Semiparametric Identification of Multidimensional Screening Models

Heleno M. Pioner

The University of Chicago

Preliminary and Incomplete. Please do not quote without permission.

August 24, 2007

Abstract

This paper considers the identification of multidimensional screening models in line with the model proposed by Rochet and Chone (1998). It shows how one can use the information presented in the characteristics of the consumers, in the attributes of the goods being sold and in the theoretical predictions of the model to recover preference and cost parameters. Moreover, it shows that using the information presented in the bunching, the property of grouping people with different characteristics by offering the same good, one may infer the distribution of unobservable attributes, such as quality.

1 Introduction

The economics literature has long been interested in models where goods are characterized by the bundle of attributes they have. For example, computers may be defined by the velocity and capacity of its processor, the capacity of its hard drive and the amount of periphericals they have; automobiles may be characterized by the power of its engine, the maximum speed they may attain, among other attributes. In modelling goods as a bundle of attributes economists intend to solve two main problems: (1) what are the welfare consequences of the introduction of a new good, i.e. what is the impact in terms of demand and supply when a new good is introduced and (2), in markets with a large amount of differentiated products, the attributes approach is a practical way to reduce the number of parameters being estimated. These ideas have been presented in the literature of since the works of Tinbergen (1956), Lancaster(1966), Gorman (1980) and Rosen (1974), and more recently, in Berry, Levinsohn and Pakes (1995) and Petrin (2002) in the context of Industrial Organization and Ekeland, Heckman and Nesheim (2004) and Heckman, Matzkin and Nesheim (2006) in the Labor literature, among others.

This paper comes to add the literature on identification of these hedonic models. It analyses the empirical content of multidimensional screening models in the light of the theory posed by Rochet and Chone (1998). The term multidimensional screening stands for markets where there is a monopolist offering a variety of goods, each one being described by a vector of attributes. The monopolist tries to customize each product to each potential costumer in order to attract as much consumers as he can and to extract the maximum surplus from each buyer. The idea is to show how can one uses the information presented in the characteristics of the consumers, in the attributes of the goods being sold and in the theoretical predictions of the model to recover preference and cost parameters and also to show how one unobservable attribute, namely quality, can be infered based on these informations.

The importance of these two points is made clear by looking at how two different fields in Economics, namely Industrial Organization and Labor Economics have been dealing with the problem of identification in hedonic models. The former is mostly concerned about how prices are formed as a function of the attributes. IO researchers are preoccupied with the fact that, as they look at oligopoly markets, firms might be acting strategically in setting the prices of their goods, taking as given the attributes each one of these goods have. In this sense, it is costly to change attributes; therefore firms would choose to keep them fixed for a long time and make all its strategy by changing prices. However, by taking attributes as given, their models do not allow to investigate how new products, or in the terminology used here, new bundles may be released by firms as cost or preference parameters change. This is a major drawback if one wants to measure the welfare impact of a new good in a given market.

The Labor Economics literature has taken other direction. Since the seminal work from Rosen (1974) Labor economists are mostly concerned on how wages for a given job (the equivalent of the good in the IO example) may be explained by different characteristics of the worker or the firm or the attributes of the job. They have paid attention in modeling how workers are matched with each firm and model the attributes of the jobs as endogenously formed, as one can see in Ekeland, Heckman and Nesheim (2004), Ekeland (2005) and Heckman, Matzkin and Nesheim (2006). In this sense, they have a more suited model to analyse welfare effects from the introduction of new goods, given the market in study can be assumed as competitive. Unfortunatelly, the problem becomes extremely complex as one treats multidimensional attributes, involving a system of partial differential equations that might be not well behaved in the applications one is interested to study. In particular, those papers have made progress in the identification and estimation of hedonic models where the attribute is scalar and observable, which is remarkable given the non parametric approach they assume, but they have limited applicability on the markets usually studied because of the intrinsical multidimensionality of the goods.

The contributions of this paper are twofold. First, it is showing how to use the predictions of Rochet and Chone (1998), which allows the attributes being multidimensional, to identify preference and cost parameters of the consumers and the firm. This is a slight progress comparing to Ekeland, Heckman and Nesheim (2004) and Heckman, Matzkin and Nesheim (2006) for allowing attributes being multidimensional but it is a major step comparing to the IO literature. Here not only attributes are endogenously determined but also it flexibilizes the assumptions about the agents while, in the traditional IO literature, it is common to assume some parametric form for the heterogeneity of the agents. This paper makes no assumption regarding the form of the heterogeneity of consumers, although keeping the same quasi-linear parameterization for preferences.

The second and most important contribution is allowing for an unobservable attribute and showing how it relates with the characteristics of the agents. In terms of the econometrics, this is a problem of identifying a function $\xi = m(x, \varepsilon)$ given that ξ is unobserved but the analyst observes $z = h(x, \varepsilon)$ and knows that z and ξ are somehow related. This is a more important depart from the results in Heckman, Matzkin and Nesheim (2006) because now one can ask questions about how much of the observed price variation is due to quality changes.

The approach followed by this article in the identification of the hedonic model is somehow different from the idea presented in Heckman, Matzkin and Nesheim (2006). The idea is to use all the information presented in the bunching region to recover most of the parameters and give guidance on how to back up the function $\xi = m(x, \varepsilon)$. The bunching region is the subset of agents that, although they have different observable and unobservable characteristics, the monopoly chooses to offer them the same good, i.e. the same bundle of attributes. It occurs because there is a need to satisfy two constraints: the participation constraint (the monopolist wants as much people consuming as possible) and incentive compatibility (he also wants to extract as much surplus from each consumer as he can).

The advantage of using the bunching region is that, in this particular subset, all the attributes are uniquely tied. In other words, in this region, once the analyst controls for the observable attributes, there is no variation in prices because the unobservable attribute is fixed. This is a major result from the model that helps to identify cost parameters and the values of the function $\xi = m(x, \varepsilon)$ on the boundary of the screening region, i.e. the subset of agents where the monopolist offers a different good for each agent. Then, using average derivative methods already devised elsewhere (Powell, Stock and Stoker (1989), Ekeland, Heckman and Nesheim (2004), Buera (2006)) it is possible to infer the non-parametric function $\xi = m(x, \varepsilon)$ as well as the distribution of the unobservable heterogeneity $f(\varepsilon)$.

The major drawback of the paper is the market setting used, i.e. the monopoly environment. In most of the interesting applications of hedonic models more than one firm offers the good in the market. However, the idea of using the information presented in bunching region is novel and could be applied in the other environments (competitive and oligopoly) as those other models get better characterization results.

The paper continues as follows: in section 2 the model considered by Rochet and Chone (1998) is presented, as well as the small extension that will allows one to take the model to data. Section 3 presents the identification of the model. Section 4 concludes.

2 Model

This section is dedicated to explain the model considered by Rochet and Chone (1998) (hereafter RC) which is the baseline of the model that will be studied later. It follows the hedonics literature and call attributes the characteristics of the goods being sold. In a labor market setting, attributes would be the riskiness of the job, or the level of responsibility involved in the job, for example. The term characteristics will be used to denote the intrinsic characteristics of the agents or the firm. Thus, in the same labor market example, the age and the education level of the agent would be characteristics. This subsection does not make differences in terms of observable and unobservable characteristics or attributes in order to simplify the exposition of the model. Section 2.1 will clarify which components are observable and which are not and explain what changes in terms of the analysis of the model.

RC consider a multiproduct monopolist who chooses a vector of attributes and a price for each good it sells. He produces just one unit of each good. The monopolist takes into account incentive compatibility and individual rationality from agents, who have the following utility function:

$$U(x, z) = x \cdot z - p(z) \tag{1}$$

where x denotes a vector of characteristics of the buyers, z is a vector of attributes and p is its price. They assume that $x \in \Omega \subset \mathbb{R}^n$ and $z \in Z \subset \mathbb{R}^n$; therefore, the agent's vector of characteristics has the same dimensionality as the vector of attributes the monopolist offers. At this moment, the properties of the function $p(z)$ cannot be infered because they will be determined in equilibrium by the model.

Each agent chooses a good $z \in Z$ that maximizes its utility, i.e.

$$\begin{aligned} U(x) &= \max_{z \in Z} \{x \cdot z - p(z)\} \qquad (2) \\ s.t. U(x) &\geq \bar{U} \end{aligned}$$

where $\bar{U}$ is the reservation utility of the agent (RC assumes $\bar{U}$ is the same accross all agents).

One could ask whether this problem has a solution, given that, at this moment, there is no other information about the properties of $p(z)$. It happens that convex analysis guarantee the function $U(x)$ to be continuous and differentiable, which is all one needs to proceed with the analysis[1].

