beta(\theta) \cdot n(\theta)d\sigma(\theta) = 1$, where abusing notation we write $\alpha(\theta) = f(\theta) + div(\frac{\partial \phi}{\sigma \nabla s(\theta)} f(\theta))$.

This can be further simplified as:
\begin{align*}
1 &= \int_{\Theta} f(\theta) d\theta + \int_{\Theta} d\text{iv} \left( \frac{\partial \phi}{\partial \nabla s(\theta)} f(\theta) \right) + \int_{\partial \Theta \cap \partial \Theta} \beta(\theta) \cdot n(\theta) d\sigma(\theta) \\
&= \int_{\Theta} f(\theta) d\theta + \int_{\Theta} f(\theta) \frac{\partial \phi}{\partial \nabla s(\theta)} \cdot n(\theta) d\sigma(\theta) - \int_{\partial \Theta \cap \partial \Theta} f(\theta) \frac{\partial \phi}{\partial \nabla s(\theta)} \cdot n(\theta) d\sigma(\theta)
\end{align*}

\begin{align*}
1 &= \int_{\Theta} f(\theta) d\theta + \int_{\partial \Theta \setminus \partial \Theta \cap \partial \Theta} f(\theta) \frac{\partial \phi}{\partial \nabla s(\theta)} \cdot n(\theta) d\sigma(\theta) \\
&= \int_{\Theta} f(\theta) d\theta + \int_{\Theta} f(\theta) \frac{\partial \phi}{\partial \nabla s(\theta)} \cdot n(\theta) d\sigma(\theta) - \int_{\partial \Theta \setminus \partial \Theta \cap \partial \Theta} f(\theta) \frac{\partial \phi}{\partial \nabla s(\theta)} \cdot n(\theta) d\sigma(\theta)
\end{align*}

where second equality follows from the Divergence theorem and substituting for \( \beta \) and the last equality follows by substituting \( \tau_0 = \theta_1 + \theta_2 \) and \( n(\theta) = (1, 1) \). \footnote{Divergence theorem implies \( \int_X d\text{iv} f(x) dx = \int_{\partial X} f(x) \cdot n(x) d\sigma(x) \).}

Recall that \( q^*(\theta) = (q_{10}, q_{20}) \), so optimal \( \tau_0 \) solves:

\begin{equation}
\int_{\theta_1}^{\theta_2} \int_{\theta_1}^{\theta_2} f(\theta_1, \theta_2) d\theta_2 d\theta_1 + \int_{\theta_1}^{\theta_2} f(\theta_1, \tau_0 - \theta_1) [\tau_0 - (b_1 - c)q_{10} - (b_2 - c)q_{20} - m_1 - m_2] d\theta_1 = 1 \quad (B.1)
\end{equation}

2) On \( \Theta_B, q^*_1(\theta) = q^*_2(\theta) = q_B(\tau) \) with \( \tau = \theta_1 + \theta_2 \) and \( \Theta_B = \{ (\theta_1, \theta_2) : \tau_0 < \theta_1 + \theta_2 < \tau_1 \} \). Since \( \int_{\Theta_B(q)} \alpha(\theta) d\theta + \int_{\partial \Theta_B(q)} \beta(\theta) \cdot n(\theta) d\sigma(\theta) = 0 \) for all \( \Theta_B(q) \), we get
Therefore, once \( \tau_0 = q_0 = \tau_1 \) is determined, \( \tau_1 \) solves

\[
0 = \int_{\theta_B} \alpha(\theta) d\theta + \int_{\partial \theta_B} \beta(\theta) \cdot n(\theta) d\sigma(\theta)
\]

\[
= \int_{\theta_B} f(\theta) d\theta + \int_{\theta_B} \text{div} \left\{ \frac{\partial \phi}{\partial \overrightarrow{s}(\theta)} f(\theta) \right\} d\theta - \int_{\partial \theta_B \cap \partial \theta} f(\theta) \cdot \frac{\partial \phi}{\partial \overrightarrow{s}(\theta)} \cdot n(\theta) d\sigma(\theta)
\]