The monopolist is interested in maximizing profits, which can be written as the total surplus minus the utility agents get, i.e. $\Pi(x, z) = S(x, z) - U(x)$. Given that profits and utility are additively separable in the price, the surplus will be given by

$$S(x, z) = x \cdot z - C(z) \qquad (3)$$

[1] The reason why $U(x)$ is C^1 comes from a familiar argument: $U(x)$ is the envelope of the problem (2) and, provided the problem is concave and C^1 (which is clearly the case in (2)) the envelope of problem (2) will also be C^1 (see Rockafellar (1972)). This point is also made clear in a paper by Carlier and Lachand-Robert (2001).

Thus, the monopolist's profit can be written as

$$\Pi(x, z) = x \cdot z - C(z) - U(x) \tag{4}$$

Finally, RC consider a quadratic cost function, in the form

$$C(z) = \frac{1}{2} \sum \theta_i z_i^2 \tag{5}$$

Rochet (1987) shows that, under the above hypothesis about the utility function, $z = \nabla U(x)$ and, substituting this on the profit function one obtains the problem the monopolist wants to solve

$$\phi(U) = \max_{U} \int_{x \in \Omega} [x \cdot \nabla U(x) - C(\nabla U(x)) - U(x)] f(x) \, dx \tag{6}$$

$$s.t.\ U(x)\ \ convex\ and$$

$$U(x) \geq \bar{U}$$

The above optimization problem with just the participation constraint is usually called in physics "an obstacle problem". The new element here is the convexity restriction. This emerges naturally from the incentive compatibility constraint on the agent's problem and it places strong restrictions on the shape of the solution. RC solve the problem (6) in two steps. In the first one they solve the relaxed problem, ignoring the concavity constraint:

$$\phi\left(U^{*}\right)=\max_{U}\int\limits_{x\in\Omega}\left[x\cdot\nabla U\left(x\right)-C\left(\nabla U\left(x\right)\right)-U\left(x\right)\right]f\left(x\right)dx \tag{7}$$

$$s.t.\ U^{*}\left(x\right)\geq\bar{U}$$

This is the obstacle problem in physics, and the authors show the solution is given by the following first-order conditions:

$$\begin{aligned}\alpha\left(x\right)&=div\left(\frac{\partial}{\partial z}S\left(x,z\left(x\right)\right)f\left(x\right)\right)+f\left(x\right)\geq0 \\ \beta\left(x\right)&=-\left[\frac{\partial}{\partial z}S\left(x,z\left(x\right)\right)f\left(x\right)\right]\cdot n\left(x\right)\geq0\end{aligned} \tag{8}$$

where $n\left(x\right)$ is the (outward) normal vector to the boundary of the set of characteristics X. The expression in the first line denotes the divergence operator applied on the vector $\frac{\partial}{\partial z}S\left(x,z\left(x\right)\right)f\left(x\right)$[2]. As an illustration, consider the case where $\Omega\subset\mathbb{R}^{2}$ and $Z\subset\mathbb{R}^{2}$. The gradient $\frac{\partial}{\partial z}S\left(x,z\left(x\right)\right)$ is simply:

$$\frac{\partial}{\partial z}S\left(x,z\left(x\right)\right)=x-\theta\cdot z\left(x\right)=x-\theta\cdot\nabla U\left(x\right) \tag{9}$$

[2] The divergence operator is simply the trace of the jacobian matrix.

because of the quadratic cost function. Thus, the divergence will be

$$div\left(\left[x-\theta\cdot\nabla U\left(x\right)\right]f\left(x\right)\right)=\left[2-\theta_1 U_{11}\left(x\right)-\theta_2 U_{22}\left(x\right)\right]f\left(x\right)+\left[x-\theta\cdot\nabla U\left(x\right)\right]\cdot\nabla f\left(x\right) \tag{10}$$

Substituting (9) and (10) in the two expressions that characterize the first-order conditions (8) one obtains the partial differential equations that characterize the solution of the relaxed problem:

$$\begin{aligned} \left[3-\theta_1 U_{11}\left(x\right)-\theta_2 U_{22}\left(x\right)\right]f\left(x\right)+\left[x-\theta\cdot\nabla U\left(x\right)\right]\cdot\nabla f\left(x\right) &= 0 \quad (11) \\ -\left[x-\theta\cdot\nabla U\left(x\right)\right]f\left(x\right)\cdot n\left(x\right) &\geq 0 \end{aligned}$$

Although these expressions may seen complicated, the intuition behind them is simple. There is a direct effect of marginally increasing the vector of attributes, keeping everything else constant: it is the density of agents who started to consume now but used to not consume before ($f\left(x\right)$). This effect is positive, as it represents more revenue for the monopolist. However, there is also an indirect effect of marginally increasing the vector of attributes: people who used to consume before can now start consuming a different good, as well as the total cost of production for this new set of consumers may increase or decrease. This is represented by the divergence of the gradient of surplus with respect to attributes. The divergence gives this net effect of increasing or decreasing marginal surplus in a given point. Thus, the optimal choice of attributes is the one which the marginal change in surplus from a marginal change in attributes cannot be less than the increase in participation from this change in attributes.

The second expression gives the boundaries of the participation region. It says that the agents located at the boundary of the set of characteristics who will be attended by the monopolist will be the ones whose gradient of the surplus with respect to the attributes is in the direction towards the interior of the set of characteristics. Those are the agents that the incentive compatibility constraint is binding for the chosen vector of attributes.

The solution for the relaxed problem gives a lot of information about the solution of the original problem. First of all, the two first-order conditions combined represent a Lebesgue measure over the characteristics space, composed by a component over the interior of the set $(\alpha(x))$ and a component over the boundary $(\beta(x))$. The next theorem makes it clear.

Proposition 1 *The first-order conditions of the problem represent a Lebesgue measure over the set of characteristics.*

Proof. Take any two subsets A, B such as $A \subset \Omega$, $B \subset \Omega$, $A \cup B \subset \Omega$ and

$A \cap B = \phi$. Call $\mu(A)$ the integral

$$\mu(A) = \int_A \alpha(x)\, dx + \int_{\partial A} \beta(x)\, d\sigma(x)$$

This will be the definition of the set function $\mu(\cdot)$. Thus, according to this definition, the value of the set function $\mu(\cdot)$ applied on the set A will be:

$$\begin{aligned}
\mu(A) &= \int_A \left[div\left(\frac{\partial}{\partial z} S(x,z) f(x) \right) + f(x) \right] dx + \int_{\partial A} - \left[\frac{\partial}{\partial z} S(x,z) f(x) \right] \cdot n(x)\, d\sigma(x) \\
&= \int_A f(x)\, dx + \int_A div\left(\frac{\partial}{\partial z} S(x,z) f(x) \right) dx - \int_{\partial A} \left[\frac{\partial}{\partial z} S(x,z) f(x) \right] \cdot n(x)\, d\sigma(x) \\
&= \int_A f(x)\, dx + \int_{\partial A} \left[\frac{\partial}{\partial z} S(x,z) f(x) \right] \cdot n(x)\, d\sigma(x) - \int_{\partial A} \left[\frac{\partial}{\partial z} S(x,z) f(x) \right] \cdot n(x)\, d\sigma(x) \\
&= \int_A f(x)\, dx = F(A)
\end{aligned}$$

where the third line comes from the divergence theorem. The same procedure can be done to evaluate $\mu(B)$, which is $\mu(B) = F(B)$. The point that remains to be done is the value of $\mu(\cdot)$ at the union $A \cup B$.

$$\begin{aligned}
\mu(A \cup B) &= \int\limits_{A \cup B} \left[div\left(\frac{\partial}{\partial z} S(x,z) f(x) \right) + f(x) \right] dx + \int\limits_{\partial A \cup B} - \left[\frac{\partial}{\partial z} S(x,z) f(x) \right] \cdot n(x)\, d\sigma(x) \\
&= \int\limits_{A} \left[div\left(\frac{\partial}{\partial z} S(x,z) f(x) \right) + f(x) \right] dx + \int\limits_{\partial A} - \left[\frac{\partial}{\partial z} S(x,z) f(x) \right] \cdot n(x)\, d\sigma(x) + \\
&\quad + \int\limits_{B} \left[div\left(\frac{\partial}{\partial z} S(x,z) f(x) \right) + f(x) \right] dx + \int\limits_{\partial B} - \left[\frac{\partial}{\partial z} S(x,z) f(x) \right] \cdot n(x)\, d\sigma(x) \\
&= F(A) + F(B)
\end{aligned}$$

where the second line comes from the fact that A and B are disjoint subsets of $\mathbb{R}^n$ and, from the linearity of integration, they can be separated as two integrals in each of the areas. The line integral over the boundary respects the same idea: as the two sets are disjoints, the contours have no intersection points, thus they can be writen as two separable integrals.