\[
= \int_{\theta_B} f(\theta) d\theta + \int_{\theta_B} f(\theta) \frac{\partial \phi}{\partial \overrightarrow{s}(\theta)} \cdot n(\theta) d\sigma(\theta) - \int_{\partial \theta_B \cap \partial \theta} f(\theta) \cdot \frac{\partial \phi}{\partial \overrightarrow{s}(\theta)} \cdot n(\theta) d\sigma(\theta)
\]

\[
= \int_{\theta_B} f(\theta) d\theta + \int_{\partial \theta_B \setminus \partial \theta_B \cap \partial \theta} f(\theta) \frac{\partial \phi}{\partial \overrightarrow{s}(\theta)} \cdot n(\theta) d\sigma(\theta) - \int_{\partial \theta_B \setminus \partial \theta_B \cap \partial \theta} f(\theta) \cdot \frac{\partial \phi}{\partial \overrightarrow{s}(\theta)} \cdot n(\theta) d\sigma(\theta)
\]

\[
1 = \int_{\theta_B} f(\theta) d\theta + \int_{\theta_B} f(\theta) d\theta + \int_{\partial \theta_B \setminus \partial \theta_B \cap \partial \theta} f(\theta) \frac{\partial \phi}{\partial \overrightarrow{s}(\theta)} \cdot n(\theta) d\sigma(\theta)
\]

\[
= \int_{\theta_B} f(\theta) d\theta + \int_{\theta_B} f(\theta) d\theta + \int_{\theta_B} f(\theta) \frac{\partial \phi}{\partial \overrightarrow{s}(\theta)} \cdot n(\theta) d\sigma(\theta)
\]

\[
= \int_{\theta_B} f(\theta) d\theta + \int_{\theta_B} f(\theta) d\theta + \int_{\theta_B} f(\theta) \frac{\partial \phi}{\partial \overrightarrow{s}(\theta)} \cdot n(\theta) d\sigma(\theta)
\]

\[
= \int_{\theta_B} f(\theta) d\theta + \int_{\theta_B} f(\theta) d\theta + \int_{\theta_B} f(\theta) \frac{\partial \phi}{\partial \overrightarrow{s}(\theta)} \cdot n(\theta) d\sigma(\theta)
\]

\[
\int_{\theta_B} f(\theta) d\theta + \int_{\theta_B} f(\theta) d\theta + \int_{\theta_B} f(\theta) \frac{\partial \phi}{\partial \overrightarrow{s}(\theta)} \cdot n(\theta) d\sigma(\theta)
\]

\[
= \int_{\theta_B} \alpha(\theta_1, \tau_1) d\theta_1 + \beta(\theta_1, \tau_1) d\theta_1 + \beta(\theta_1, \tau_1) d\theta_1 = 1.
\]

Therefore, once \( \tau_0 \) is determined, \( \tau_1 \) solves

\[
\int_{\theta_B} \alpha(\theta_1, \tau_1) d\theta_1 + \beta(\theta_1, \tau_1) d\theta_1 + \beta(\theta_1, \tau_1) d\theta_1 = 1.
\]

For each bunch, \( q_B(\tau) \) for \( \tau_0 < \tau < \tau_1 \) solves

\[
\int_{\theta_B} \alpha(\theta_1, \tau_1) d\theta_1 + \beta(\theta_1, \tau_1) d\theta_1 + \beta(\theta_1, \tau_1) d\theta_1 = 0.
\]
Recall that $\alpha(\theta) = f(\theta) + \sum_{i=1}^{2} \frac{\partial}{\partial \theta_i}(\theta_i - b_i q_i(\theta) + c q_j(\theta) - m_i) f(\theta)$ but because $q_1(\theta) = q_2(\theta) = q_B(\tau)$ we get