By induction, one can see that $\mu\left(\overset{\infty}{\underset{i=1}{\cup}} A_i\right) = \sum\limits_{i=1}^{\infty} \mu(A_i)$ and, because $\mu(A) = F(A)$ numerically, then $\mu(\phi) = 0$ and $\mu(\Omega) = 1$. Therefore, $\mu(\cdot)$ is a Lebesgue measure over Ω. ■

Given that $\mu(x)$ induces a measure on the set Ω, it is possible to separate the set of characteristics in two disjoint subsets: Ω_0 were all the agents in this set do not participate and Ω_1, were all the agents in this set participate. The important thing to know is that, in the participation set, the first-order conditions are binding; thus, for all $x \in \Omega_1, \alpha(x) = \beta(x) = 0$. This is a major result because, as Ω_0 and Ω_1 are disjoint sets and $\mu(x)$ is a Lebesgue measure, then

$$\Omega_0 \cup \Omega_1 = \Omega \Rightarrow \mu(\Omega) = \mu(\Omega_0) + \mu(\Omega_1) = 1 \tag{12}$$

Using the fact that the integrands in the participation region are equal to zero, one obtains

$$\mu(X_1) = \int\limits_{\Omega_1} \left[div \left(\frac{\partial}{\partial z} S(x,z) f(x) \right) + f(x) \right] dx + \int\limits_{\partial\Omega_1} -\left[\frac{\partial}{\partial z} S(x,z) f(x) \right] \cdot n(x)\, d\sigma(x) = 0 \tag{13}$$

Therefore (12) and (13) implies $\mu(X_0) = 1$, i.e.

$$\int\limits_{\Omega_0} \left[div \left(\frac{\partial}{\partial z} S(x,z) f(x) \right) + f(x) \right] dx + \int\limits_{\partial\Omega_0} -\left[\frac{\partial}{\partial z} S(x,z) f(x) \right] \cdot n(x)\, d\sigma(x) = 1 \tag{14}$$

The solution to (14) gives the upper bound on the non-participation region X_0 and, for this relaxed problem, it may assume any shape. However, the shape of the boundary between Ω_0 and Ω_1 is crucial in terms of the properties of the function $U(x)$, the argument of the problem (6). The complete problem includes a restriction that $U(x)$ must be concave. In the parametric structure used by RC in (1) this means that the boundary must be a straight line, otherwise the incentive compatibility constraint will be violated. Proposition 2 states this result.

Proposition 2 *(page 802 in RC) The boundary between the participation and non-participation must be a straight line.*

Proof. Take two points, x and x' in the boundary. Clearly, $z(x) = z(x') = 0$ and the corresponding prices is also zero. Now consider the conxex combination $x'' = tx + (1-t)x'$, $t \in [0,1]$. Call $z'' = z(x'')$. Incentive compatibility says that

$$\begin{aligned} x'' \cdot z'' - p\left(z''\right) &\geq x'' \cdot z - p(z) = 0 \\ x'' \cdot z'' - p\left(z''\right) &\geq x'' \cdot z' - p\left(z'\right) = 0 \end{aligned}$$

The same applies for the agents x and x':

$$\begin{aligned} x \cdot z - p(z) &\geq x \cdot z'' - p\left(z''\right) \\ 0 &\geq x \cdot z'' - p\left(z''\right) \end{aligned}$$

$$\begin{aligned} x' \cdot z' - p\left(z'\right) &\geq x' \cdot z'' - p\left(z''\right) \\ 0 &\geq x' \cdot z'' - p\left(z''\right) \end{aligned}$$

Taking the convex combination of these two equations we get

$$
\begin{aligned}
0 &\geq \left[tx + (1-t)x'\right] \cdot z'' - p\left(z''\right) \\
&= x'' \cdot z'' - p\left(z''\right)
\end{aligned}
$$

Therefore $x'' \cdot z'' - p\left(z''\right) = 0$, which means that x'' also belongs to the boundary. Thus, the boundary is a straight line. ∎

The paper follows by showing how the results obtained for the relaxed problem can be extended for the original problem (6). RC show that the idea becomes to find a function $U^*\left(x\right)$ that follows the first-order conditions estabilished in (8) who also respects the concavity restriction. The only way to admit this combination is by allowing bunching on the types. This means that agents with different characteristics will consume the same good (same attributes) in equilibrium. The monopolist will not screen these different agents because there is no way to make them participate and, at the same time, to separate them with different goods. Therefore, as RC explain, bunching occurs as a form to conciliate the individual rationality constraint and the incentive compatibility restriction.

Mathematically, bunching occurs as a form to conciliate the straight line shape of the boundary between participation and non-participation regions. As

RC explain, the first-order conditions of the relaxed problem continue to hold; also, the measure of the non-participation region is still unitary, because the first-order conditions for the participation region are binding and equal to zero. However, the free shape of the boundary does not hold anymore; it was seen in proposition 2 (and page 802 in RC) the boundary between Ω_0 and Ω_1 must be a straight line. This means that a subset of Ω which was previously on Ω_0 is now participating and other subset of Ω_1 may now be not participating. The idea is take these agents who were close to the previous boundary between the two subsets Ω_0, Ω_1 and reallocate them in the bunching region. This will change the properties of the measure $\mu(\cdot)$ in the bunching region.

RC borrow from the potential theory literature the concept of sweeping operator. The sweeping operator works by reallocating the density along the area but preserving the distribution. It is a generalization of the mean preserving spread used by Rothschild and Stiglitz (1970). In terms of the solution of the original problem (7), RC show that the first-order conditions, at the bunch must satisfy two other conditions:

$$\begin{aligned} \int_{\Omega(z)} d\mu(x) &= 0 \qquad (15) \\ \int_{\Omega(z)} x d\mu(x) &= 0 \end{aligned}$$

where $\Omega(z) = \{x \in \Omega \,|x \cdot z - p(z) \geq x \cdot z' - p(z'),$ for every $z' \in Z \cup \{0\}\}$, i.e. $\Omega(z)$ is the set of all agents who are bunched consuming the good z. These conditions mean that possibly negative values that the $\alpha(x)$ function may assume in the bunch must be compensated by positive ones in the boundary function $\beta(x)$. This clearly respects the condition that the measure in the participation region must be zero and allows the boundary between the regions being a straight line.

Therefore, the solution for the monopolist problem as given by RC is given by three different regions, which are:

1. The non-participation region Ω_0, where agents consume $z = 0$ and its frontier is characterized by the following conditions:

$$\int_{x\in\Omega_0} \left[div\left(\frac{\partial}{\partial z}S(x,z)\,f(x)\right) + f(x)\right]dx - \int_{\partial\Omega_-} \frac{\partial}{\partial z}S(x,z)\cdot n(x)\,d\sigma(x) = 0 \quad (16)$$

$$and \; -\frac{\partial}{\partial z}S(x,z)\cdot n(x,\varepsilon) > 0$$

2. The perfect sreening region, Ω_1 where each consumer type x consumes a customized good $z(x)$ and the first-order conditions characterizing the

function $z(x)$ are:

$$\begin{aligned} div\left(\frac{\partial}{\partial z}S(x,z)\,f(x)\right)+f(x) &= 0 \\ and\ -\frac{\partial}{\partial z}S(x,z)\,f(x)\cdot n(x) &= 0 \end{aligned} \tag{17}$$

3. The bunching region, where each bunch is characterized by a straight line parallel to the boundary of the region Ω_0 and the sweeping conditions hold

$$\begin{aligned} \int\limits_{\Omega(z)}\left[div\left(\frac{\partial}{\partial z}S(x,z)\,f(x)\right)+f(x)\right]dx-\int\limits_{\partial\Omega(z)}\frac{\partial}{\partial z}S(x,z)\cdot n(x)\,d\sigma(x) &= 0 \\ \int\limits_{\Omega(z)}x\left[div\left(\frac{\partial}{\partial z}S(x,z)\,f(x)\right)+f(x)\right]dx-\int\limits_{\partial\Omega(z)}x\frac{\partial}{\partial z}S(x,z)\cdot n(x)\,d\sigma(x) &= 0 \end{aligned} \tag{18}$$

Notice that, up to now, there was no mention on the price function $p(z)$. This comes from the fact that as $z(x)$ asssumes different shapes in the regions Ω_1 and Ω_B, so does $p(z)$. However, once $z(x)$ becomes determined then one can go back to the agent's problem and solve

$$\begin{aligned} U(x) &= \max_{z\in Z}\{x\cdot z-p(z)\} \\ s.t.U(x) &\geq \bar{U} \end{aligned} \tag{19}$$

The solution for this problem gives a partial differential equation for $p(z)$ and this equation will be extremely helpful in combining the information presented in prices to the parameters of interest in order to identify the model. This will be explored in detail later.