$$\alpha(\theta_1, \tau - \theta_1) = f(\theta_1, \tau - \theta_1) \left\{ 1 + \sum_{i=1}^{2} (1 - (b_i - c) q_B^i(\tau) - m_i) \right\} + (\theta_1 - (b_1 - c) q_B(\tau) - m_1) \frac{\partial f(\theta_1, \tau - \theta_1)}{\partial \theta_1}$$

$$(\tau - \theta_1 - (b_2 - c) q_B(\tau) - m_2) \frac{\partial f(\theta_1, \tau - \theta_1)}{\partial \theta_2}$$

$$\beta(\theta_1, \tau - \theta_1) = f(\theta_1, \tau - \theta_1) \{ \tau - (b_1 - c) q_B(\tau) - (b_2 - c) q_B(\tau) - m_1 - m_2 \}$$

$$\beta(\tau - \theta_2, \theta_2) = f(\tau - \theta_2, \theta_2) \{ \tau - (b_1 - c) q_B(\tau) - (b_2 - c) q_B(\tau) - m_1 - m_2 \}.$$  

Therefore,

$$\left\{ 1 + \sum_{i=1}^{2} (1 - (b_i - c) q_B^i(\tau) - m_i) \right\} \int_{\theta_1}^{\tau-\theta_1} f(\theta_1, \tau - \theta_1) d\theta_1 + \int_{\theta_1}^{\tau-\theta_1} (\theta_1 - (b_1 - c) q_B(\tau) - m_1) \frac{\partial f(\theta_1, \tau - \theta_1)}{\partial \theta_1} d\theta_1$$

$$+ \int_{\theta_1}^{\tau-\theta_1} (\tau - \theta_1 - (b_2 - c) q_B(\tau) - m_2) \frac{\partial f(\theta_1, \tau - \theta_1)}{\partial \theta_2} d\theta_1 + (f(\theta_1, \tau - \theta_1) + f(\tau - \theta_2, \theta_2))$$

$$\times \{ \tau - (b_1 - c) q_B(\tau) - (b_2 - c) q_B(\tau) - m_1 - m_2 \} = 0,$$

or equivalently

$$\int_{\theta_1}^{\tau-\theta_1} f(\theta_1, \tau - \theta_1) d\theta_1 (b_1 + b_2 - 2c) q_B^1(\tau) + \left( b_1 - c \right) \int_{\theta_1}^{\tau-\theta_1} f_1(\theta_1, \tau - \theta_1) d\theta_1 + (b_2 - c) \int_{\theta_1}^{\tau-\theta_1} f_2(\theta_1, \tau - \theta_1) d\theta_1$$

$$- \left\{ f(\theta_1, \tau - \theta_1) + f(\tau - \theta_2, \theta_2) \right\} (b_1 + b_2 - 2c) q_B(\tau) - \int_{\theta_1}^{\tau-\theta_1} \left\{ (\theta_1 - m_1) f_1(\theta_1, \tau - \theta_1) \right\}$$

$$+ (\tau - \theta_1 - m_2) f_2(\theta_1, \tau - \theta_1) \right\} d\theta_1 - \left\{ f(\theta_1, \tau - \theta_1) + f(\tau - \theta_2, \theta_2) \right\} (\tau - m_1 - m_2)$$

$$- (3 - m_1 - m_2) \int_{\theta_1}^{\tau-\theta_1} f(\theta_1, \tau - \theta_1) d\theta_1 = 0.$$  

This is a linear ordinary differential equation in $q_0(\tau)$ of the form $v_1(\tau) q_B^i(\tau) = -v_2(\tau) q_B(\tau) + v_3(\tau)$, and the solution is given by

$$q_B(\tau) = \exp \left( \int_{\tau_0}^{\tau} \frac{v_2(x)}{v_1(x)} dx \right) \left( k + \int_{\tau_0}^{\tau} \exp \left( \int_{\tau_0}^{t} \frac{v_2(t)}{v_1(t)} dt \right) \frac{v_3(x)}{v_1(x)} dx \right),$$
where \( k \) is the constant.

3) In \( \Theta_1 \) the surplus function \( s(\theta) \) is strictly convex and is determined by the following Euler’s equation:

\[
f(\theta) + \sum_{i=1}^{2} \frac{\partial}{\partial \theta_i} \left\{ (\theta_i - b_i \nabla_i s(\theta) + c \nabla_j s(\theta) - m_i) f(\theta) \right\} = 0,
\]

together with the boundary conditions:

\[
\beta(\theta) = 0
\]
on the upper boundary of \( \Theta_1 \) and which is smooth pasting condition.

### 2.9 Appendix C: Exit

In this section we characterize optimal nonlinear price when after merger, the monopoly restricts the product sold to only \( q \). In view of that, the utility function should be redefined as \( u(q, \theta) = \theta_1 q + \theta_2 q - \frac{b_1}{2} q^2 - \frac{b_2}{2} q^2 + c q^2 = (\theta_1 + \theta_2)q - \frac{b}{2} q^2 \), where \( b = b_1 + b_2 - 2c \). Then the net (indirect) utility from consuming \( q \) is

\[
W(q, \theta_1, \theta_2; T(\cdot)) = s(q(\theta); \theta) = \max_{q \geq q_0} \{(\theta_1 + \theta_2)q - \frac{b}{2} q^2 - T(q)\}. \tag{C.1}
\]

Let \( s(\hat{\theta}, \theta) \) be the utility of a type \( \theta \) when he pretends to be \( \hat{\theta} \) and is given by \( s(\hat{\theta}, \theta) = u(q(\hat{\theta}), \theta) - T(q(\hat{\theta})) \) and slightly abusing the notation we get

\[
T(\theta) = (\theta_1 + \theta_2)q - \frac{b}{2} q^2 - s(\theta).
\]

The monopolist’s dual problem is to choose a convex (surplus) function \( s(\cdot) \) for each \( \theta \) such that \( s(\theta) \geq s_0(\theta) \) and

\[
s(\cdot) = \arg \max \int \int \left\{ (\theta_1 + \theta_2)q - \frac{b}{2} q^2 - K - mq - s(\theta) \right\} f(\theta_1, \theta_2) d\theta_1 d\theta_2 - K_1.
\]
Ignoring the constraints, and assuming that the reservation utility is type independent and normalized to $s_0(\theta) = 0$, the Hamiltonian for the control problem is

$$H(p) = \left(\sum_{i=1}^{2} \theta_i p_i - \frac{b}{2} p_1^2 - mp_1 - s\right) f(\theta_1, \theta_2) + \sum_{i=1}^{2} \lambda_i(\theta) p_i + \mu(\theta)(p_1 - p_2) - K,$$

where $p_i = \frac{\partial s}{\partial \theta_i}$ and our technological constraint $a(p_1, p_2) = 0$ is given by $p_1 - p_2 = 0$. For simplicity, $K + \frac{b}{2} h_1^2 + mh_1$ can be thought of as cost of providing $p_1$ level of utils. Then the solution to the above problem is given by the following conditions

$$\text{div}\lambda + \frac{dH}{ds} \leq 0, \quad a.e.\ \Theta$$

$$<\lambda, \nu> \geq 0, \quad a.e. \ \partial\Theta$$

$$p_i \in \arg\max H(z).$$

Then we have the following conditions:

$$\frac{\partial\lambda_1}{\partial\theta_1} + \frac{\partial\lambda_2}{\partial\theta_2} - f(\theta) \leq 0 \quad \text{(C.2)}$$

$$<\lambda, \nu> \geq 0, \quad \text{(C.3)}$$

$$\frac{\partial H}{\partial p_1} = 0 \Rightarrow \lambda_1 = (bp_1 + m - \theta_1)f(\theta) - \mu \quad \text{(C.4)}$$

$$\frac{\partial H}{\partial p_2} = 0 \Rightarrow \lambda_2 = \mu - \theta_2 f(\theta) \quad \text{(C.5)}$$