This section has been treating the original model proposed by RC, its assumptions and results. It is clear that some simplifications in the original model are strong or not very useful if one considers to take this model to data. The next subsection will present a slightly different model that tries to fill these drawbacks and show that the main results of RC continue to be valid in a more flexible model.

2.1 Empirical Model

The model in the previous section lacks some important features necessary for its application in empirical problems. For example, which characteristic interact with each attribute? What are the observables and unobservable variables of the problem? In order to accomodate these points, this section will pose the model that will be taken to the data and show that the minor modifications over the RC model will not change the main results of the previous section.

The agent's characteristics will now be represented by the vector (x, ε), where $x \in X \subset \mathbb{R}^n$ is a vector of observable characteristics and $\varepsilon \in \mathcal{E} \subset \mathbb{R}$

is a (scalar) unobservable characteristic of the agent. In terms of the set of characteristics, it means that $\Omega = X \times \mathcal{E} \subset \mathbb{R}^{n+1}$, convex. The attributes vector has turned into (z, ξ), where $z \in Z \subset \mathbb{R}^n$ is a vector of observable attributes and $\xi \subset \mathcal{J} \subset \mathbb{R}$ is a (scalar) unobservable attribute. Thus, the set of attributes is given by $\Psi = Z \times \mathcal{J} \subset \mathbb{R}^{n+1}$. This representation is useful because it allows the presence of attributes such as quality, which is present in most application and may cause serious biases in the estimation if not treated properly.

The utility function to be considered has the following parameterization

$$\begin{aligned} U(x, \varepsilon) &= \max_{(z,\xi)\in\Psi} \{u(x, \varepsilon, z, \xi) - p(z, \xi)\} \\ &\quad \max_{(z,\xi)\in\Psi} \left\{x' A z + x' B \xi + \varepsilon C' z + D \varepsilon \xi - p(z, \xi)\right\} \end{aligned} \tag{20}$$

where A is a $n \times n$ matrix, B and C are $n \times 1$ vectors and D is a scalar. In this representation every characteristic may interact with each attribute. This flexibility is essential because, in the end, it allows the data to say what interactions between characteristics and attributes are significant and which ones are not. Finally, the cost function becomes

$$C(z, \xi) = z' \Gamma z + 2 z' \Lambda \xi + \Phi \xi^2 \tag{21}$$

where Γ is a $n \times n$ matrix, Λ is a $n \times 1$ vector and Φ is a scalar. Again, by allowing interactions between attributes in the cost function the goal is to capture any sort of complementarity or substitutability in the technology. Given that this is a monopoly setting, those effects should be presented if one wants to test for the presence of natural monopoly, for example.

The first point to be made precise is that the results from Rochet (1987) are still valid on this representation, that is, the attributes vector can be written as the gradient of the utility function. Proposition 3 states this result.

Proposition 3 *Let $\Omega = X \times \mathcal{E} \subset \mathbb{R}^{n+1}$, convex and $u(x, \varepsilon, z, \xi)$ as defined in (20). Then the attributes functions $z(x, \varepsilon)$ and $\xi(x, \varepsilon)$ can be written as a linear function of the gradient of $U(x, \varepsilon)$.*

Proof. This is an immediate consequence of proposition 2 in Rochet (1987). Given that Ω is a convex subset of $\mathbb{R}^{n+1}$ and $u(x, \varepsilon, z, \xi)$ is (1) - a linear function in the characteristics vector (x, ε) and (2) - a C^1 function in the attributes vector (z, ξ) then Rochet (1987) shows that

$$\forall (x, \varepsilon) \in \Omega, \ \nabla_{(x,\varepsilon)} u(x, \varepsilon, z, \xi) = \nabla_{(x,\varepsilon)} U(x, \varepsilon)$$

Substituting the parametric structure of $u(x, \varepsilon, z, \xi)$ in the result above. In order to simplify the notation call $\Theta = \begin{bmatrix} A & B \\ C' & D \end{bmatrix}$ and $\nabla_{(x,\varepsilon)} U(x,\varepsilon) = \nabla U(x,\varepsilon)$. Then

$$\begin{aligned} \Theta \cdot \begin{bmatrix} z \\ \xi \end{bmatrix} &= \nabla U(x,\varepsilon) \qquad (22) \\ \begin{bmatrix} z \\ \xi \end{bmatrix} &= \Theta^{-1} \cdot \nabla U(x,\varepsilon) \end{aligned}$$

which is a linear combination of the gradient of $U(x,\varepsilon)$. ■

The solution for the monopolist's problem is similar to what was already discussed in the previous section. Substituting (22) in the monopolist's objective function (5) implies that

$$\begin{aligned} \Pi(x,\varepsilon) &= \begin{bmatrix} x' & \varepsilon \end{bmatrix} \Theta\Theta^{-1}\nabla U(x,\varepsilon) - C\left(\Theta^{-1}\nabla U(x,\varepsilon)\right) - U(x,\varepsilon) \qquad (23) \\ &= \begin{bmatrix} x' & \varepsilon \end{bmatrix} \nabla U(x,\varepsilon) - C\left(\Theta^{-1}\nabla U(x,\varepsilon)\right) - U(x,\varepsilon) \end{aligned}$$

Therefore, the objective function the monopolist has is almost the same RC has for their problem. The difference in terms of the cost function can be easily handled because of the quadratic form assumed in (21). Thus, the new problem the monopolist has to solve is given by (24)

$$\phi(U) = \max_U \int_{\varepsilon \in \mathcal{E}} \int_{x \in X} \left(\begin{bmatrix} x' & \varepsilon \end{bmatrix} \nabla U(x, \varepsilon) - C\left(\Theta^{-1} \nabla U(x, \varepsilon)\right) - U(x, \varepsilon) \right) f(x, \varepsilon)\, dx d\varepsilon \tag{24}$$

$$s.t.\ U(x, \varepsilon)\ \ convex\ and$$

$$U(x, \varepsilon) \geq \bar{U}$$

and the first-order conditions that characterize the relaxed problem, i.e. ignoring the concavity constraint are the same as RC have obtained

$$\begin{aligned} \alpha(x, \varepsilon) &= div\left(\nabla_{(z,\xi)} S((x, \varepsilon), z(x, \varepsilon), \xi(x, \varepsilon)) f(x, \varepsilon)\right) + f(x, \varepsilon) \geq 0 \\ \beta(x, \varepsilon) &= -\left[\nabla_{(z,\xi)} S(x, \varepsilon, z(x, \varepsilon), \xi(x, \varepsilon)) f(x, \varepsilon)\right] \cdot n(x, \varepsilon) \geq 0 \end{aligned}$$

Substituting the definitions of the utility and cost functions, the above expressions become

$$\begin{aligned} \alpha(x, \varepsilon) &= div\left(\left(\begin{bmatrix} x' & \varepsilon \end{bmatrix} - \Upsilon \Theta^{-1} \nabla U(x, \varepsilon)\right) f(x, \varepsilon)\right) + f(x, \varepsilon) \geq 0 \quad (25) \\ \beta(x, \varepsilon) &= -\left(\begin{bmatrix} x' & \varepsilon \end{bmatrix} - \Upsilon \Theta^{-1} \nabla U(x, \varepsilon)\right) f(x, \varepsilon) \cdot n(x, \varepsilon) \geq 0 \end{aligned}$$

where Υ is a matrix derived from the cost function parameters.

The property on how the first-order conditions induce a measure on the space of characterisitics has not changed, given that the first-order conditions are the same as in RC. In particular, the measure of the non-participation region is still unitary (the argument is the same as used in RC and proposition 1). The point that remains to be shown is how the frontier between non-participation and participation regions (Ω_0, Ω_1) continues to be a straight line in the general problem with concavity restriction. Proposition 4 makes this point clear.