Recalling that $p_1 = s_1(\theta)$ and differentiating (B.4) and (B.5) with respect to $\theta_1$ and $\theta_2$, respectively gives

$$\frac{d\lambda_1}{d\theta_1} = (bs_{11}(\theta) - 1)f(\theta) + (bs_1 + m - \theta_1)f_1(\theta) - \mu_1(\theta); \quad \frac{d\lambda_2}{d\theta_2} = \mu_2(\theta) - f(\theta) - \theta_2 f_2(\theta),$$

which for all firms with $s(\cdot) > 0$ in (B.2) gives

$$(bs_{11}(\theta) - 1)f(\theta) + (bs_1(\theta) + m - \theta_1)f_1(\theta) - \mu_1(\theta) + \mu_2(\theta) - f(\theta) - \theta_2 f_2(\theta) - f(\theta) = 0.$$
From (B.3), at the boundary the inequality binds with equality (no distortion on top), and since $\nu \gg 0$, we have

\[
(bs_1(\bar{\theta}_1, \theta_2) + m - \bar{\theta}_1)f(\bar{\theta}_1, \theta_2) - \mu(\bar{\theta}_1, \theta_2) = 0
\]

\[
\mu(\theta_1, \bar{\theta}_2) - \bar{\theta}_2f(\theta_1, \bar{\theta}_2) = 0.
\]

From the last equation we conjecture $\mu(\theta) = \theta_2f(\theta)$ and let $s_1(\theta) = y(\theta)$, then we can write the optimality condition as

\[
(by_1(\theta) - 1)f(\theta) + (by(\theta) + m - \theta_1)f_1(\theta) - \theta_2f_1(\theta) + f(\theta_1, \theta_2) - f(\theta) - \theta_2f_2(\theta) - f(\theta) = 0
\]

\[
(by_1(\theta) - 1)f(\theta) + (by(\theta) + m - \theta_1)f_1(\theta) - \theta_2f_1(\theta) - f(\theta) = 0.
\]

Therefore the optimal contract is determined by the following boundary valued PDE

\[
\frac{\partial}{\partial \theta_1}[(by(\theta) + m - \theta_1 - \theta_2)f(\theta)] - f(\theta) = 0 \quad (C.6)
\]

with the boundary condition $(by(\bar{\theta}_1, \theta_2) + m - \bar{\theta}_1 - \theta_2)f(\bar{\theta}_1, \theta_2) = 0$. Integrating (B.6) with respect to $\theta_1$ and simplifying the expression gives

\[
q(\theta) = h_1 = s_1(\theta) = y(\theta) = \frac{\theta_1 + \theta_2 - m - \int_{\theta_1}^{\bar{\theta}_1} f(t, \theta_2) dt}{b} \quad (C.7)
\]

Next step is to determine the tariff rule that will implement the above allocation. To that end we begin by noting that $\frac{\partial s(\theta, \theta_2)}{\partial \theta_1} = q(\theta)$, so integrating with respect to $\theta_1$ we get

\[
s(\theta) - s(\bar{\theta}_1, \theta_2) = \frac{1}{b} \int_{\theta_1}^{\bar{\theta}_1} \left[ \theta_1 + \theta_2 - m - \int_{\theta_1}^{\bar{\theta}_1} f(t, \theta_2) dt \right] d\theta_1.
\]
Table 2.1: Comparison of Two Partial Price Schedules

<table>
<thead>
<tr>
<th>Listing</th>
<th>VZ Picas</th>
<th>VZ Percentage</th>
<th>OG Picas</th>
<th>OG Percentage</th>
<th>Color Category 1</th>
<th>Color Category 2</th>
<th>Color Category 3</th>
<th>Color Category 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VZ Price</td>
<td>OG Price</td>
<td>VZ Price</td>
<td>OG Price</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.4%</td>
<td>9</td>
<td>0.5%</td>
<td></td>
<td>$0</td>
<td>$0</td>
<td>$151</td>
<td>$114</td>
</tr>
<tr>
<td>18</td>
<td>0.6%</td>
<td>12</td>
<td>0.66%</td>
<td></td>
<td>$151</td>
<td>$114</td>
<td>$492</td>
<td>$428</td>
</tr>
<tr>
<td>27</td>
<td>0.89%</td>
<td>15</td>
<td>0.83%</td>
<td></td>
<td>$280</td>
<td>$240</td>
<td>$942</td>
<td>$828</td>
</tr>
<tr>
<td>36</td>
<td>1.19%</td>
<td></td>
<td></td>
<td></td>
<td>$492</td>
<td></td>
<td>$845</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space Listing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>1.79%</td>
<td>46</td>
<td>2.49%</td>
<td></td>
<td>$504</td>
<td>$490</td>
<td>$1,789</td>
<td>$1,096</td>
</tr>
<tr>
<td>72</td>
<td>2.38%</td>
<td>92</td>
<td>4.98%</td>
<td></td>
<td>$784</td>
<td>$587</td>
<td>$2,873</td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>3.58%</td>
<td>138</td>
<td>7.46%</td>
<td></td>
<td>$1,134</td>
<td>$1,008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>144</td>
<td>4.77%</td>
<td>184</td>
<td>9.95%</td>
<td></td>
<td>$1,436</td>
<td>$1,154</td>
<td>$2,242</td>
<td>$1,231</td>
</tr>
<tr>
<td>216</td>
<td>7.15%</td>
<td>230</td>
<td>12.44%</td>
<td></td>
<td>$2,080</td>
<td>$1,276</td>
<td>$3,289</td>
<td>$1,363</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Display</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>174</td>
<td>5.76%</td>
<td>211</td>
<td>11.43%</td>
<td></td>
<td>$1,638</td>
<td>$1,118</td>
<td>$2,458</td>
<td>$1,308</td>
</tr>
<tr>
<td>208</td>
<td>6.96%</td>
<td></td>
<td></td>
<td></td>
<td>$1,915</td>
<td>$2,861</td>
<td>$3,049</td>
<td>$3,326</td>
</tr>
<tr>
<td>355</td>
<td>11.77%</td>
<td>438</td>
<td>23.74%</td>
<td></td>
<td>$3,074</td>
<td>$1,722</td>
<td>$4,612</td>
<td>$2,254</td>
</tr>
<tr>
<td>537</td>
<td>17.77%</td>
<td></td>
<td></td>
<td></td>
<td>$4,473</td>
<td>$6,703</td>
<td>$7,145</td>
<td>$7,812</td>
</tr>
<tr>
<td>735</td>
<td>24.34%</td>
<td></td>
<td></td>
<td></td>
<td>$5,872</td>
<td>$8,808</td>
<td>$9,388</td>
<td>$10,256</td>
</tr>
<tr>
<td>1,110</td>
<td>36.76%</td>
<td></td>
<td></td>
<td></td>
<td>$8,341</td>
<td>$12,512</td>
<td>$13,344</td>
<td>$14,579</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,485</td>
<td>49.18%</td>
<td>908</td>
<td>49.19%</td>
<td></td>
<td>$10,093</td>
<td>$3,372</td>
<td>$15,133</td>
<td>$4,420</td>
</tr>
<tr>
<td>3,020</td>
<td>100.00%</td>
<td>1,845</td>
<td>100.00%</td>
<td></td>
<td>$18,510</td>
<td>$6,324</td>
<td>$27,770</td>
<td>$8,290</td>
</tr>
<tr>
<td>6,039</td>
<td>200.00%</td>
<td></td>
<td></td>
<td></td>
<td>$34,272</td>
<td>$51,434</td>
<td>$54,835</td>
<td>$60,002</td>
</tr>
</tbody>
</table>
Table 2.2: Number of Purchases and Revenue by Sizes