Proposition 4 *The frontier between Ω_0 and Ω_1 is a straight line.*

Proof. The proof is exactly the same as in proposition 2. In order to simplify notation, call $t = (x, \varepsilon) \in \Omega$, $s = (z, \xi) \in Z$ and Θ the coefficients matrix presented in te utility function. Take two points, t and $\tilde{t}$ in the boundary. Clearly, $s(t) = s\left(\tilde{t}\right) = 0$ and the corresponding prices is also zero. Now consider the conxex combination $\bar{t} = \alpha t + (1-\alpha)\tilde{t}$, $\alpha \in [0,1]$. Call $\bar{s} = s(\bar{t})$. Incentive compatibility says that

$$\begin{aligned} \bar{t}'\Theta\bar{s} - p(\bar{s}) &\geq \bar{t}'\Theta s - p(s) = 0 \\ \bar{t}'\Theta\bar{s} - p(\bar{s}) &\geq \bar{t}'\Theta\tilde{s} - p(\tilde{s}) = 0 \end{aligned}$$

The same applies for the agents t and $\tilde{t}$:

$$\begin{aligned} t'\Theta s - p(s) &\geq t'\Theta\bar{s} - p(\bar{s}) \\ 0 &\geq t'\Theta\bar{s} - p(\bar{s}) \end{aligned}$$

$$\begin{aligned} \tilde{t}'\Theta\tilde{s} - p(\tilde{s}) &\geq \tilde{t}'\Theta\bar{s} - p(\bar{s}) \\ 0 &\geq \tilde{t}'\Theta\bar{s} - p(\bar{s}) \end{aligned}$$

Taking the convex combination of these two equations we get

$$\begin{aligned} 0 &\geq \left[\alpha t + (1-\alpha)\tilde{t}\right]' \Theta\bar{s} - p(\bar{s}) \\ 0 &\geq \bar{t}'\Theta\bar{s} - p(\bar{s}) \end{aligned}$$

Therefore $\bar{t}'\Theta\bar{s} - p(\bar{s}) = 0$, which means that $\bar{t}$ also belongs to the boundary. Thus, the boundary is a straight line. ■

Having showing the linearity of the boundary between participation and non-participation, the other results in RC go along in the same way.

The next section enters on the real objective of this article, which is to show that the above model is identified using the implications from its solution. That will be the goal of the next pages.

3 Identification

The identification problem consists in recover the parameter matrices A, B, C and D, the attribute vector functions $z(x,\varepsilon)$, $\xi(x,\varepsilon)$, the cost matrices Γ, Λ and Φ and the distribution of the unobservable F_ε given that the observable objects are the vector of consumer's characteristics X, the vector of attributes these consumers are buying z and the prices they are paying p. The analysis will be done assuming $x \in \mathbb{R}$, $\varepsilon \in \mathbb{R}$, $z \in \mathbb{R}$ and $\xi \in \mathbb{R}$. It can be extended for the case when $x \in \mathbb{R}^k$,and $z \in \mathbb{R}^k$, as long as the unobservable attributes remain scalar. Finally, x is assumed to be independently distributed from ε. Thus, the model described in equation (20) can be written as:

$$\begin{aligned} U(x,\varepsilon,z,\xi) &= (\theta_1 x + \theta_2 \varepsilon) z + (\theta_3 x + \theta_4 \varepsilon) \xi - p(z,\xi) \qquad (26) \\ C(z,\xi) &= \frac{1}{2}\left(\gamma_1 z^2 + \gamma_2 \xi^2\right) + \gamma_3 z \xi \\ & x \perp\!\!\!\perp \varepsilon \end{aligned}$$

Assume the economist observes whether people participate or not consuming the good. This allow one to identify the Ω_0 region, using a well known result from Cosslett (1983) and Manski (1988). This result is summarized in proposition 5.

Proposition 5 *Assume the choice of consuming or not the good by each agent in the sample is observed. Then, the slope of the line that separates* Ω_0 *and* Ω_B *and* Ω_B *and* Ω_1 *is identified (up to a scale) and the distribution of the unobservable characteristic of the consumer,* ε*, is also identified in the support of* Ω_0*.*

Proof. The above result is a direct application of Cosslett (1983) and Manski (1988) result. As it was shown in proposition 4, the frontier that separates Ω_0 and Ω_B is given by a straight line. Thus, define a indicator variable I as

$$I = \begin{cases} 1, \ if \ \ (x, \varepsilon) \in \Omega_0 \\ \\ \ \ 0, \ otherwise \end{cases} \tag{27}$$

Call τ_0 as the maximum value ε admits in Ω_0 when $x = 0$. This is the point where the frontier between Ω_0 and Ω_B crosses the ε axis. The indicator function in (27) can be written as

$$\begin{aligned} I &= 1\left[\varepsilon + bx \leq \tau_0\right] \qquad (28) \\ &= 1\left[\varepsilon + bx - \tau_0 \leq 0\right] \end{aligned}$$

Normalizing $\tau_0 = 1$, it is possible to identify b and F_ε in the support of Ω_0. As the model also tells, the line that separates Ω_B and Ω_1 is parallel to the curve $\varepsilon + bx = \tau_0$. Therefore, the slope of the frontier between bunching and perfect screening is also identified (up to a scale). ■

The model exposed in the last section also tells something about the properties of the non-participation region: it must have a unit measure. This condition will allow the identification of the utility parameters θ_i, provided the economist have access to enough cross-sections. This result is shown in proposition 6.

Proposition 6 *Assume the analyst has access to, at least, four cross-sections of the data, each of them with different values for the b that characterizes the boundary of Ω_0. Then, the utility parameters θ_i are identified.*

Proof. To prove this result, one have to go back to (14), which characterizes the measure of the non-participation region. The surplus function is given by

$$S\left(x, \varepsilon, z, \xi\right) = \left(\theta_1 x + \theta_2 \varepsilon\right) z + \left(\theta_3 x + \theta_4 \varepsilon\right) \xi - \frac{1}{2}\left(\gamma_1 z^2 + \gamma_2 \xi\right) - \gamma_3 z \xi$$

thus, the gradient of the surplus with respect to the quality vector (z, ξ) can be calculated as

$$\nabla_{(z,\xi)} S\left(x, \varepsilon, z, \xi\right) = \begin{bmatrix} \theta_1 x + \theta_2 \varepsilon - \left(\gamma_1 z_1 + \gamma_3 \xi\right) \\ \theta_3 x + \theta_4 \varepsilon - \left(\gamma_3 z_1 + \gamma_2 \xi\right) \end{bmatrix}$$

Call the vector $\nabla_{(z,\xi)} S(x, \varepsilon, z(x, \varepsilon), \xi(x, \varepsilon)) f_x(x) f_\varepsilon(\varepsilon) = \nu(x, \varepsilon)$. The divergence of $\nu(x, \varepsilon)$ is given by

$$div(\nu(x, \varepsilon)) = \left(\theta_1 - \left(\gamma_1 \frac{\partial z}{\partial x} + \gamma_3 \frac{\partial \xi}{\partial x}\right) + \theta_4 - \left(\gamma_3 \frac{\partial z}{\partial \varepsilon} + \gamma_2 \frac{\partial \xi}{\partial \varepsilon}\right)\right) f_x(x) f_\varepsilon(\varepsilon) +$$

$$+ (\theta_1 x + \theta_2 \varepsilon - (\gamma_1 z + \gamma_3 \xi)) \frac{\partial}{\partial x} f_x(x) f_\varepsilon(\varepsilon) + (\theta_3 x + \theta_4 \varepsilon - (\gamma_3 z + \gamma_2 \xi)) f_x(x) \frac{\partial}{\partial \varepsilon} f_\varepsilon(\varepsilon)$$

The contour integral presented in (14) can be easily calculated because Ω_0 is given by a subset of the upper-right quadrant of the $\mathbb{R}^2$ plane. Thus, the boundary of Ω_0 is given by the curves $\sigma(x, \varepsilon) = x$ and $\sigma(x, \varepsilon) = \varepsilon$ and the correspondent normal to these curves are, $(0, -1)$ and $(-1, 0)$ respectively. Using the definition of $\beta(x, \varepsilon)$ in (9) it is possible to write the integrand of the countour integral in (14) as

$$\begin{aligned} \beta(x, 0) &= [\theta_3 x - (\gamma_3 z(x, 0) + \gamma_2 \xi(x, 0))] f_x(x) \\ \beta(0, \varepsilon) &= [\theta_2 \varepsilon - (\gamma_1 z(0, \varepsilon) + \gamma_3 \xi(0, \varepsilon))] f_\varepsilon(\varepsilon) \end{aligned}$$

In the non-participation region, agents consume zero units of the good. Therefore, $z = \xi = 0$, for every (x, ε) in Ω_0. This means that $\nabla_{(z,\xi)} z(x, \varepsilon) = \nabla_{(z,\xi)} \xi(x, \varepsilon) = 0$, for every (x, ε) in Ω_0.Therefore, the measure of U^* in Ω_0

becomes:

$$\int_0^1 \int_0^{1-bx} \left\{ (\theta_1 + \theta_4 + 1) f_x(x) f_\varepsilon(\varepsilon) + (\theta_1 x + \theta_2 \varepsilon) \frac{\partial}{\partial x} f_x(x) f_\varepsilon(\varepsilon) + (\theta_3 x + \theta_4 \varepsilon) f_x(x) \frac{\partial}{\partial \varepsilon} f_\varepsilon(\varepsilon) \right\} dx d\varepsilon + \tag{29}$$

$$+ \int_0^1 \theta_3 x f_x(x) dx + \int_0^{1/b} \theta_2 \varepsilon f_\varepsilon(\varepsilon) d\varepsilon = 1$$

Except for the θ_i, all the other objects in the above expression are observed or already identified: b can be identified for each cross section using proposition 5, as well as f_ε for the ε in Ω_0. Thus (29) is a linear system on θ_i and provided there is enough variation across each cross section, all the θ_i can be identified.