<table>
<thead>
<tr>
<th>VZ</th>
<th># Purchases</th>
<th>Percent of Purchases</th>
<th>Revenue</th>
<th>Percent of Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Listing</td>
<td>2,152</td>
<td>31.54%</td>
<td>$0</td>
<td>0%</td>
</tr>
<tr>
<td>Listing</td>
<td>2,471</td>
<td>36.22%</td>
<td>$648,127</td>
<td>10.77%</td>
</tr>
<tr>
<td>Space listing</td>
<td>1,486</td>
<td>21.78%</td>
<td>$1,132,532</td>
<td>18.82%</td>
</tr>
<tr>
<td>Display</td>
<td>714</td>
<td>10.46%</td>
<td>$4,236,973</td>
<td>70.41%</td>
</tr>
<tr>
<td>Total</td>
<td>6,823</td>
<td>100.00%</td>
<td>$6,017,632</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OG</th>
<th># Purchases</th>
<th>Percent of Purchases</th>
<th>Revenue</th>
<th>Percent of Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Listing</td>
<td>5,913</td>
<td>86.66%</td>
<td>$0</td>
<td>0%</td>
</tr>
<tr>
<td>Listing</td>
<td>484</td>
<td>7.09%</td>
<td>$105,805</td>
<td>12.75%</td>
</tr>
<tr>
<td>Space listing</td>
<td>167</td>
<td>2.45%</td>
<td>$98,341</td>
<td>11.85%</td>
</tr>
<tr>
<td>Display</td>
<td>259</td>
<td>3.80%</td>
<td>$625,441</td>
<td>75.40%</td>
</tr>
<tr>
<td>Total</td>
<td>6,823</td>
<td>100.00%</td>
<td>$829,587</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Table 2.3: Summary Statistics by Consumption Pattern

<table>
<thead>
<tr>
<th># Purchases</th>
<th>VZ only</th>
<th>OG only</th>
<th>Both</th>
<th>Neither</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most popular headings</td>
<td>3,721</td>
<td>110</td>
<td>800</td>
<td>2,192</td>
</tr>
<tr>
<td>Attorneys (3.63%)</td>
<td>$1,050</td>
<td>$812</td>
<td>$2,480</td>
<td>$3,405</td>
</tr>
<tr>
<td>Physicians (3.14%)</td>
<td>$2,672</td>
<td>$1,531</td>
<td>$5,311</td>
<td>$5,905</td>
</tr>
<tr>
<td>Psychologists (2.04%)</td>
<td>90</td>
<td>120</td>
<td>204</td>
<td>340</td>
</tr>
<tr>
<td>Automobile Repair &amp; Service (2.73%)</td>
<td>237</td>
<td>282</td>
<td>476</td>
<td>609</td>
</tr>
<tr>
<td>Beauty Salons (2.37%)</td>
<td>VAZ OG Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average price</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std of price</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std of size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.4: Estimation of the Parameters

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimates</th>
<th>Estimates</th>
<th>Estimates</th>
<th>Estimates</th>
<th>Estimates</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_1$</td>
<td>10.3</td>
<td>1.3</td>
<td>-0.001</td>
<td>10.41</td>
<td>3.4</td>
<td>2363.5</td>
</tr>
</tbody>
</table>
Table 2.5: OLS for Effect of Competition