■

Propositions 5 and 6 represent a remarkable result if one compares to already estabilished results in the empirical IO literature. They show that, if one wants to assume the same quasilinear parametric specification defined in (20) then identification of the preference parameters and the distribution of the heterogeneity in the non-participation region is assured without assuming any parametric form for the distribution of the heterogeneity. It just needs to use the economics of the model to guarantee identification. Beggining with Berry, Levinsohn and Pakes (1995), a whole history of articles have been assuming parametric structures for the utility in the same way as described in (20) - for example, Petrin (2002), Goolsbee and Petrin (2004), Crawford and Shum

(2006) among many others. In all of these cases, the authors impose some parametric form in order to assure identification of the preference parameters (they usually assume the heterogeneity is given by a logit or multivariate normal distributions). The above results show that it is possible to relax the distribution assumption and let the data tell us about the shape of the heterogeneity.

Up to this moment there was no need to deal with the fact that only one component of the quality vector is observed (z). However, if one wants to back up the distribution of the unobservable component of the quality ξ, or the distribution of the unobservable individual characteristic ε over its full support, then this issue needs to be addressed. As it was shown in section 2, inside the bunching region, each bunch can be characterized by affine lines parallel to the bondary between Ω_0 and Ω_B. Thus, they can be summarized by an index, in the form $\varepsilon = t - bx$, where t is the value in the plane (x, ε) that the bunch crosses the vertical axis. Rearranging terms, a bunch could be defined as $t = \varepsilon + xb$, i.e. all the values in the Ω_B subset such that $t = \varepsilon + xb$ represent this bunch. For each of these bunches, (z, ξ) is constant, by definition. It means that, for the bunching region, (z, ξ) can be written as a function of t only. In other words

$$\begin{aligned} z &= z\left(t\right) = z\left(\varepsilon + xb\right) \\ \xi &= \xi\left(t\right) = \xi\left(\varepsilon + xb\right) \end{aligned} \tag{30}$$

Two important conclusions can be made based on (30). The first one is that, as (z, ξ) is a monotonic function of t (remember that the vector (z, ξ) is a linear combination of the gradient of U, which is a concave function), one can invert the z function and express $\xi = g(z)$. This gives a (unknown) relation between the observable component of the quality and the unobservable component of the attribute vector ξ which will be explored lately to help identify the distribution of ξ. The second result is build upon this last observation: once ξ is a one-to-one map from z to $\mathbb{R}$, then it must be the case that, in the bunching region, one should not observe variation in the price once z was fixed. This means that, if someone observes people with different characteristics x demanding a good with the same attribute z and the price of this good has no variation, it means that this good has the same unobservable attribute ξ and all of these consumers are bunched on the quality (z, ξ).

Using the fact that the boundary between the bunching region Ω_B and the screening region Ω_1 has the same slope as the boundary between Ω_0 and Ω_B, one may use the information on who belongs to the screening region Ω_1 and who does not in order to identify the ordinate τ of the last bunch on Ω_B. This is stated in the next proposition.

Proposition 7 *The boundary between the bunching region Ω_B and the screening region Ω_1 is identified.*

Proof. As stated previously, one can determine all the agents that are in the bunching region by determining the attributes that belong to the bunching region (i.e. those z whose prices do not present variation) and taking the agents that are consuming these goods. Then following indicator variable can be defined:

$$\begin{aligned} I &= \begin{cases} 1, \ if \ \ (x,\varepsilon) \in \Omega_0 \cup \Omega_B \\ \quad 0, \ otherwise \end{cases} \\ &= 1\left[\varepsilon + bx \leq \tau\right] \end{aligned}$$

Using again the result in Cosslett (1983) and Manski (1988), is possible to identify the distribution of the unobservable ε over the support of $\Omega_0 \cup \Omega_B$ and τ, remembering that the coefficient b was already identified in proposition 5. ■

The identification of the distribution of ε over the support of $\Omega_0 \cup \Omega_B$ and the parameter τ can now be used to identify the function $z = m(x,\varepsilon)$ in the bunching region. This is done using a result in Matzkin (2003).

Proposition 8 *The function $z = m(x,\varepsilon)$ is identified in the bunching region.*

Proof. From the previous discussion, z can be summarized by an index function:

$$z = m(x,\varepsilon) = m(\varepsilon + xb) = m(\tau)$$

m is a monotonic function of τ, an index. As τ was already identified in the last proposition, all the conditions required in Matzkin (2003) to identify the function $m(.,.)$ are satisfied. Therefore, $z = m(x, \varepsilon)$ is identified. ■

The next step is to identify the function that relates ξ to z. This is done in the next proposition.

Proposition 9 *The solution for the agent's problem in the bunching region allows one to identify the function* $\xi = g(z)$.

Proof. The agent solves the following problem:

$$\max_{z_1} (\theta_1 x + \theta_2 \varepsilon) z + (\theta_3 x + \theta_4 \varepsilon) g(z) - p(z)$$

The first-order conditions for this problem are given by

$$(\theta_1 x + \theta_2 \varepsilon) + (\theta_3 x + \theta_4 \varepsilon) g'(z) = p'(z)$$

From the previous proposition, one can substitute $\varepsilon = m^{-1}(z_1) - xb$. This implies that

$$\left[\theta_1 x + \theta_2 \left(m^{-1}(z) - xb\right)\right] + \left[\theta_3 x + \theta_4 \left(m^{-1}(z) - xb\right)\right] g'(z) = p'(z)$$

The function $p(z)$ can be traced using the prices and the attribute z that belong to the bunching region. Thus $p'(z)$ is identified. If one fix x and varies z for the values in Ω_B then he can obtain the function $g'(z)$, given that all the other parameters and functions in the expression above were already identified. Finally, as $\xi = 0$ when $z = 0$, then the function $g(z)$ is known because the value of the derivative of g is known for all the points in Ω_B and the scale of g is known at $z = 0$. ∎

The identification of ξ in the bunching region allows one to recover the cost parameters using the FOC of the monopolist in the bunching region. Remember that

$$\begin{aligned} \int\limits_{\Omega(q)} \alpha(t)\,dt + \int\limits_{\partial\Omega(q)\cap\partial\Omega} \beta(t)\,d\sigma(t) &= 0 \\ \int\limits_{\Omega(q)} t\alpha(t)\,dt + \int\limits_{\partial\Omega(q)\cap\partial\Omega} t\beta(t)\,d\sigma(t) &= 0 \end{aligned} \tag{31}$$

Substituting the expressions for $\alpha(t)$ and $\beta(t)$ we have

$$\begin{aligned} &\int\limits_0^\tau \left(1 + \theta_1 + \theta_4 - z'\left[b\left(\gamma_1 + \gamma_3 g'\right) + \left(\gamma_3 + \gamma_2 g'\right)\right]'\right) f_x(x)\, f_\varepsilon(\varepsilon)\, dx + \\ &\int\limits_0^\tau \left(\theta_1 x + \theta_2(\tau - bx) - (\gamma_1 z + \gamma_3 \xi)\right) \frac{\partial}{\partial x} f_x(x)\, f_\varepsilon(\varepsilon)\, dx + \\ \int\limits_0^\tau \quad &\left(\theta_3 x + \theta_4(\tau - bx) - (\gamma_3 z + \gamma_2 \xi)\right) f_x(x) \frac{\partial}{\partial \varepsilon} f_\varepsilon(\varepsilon)\, dx + \end{aligned} \tag{32}$$

$$\left[\theta_3\frac{\tau}{b}-(\gamma_3 z+\gamma_2\xi)\right]f_x(x)+\left[\theta_2\tau-(\gamma_1 z+\gamma_3\xi)\right]f_\varepsilon(\varepsilon)=0$$

In the bunch, z and ξ are constants. The only unknowns are the γ_i parameters and it is a linear equation on them. With 3 different cross-sections we identify the γ_i parameters.

The last part that remains to be identified is the distribution of the unobservable characteristic as well as the quality functions $z(x,\varepsilon)$ and $\xi(x,\varepsilon)$ in the screening region Ω_1, i.e. to have these functions identified in the remaining of the support of Ω. The next two lemmas will show that using average derivative methods as the ones used by Powell, Stock and Stoker (1989), Ekeland, Heckman and Nesheim (2004), Buera (2006) it is possible to recover the derivatives of $z(x,\varepsilon)$ and $\xi(x,\varepsilon)$ in the domain of Ω_1. Then, proposition 12 will use the results of these two lemmas and the first-order conditions of the agent's problem in the screening region to show that $z(x,\varepsilon)$, $\xi(x,\varepsilon)$ and the distribution $F_\varepsilon(\varepsilon)$ in Ω_1 are identified.