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

Table 2.6: Incomplete Information vs. Complete Information

<table>
<thead>
<tr>
<th>Qt: Incomplete Info.</th>
<th>Complete Info.</th>
<th># Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>(101, 210)</td>
<td>(104, 210)</td>
<td>230</td>
</tr>
<tr>
<td>(106, 231)</td>
<td>(108, 243)</td>
<td>53</td>
</tr>
<tr>
<td>(137, 231)</td>
<td>(139, 243)</td>
<td>27</td>
</tr>
<tr>
<td>(137, 248)</td>
<td>(139, 260)</td>
<td>31</td>
</tr>
<tr>
<td>(137, 288)</td>
<td>(139, 300)</td>
<td>28</td>
</tr>
<tr>
<td>(237, 432)</td>
<td>(240, 442)</td>
<td>9</td>
</tr>
<tr>
<td>(572, 517)</td>
<td>(575, 527)</td>
<td>4</td>
</tr>
<tr>
<td>(572, 843)</td>
<td>(575, 851)</td>
<td>1</td>
</tr>
<tr>
<td>(1709, 697)</td>
<td>(1711, 706)</td>
<td>6</td>
</tr>
<tr>
<td>(1709, 1154)</td>
<td>(1711, 1160)</td>
<td>3</td>
</tr>
<tr>
<td>(3171, 2153)</td>
<td>(3173, 2153)</td>
<td>1</td>
</tr>
<tr>
<td>(6330, 1621)</td>
<td>(6330, 1624)</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.7: Summary Statistics of $(\hat{\theta}_1, \hat{\theta}_2)$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>VZ or OG</td>
<td>$\hat{\theta}_1$</td>
<td>4361</td>
<td>2166.68</td>
<td>3221.67</td>
<td>978.5</td>
</tr>
<tr>
<td>VZ or OG</td>
<td>$\hat{\theta}_2$</td>
<td>4361</td>
<td>338.65</td>
<td>158.38</td>
<td>298.8</td>
</tr>
<tr>
<td>VZ and OG</td>
<td>$\hat{\theta}_1$</td>
<td>800</td>
<td>3314.679</td>
<td>5207.623</td>
<td>1073.343</td>
</tr>
<tr>
<td>VZ and OG</td>
<td>$\hat{\theta}_2$</td>
<td>800</td>
<td>504.56</td>
<td>298.99</td>
<td>326.7</td>
</tr>
</tbody>
</table>
Fig. 2.1: Demand Pattern

Fig. 2.2: Multiproduct nonlinear pricing.
Fig. 2.3: Joint Density of $(z_1, z_2)$
Bibliography


Vita
Yao Huang

Education

The Pennsylvania State University State College, Pennsylvania 2004–Present
Ph.D. in Economics, expected in August 2010
Thesis: "Structural Estimation of Nonlinear Pricing Models in the Yellow Pages Industry
Thesis Advisor: Professor Isabelle Perrigne and Professor Quang Vuong

Syracuse University Syracuse, NY 2001–2003
M.S. in Statistics and M.P.A.

Dongbei University of Finance and Economics Dalian, China 1997–2001
B.A. in Economics

Awards and Honors
The Bates White Graduate Fellowship, 2006-2007
Huron-Liaoning North American Study Fellowship 2000-2001

Research Experience

Research Assistant 2005-2008
Professor Perrigne and Professor Vuong under NSF grant:SES-0452154
Develop theoretical models and implement nonparametric methods to estimate consumer preferences with unobserved consumer heterogeneity

Summer Senior Consultant Bates White LLC Summer 2006
Utilized Random-Coefficient Logit and Nested Logit Models to study demand for computer processors.

Research Assistant Summer 2005
Professor Pinke
Comparison of estimation results of Probit model and Linear Probability model

Research Assistant 2001-2003
Professor Yinger
Research on Litigations on Education Finance

Teaching Experience

Instructor 3 summer 2005-2010
Introductory Econometrics (Econ 490)
Statistical Foundations of Econometrics (Econ 390)

Teaching Assistant 3 semesters 2004-2006
Introductory Econometrics (Econ 490)

Recitation Instructor 4 semesters 2001-2003
Managerial Econometrics