Take the joint distribution of the observed attribute z and x in the screening region. Lemma 1 shows that, given the joint distribution $F(z,x)$ the derivative $\frac{\partial z}{\partial x}$ is identified.

Lemma 10 *The partial derivative of the observed attribute z with respect to the observed characteristc x, $\left(\frac{\partial z}{\partial x}\right)$ is identified in the support of Ω_1 up to a* *normalization.*

Proof. Call the function $z(x,\varepsilon)$ as $g(x,\varepsilon)$. The model says $g(x,\varepsilon)$ is the gradient of a concave function, therefore it is monotone in its coordinates. Thus, g can be inverted, i.e.

$$z = g(x,\varepsilon) \rightarrow \varepsilon = g^{-1}(x,z)$$

Now, take the joint distribution of (z,x) in Ω_1. It can be written as

$$\begin{aligned}
F_{z\in\Omega_1}(z,x) &= F^1(z,x) = \Pr\left(Z \le z \,| z \in \Omega_1\right) \\
&= \Pr\left(g(x,\varepsilon) \le z \,| z \in \Omega_1\right) \\
&= \Pr\left(\varepsilon \le g^{-1}(x,z) \,| \varepsilon \in \Omega_1\right) \\
&= \int_{\tau-bx}^{g^{-1}(x,z)} f_\varepsilon(\varepsilon)\, d\varepsilon
\end{aligned}$$

The derivatives of F^1 with respect to z and x are:

$$\begin{aligned}
F_z^1 &= \left(\frac{\partial}{\partial z} g^{-1}(x,z)\right) f_\varepsilon\left(g^{-1}(x,z)\right) \\
F_x^1 &= \left(\frac{\partial}{\partial x} g^{-1}(x,z)\right) f_\varepsilon\left(g^{-1}(x,z)\right) + b f_\varepsilon(\tau - bx)
\end{aligned}$$

We already identified $bf_\varepsilon(\tau - bx)$ in the bunching region. Thus, if we divide the two equations above we can identify the ratio

$$\frac{\frac{\partial}{\partial z} g^{-1}(x,z)}{\frac{\partial}{\partial x} g^{-1}(x,z)} = \frac{F_z^1}{F_x^1 - b f_\varepsilon(\tau - bx)}$$

However, using the implicit function theorem, the expression above can be related to the original function $g(x,\varepsilon)$. It says that

$$\frac{\partial \varepsilon}{\partial z} = \frac{\partial}{\partial z} g^{-1}(x,z) = \frac{1}{\frac{\partial}{\partial \varepsilon} g(x,\varepsilon)}$$

$$\frac{\partial \varepsilon}{\partial x} = \frac{\partial}{\partial x} g^{-1}(x,z) = -\frac{\frac{\partial}{\partial x} g(x,\varepsilon)}{\frac{\partial}{\partial \varepsilon} g(x,\varepsilon)}$$

Thus, $\frac{\partial}{\partial x} g(x,\varepsilon)$ is identified. In the previous notation, this is equal to $\frac{\partial z_1}{\partial x}$. The missing part is how to relate z to ε obtained here. It is known how ε relates to z in the boundary between Ω_B and Ω_1. However, for the other points in the domain, $dz = g_\varepsilon(x,\varepsilon)\, d\varepsilon$. So, unless one identifies $g_\varepsilon(x,\varepsilon)$, $\frac{\partial z}{\partial x}$ is identified just up to a normalization. ■

Lemma 11 *Given the joint distribution of prices, attribute z and characteristics x, the derivative of prices with respect to the observed quality is identified. Also, the relation between the derivative of price with respect to ξ and the derivative $\frac{\partial \xi}{\partial x}$ are identified.*

Proof. Call the function $p(z,\xi)$ as $h(z,\xi)$. Again, this function is increasing in both coordinates (otherwise, it would rule out the optimality of the solution of the monopoly). Thus,

$$p = h(z, \xi) \rightarrow \xi = h^{-1}(z, p)$$

Also, call the function $\xi(x, \varepsilon)$ as $s(x, \varepsilon)$. Using the same arguments used in lemma 1, write

$$\xi = s(x, \varepsilon) \rightarrow \varepsilon = s^{-1}(x, \xi)$$

Moreover, the joint distribution of (p, z, x) in Ω_1 can be written as

$$\begin{aligned}
F_{(p,z)\in\Omega_1}(z, x) &= F^2(p, z, x) = \Pr(P \leq p \,|(p, z) \in \Omega_1) \\
&= \Pr(h(z, \xi) \leq p \,|(p, z) \in \Omega_1) \\
&= \Pr\left(\xi \leq h^{-1}(z, p) \,|\xi \in \Omega_1\right) \\
&= \Pr\left(s(x, \varepsilon) \leq h^{-1}(z, p) \,|\xi \in \Omega_1\right) \\
&= \Pr\left(\varepsilon \leq s^{-1}\left(x, h^{-1}(z, p)\right) |\varepsilon \in \Omega_1\right) \\
&= \int_{\tau - bx}^{s^{-1}\left(x, h^{-1}(z,p)\right)} f_\varepsilon(\varepsilon)\, d\varepsilon
\end{aligned}$$

Repeating the procedure done in lemma 10, one can differentiate F^2 with respect to p, z and x. This gives the following expressions:

$$\begin{aligned}
F_p^2 &= \left(\frac{\partial}{\partial \xi} s^{-1}\left(x, h^{-1}(z,p)\right)\right)\left(\frac{\partial}{\partial p} h^{-1}(z,p)\right) f_\varepsilon\left(s^{-1}\left(x, h^{-1}(z,p)\right)\right) \\
F_{z_1}^2 &= \left(\frac{\partial}{\partial \xi} s^{-1}\left(x, h^{-1}(z,p)\right)\right)\left(\frac{\partial}{\partial z_1} h^{-1}(z,p)\right) f_\varepsilon\left(s^{-1}\left(x, h^{-1}(z,p)\right)\right) \\
F_x^2 &= \left(\frac{\partial}{\partial x} s^{-1}\left(x, h^{-1}(z,p)\right)\right) f_\varepsilon\left(s^{-1}\left(x, h^{-1}(z,p)\right)\right) + b f_\varepsilon(\tau - bx)
\end{aligned}$$

From lemma 10 $b f_\varepsilon(\tau - bx)$ its already known. Thus, the above system allow us to identify two expressions:

$$\frac{\frac{\partial}{\partial p} h^{-1}(z,p)}{\frac{\partial}{\partial z} h^{-1}(z,p)} = \frac{F_p^2}{F_z^2}$$

and

$$\frac{\left(\frac{\partial}{\partial z_2} s^{-1}\left(x, h^{-1}(z_1,p)\right)\right)\left(\frac{\partial}{\partial z_1} h^{-1}(z_1,p)\right)}{\frac{\partial}{\partial x} s^{-1}\left(x, h^{-1}(z_1,p)\right)} = \frac{F_{z_1}^2}{F_x^2 - b f_\varepsilon(\tau - bx)}$$

Using the same arguments as in lemma 10, the first expression is the same as $-\frac{1}{p_1}$, i.e.

$$\frac{\frac{\partial}{\partial p} h^{-1}(z_1,p)}{\frac{\partial}{\partial z_1} h^{-1}(z_1,p)} = -\frac{1}{\frac{\partial p}{\partial z_1}}$$

so, $\frac{\partial p}{\partial z_1}$ is identified. The second expression gives us

$$\frac{\left(\frac{\partial}{\partial \xi} s^{-1}\left(x, h^{-1}(z,p)\right)\right)\left(\frac{\partial}{\partial z} h^{-1}(z,p)\right)}{\frac{\partial}{\partial x} s^{-1}\left(x, h^{-1}(z,p)\right)} = \frac{\frac{\partial h}{\partial z}}{\frac{\partial h}{\partial \xi}\frac{\partial s}{\partial x}}$$

In the previous notation, $\frac{\partial h}{\partial z} = \frac{\partial p}{\partial z}$, $\frac{\partial h}{\partial \xi} = \frac{\partial p}{\partial \xi}$ and $\frac{\partial s}{\partial x} = \frac{\partial \xi}{\partial x}$. ■

The previous lemmas have shown that some important functions can be infered non parametrically, using only the data we observe, even though there is still missing a normalization to link variations in z to ε and variations in p and z to ξ. This is finally obtained if one takes into account the first-order conditions of the agents in the screening region. That is the last step needed to fully identify this model and is presented in the next proposition.

Proposition 12 *The distribution of the unobservable characteristic ε is identified.*

Proof. The FOC for the agent in the bunching region are:

$$\begin{aligned} \theta_1 x + \theta_2 \varepsilon &= \frac{\partial p}{\partial z} \\ \theta_3 x + \theta_4 \varepsilon &= \frac{\partial p}{\partial \xi} \end{aligned}$$

and, for this region, it is possible to differentiate again with respect to (x, ε)

$$\begin{aligned}\theta_1 &= p_{11}\frac{\partial z}{\partial x} + p_{12}\frac{\partial \xi}{\partial x} \\ \theta_2 &= p_{11}\frac{\partial z}{\partial \varepsilon} + p_{12}\frac{\partial \xi}{\partial \varepsilon} \\ \theta_3 &= p_{21}\frac{\partial z}{\partial x} + p_{22}\frac{\partial \xi}{\partial x} \\ \theta_4 &= p_{21}\frac{\partial z}{\partial \varepsilon} + p_{22}\frac{\partial \xi}{\partial \varepsilon}\end{aligned} \tag{33}$$

Given that p_1 was identified in lemma 11, it is easy to calculate all its derivatives. Thus, using θ_1 and $\frac{\partial z}{\partial x}$ already identified before and the derivatives of p_1 one obtains $\frac{\partial \xi}{\partial x}$ using the first equation in (33). The implementation result in (22) shows that

$$\begin{aligned}\begin{bmatrix} z \\ \xi \end{bmatrix} &= \Theta^{-1}\nabla U(x, \varepsilon) \\ \begin{bmatrix} z \\ \xi \end{bmatrix} &= \frac{1}{\theta_1\theta_4 - \theta_2\theta_3}\begin{bmatrix} \theta_4 & -\theta_2 \\ -\theta_3 & \theta_1 \end{bmatrix}\begin{bmatrix} U_x(x, \varepsilon) \\ U_\varepsilon(x, \varepsilon) \end{bmatrix}\end{aligned} \tag{34}$$

Differentiating (34) with respect to the vector (x, ε) implies the following system

$$\begin{aligned}
\frac{\partial z}{\partial x} &= \frac{1}{\theta_1\theta_4 - \theta_2\theta_3}\left(\theta_4 U_{xx}(x,\varepsilon) - \theta_2 U_{\varepsilon x}(x,\varepsilon)\right) \\
\frac{\partial \xi}{\partial x} &= \frac{1}{\theta_1\theta_4 - \theta_2\theta_3}\left(\theta_1 U_{\varepsilon x}(x,\varepsilon) - \theta_3 U_{xx}(x,\varepsilon)\right) \\
\frac{\partial z}{\partial \varepsilon} &= \frac{1}{\theta_1\theta_4 - \theta_2\theta_3}\left(\theta_4 U_{x\varepsilon}(x,\varepsilon) - \theta_2 U_{\varepsilon\varepsilon}(x,\varepsilon)\right) \\
\frac{\partial \xi}{\partial \varepsilon} &= \frac{1}{\theta_1\theta_4 - \theta_2\theta_3}\left(\theta_1 U_{\varepsilon\varepsilon}(x,\varepsilon) - \theta_3 U_{x\varepsilon}(x,\varepsilon)\right)
\end{aligned} \tag{35}$$

Using symmetry arguments, $U_{\varepsilon x}(x,\varepsilon) = U_{x\varepsilon}(x,\varepsilon)$. Thus, the values of $\frac{\partial z}{\partial x}$ and $\frac{\partial \xi}{\partial x}$ combined with the two first equations in (35) and the symmetry of the U function allows one to identify $\frac{\partial z}{\partial \varepsilon}$ and $\frac{\partial \xi}{\partial \varepsilon}$. Finally, using these last two derivatives, one can substitute in the first equation in (33) and calculate the price derivatives with respect to the unobservable attribute p_{21} and $p_{22.}$

As a last comment $\frac{\partial z}{\partial \varepsilon}$ is the scale factor needed to fully identify $\frac{\partial z}{\partial x}$ in lemma 10 ($\frac{\partial z}{\partial \varepsilon} = g_\varepsilon(x,\varepsilon)$). Therefore, lemmas 10 and 11 and the last proposition constitute a recursive procedure to fully identify $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial \varepsilon}$ in the whole screening region Ω_1 (which implies $\frac{\partial \xi}{\partial x}$ and $\frac{\partial \xi}{\partial \varepsilon}$ are also identified in Ω_1). Moreover, the distribution of ε becomes identified because

$$\begin{aligned} F_{\varepsilon}\left(\varepsilon \left| \varepsilon \in \Omega_1 \right.\right) &= \Pr\left(\varepsilon \leq e \left| \varepsilon \in \Omega_1 \right.\right) \\ &= \Pr\left(g^{-1}\left(x, z\right) \leq e \left| \varepsilon \in \Omega_1 \right.\right) \\ &= \Pr\left(z \leq g\left(x, e\right) \left| \varepsilon \in \Omega_1 \right.\right) \\ &= F_z\left(g\left(x, e\right)\right) \end{aligned}$$

■

A final remark might be done here that is, even though the characterization of the equilibrium in the screening region involves a partial differential equation, proposition 12 shows that it is not necessary to solve it in order to obtain identification in this region. This is a very nice result provided that the PDE that characterizes the solution (17) can be a very complex equation, depending on the shape of the distribution function $F_x\left(x\right)$ and $F_{\varepsilon}\left(\varepsilon\right)$. This is also helpful in terms of the estimation algorithm that will be develop for this model. Most of the empirical IO literature models involve computing complicated maximum likelihoods that, in some part of the algorithm, requires the solution of the first-order conditions. This is not necessary here and, as one can see, most of the parameters and functions could be estimated with standard nonparametric estimation procedures.

4 Conclusion

This paper contributes for the literature on identification of hedonic models. It shows that using all the economics of the underlying model it is possible to identify preference and cost parameters, even in the presence of unobservable attributes. The approach used here differs from previous papers in the literature by focusing in the information the bunching region contains. By using the equilibrium conditions that generates the bunching region it is possible to recover most of the parameters of interest and, more importantly, it places restrictions on the values the unobserved attributes may attain.

However, there are some caveats that one might be concerned. In particular, these results applies to markets where a monopoly can screen its consumers by differentiation. More interesting market structures, as it is the case of competitive and oligopoly settings, might have differences in the way the equilibrium conditions would help in the identification of the model. In particular, the identification results provided by the bunching restrictions may be different in these other structures. This is an open question that should be addressed.

References

[1] Berry, Steven, James Levinsohn and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63: 841-890.

[2] Carlier, Guillaume and T. Lanchand-Robert. 2001. "Regularity of Solutions for some Variational Problems Subject to a Convexity Constraint." Unpublished manuscript, Universite Paris IX Dauphine, CEREMADE.

[3] Cosslett, Stephen. 1983. "Distribution-Free Maximum Likelihood Estimation of Binary Choice Model." *Econometrica* 51: 765-782.

[4] Ekeland, Ivar. 2005. "Existence, Uniqueness and Efficiency of Equilibrium in Hedonic Markets with Multidimensional Types." Unpublished manuscript, University of British Columbia, Pacific Institute of Mathematical Sciences.

[5] Ekeland, Ivar, James J. Heckman and Lars Nesheim. 2004. "Identification and Estimation of Hedonic Models." *Journal of Political Economy* 112: 60-108.

[6] Gorman, William. 1980. "A Possible Procedure for Analyzing Quality Differentials in the Egg-Market." *Review of Economic Studies* 47: 843-856.

[7] Heckman, James J., Rosa Matzkin and Lars Nesheim. 2006. "Nonpara-

metric Estimation of Nonadditive Hedonic Models." Unpublished manuscript, The University of Chicago, Department of Economics.

[8] Lancaster, Kelvin. 1966. "A New Approach to Consumer Theory." *Journal of Political Economy* 74: 132-157.

[9] Manski, Charles. 1988. "Identification of Binary Response Models." *Journal of the American Statistical Association* 83: 729-738.

[10] Petrin, Amil. 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy* 110: 705:729.

[11] Powell, James, James Stock and Thomas Stocker. 1989. "Semiparametric Estimation of Index Coefficients." *Econometrica* 57: 1403-1430

[12] Rochet, Jean-Charles. 1987. "A Necessary and Sufficient Condition for Rationalizability in a Quasi-Linear Context." *Journal of Mathematical Economics* 16: 191-200.

[13] Rochet, Jean-Charles and Philippe Chone. 1998. "Ironing, Sweeping and Multidimensional Screening." *Econometrica* 66: 783-826.

[14] Rockafellar, Ralph. 1972. "Convex Analysis." Princeton University Press, Princeton, NJ.

[15] Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82:

32-50.

[16] Tinbergen, Jan. 1956. "On the Theory of Income Distribution." *Weltwirtschaftliches Archiv* 77: 155-173