CEPS INSTEAD

WORKING PAPERS

Entropy methods for identifying hedonic models

Arnaud DUPUY[1]
Alfred GALICHON[2]
Marc HENRY[3]

CEPS/INSTEAD, Luxembourg[1]
Sciences Po, France[2]
Penn State University, United States[3]

Working Paper No 2014-07
June 2014

Entropy methods for identifying hedonic models

Arnaud Dupuy*
CEPS/INSTEAD

Alfred Galichon†
Sciences Po

Marc Henry‡
Penn State University

June 12, 2014
Dedicated to Ivar Ekeland on his 70th birthday.

Abstract

This paper contributes to the literature on hedonic models in two ways. First, it makes use of Queyranne's reformulation of a hedonic model in the discrete case as a network flow problem in order to provide a proof of existence and integrality of a hedonic equilibrium and efficient computational techniques of hedonic prices. Second, elaborating on entropic methods developped in Galichon and Salanié (2014), this paper proposes a new identification strategy for hedonic models in a single market. This methodology allows one to introduce heterogeneities in both consumers' and producers' attributes and to recover producers' profits and consumers' utilities based on the observation of production and consumption patterns and the set of hedonic prices.

Keywords: Hedonic models, Entropic methods, Identification.

JEL codes: D12, J3, L11.

*E-mail: arnaud.dupuy@ceps.lu.

†Corresponding author. E-mail: alfred.galichon@sciencespo.fr. Galichon's research has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no 313699, and from FiME, Laboratoire de Finance des Marchés de l'Energie.

‡E-mail: marc.henry@psu.edu. Henry's research is supported by SSHRC Grant 435-2013-0292 and NSERC Grant 356491-2013.

1 Introduction

Starting with Court (1941), Griliches (1961) and Lancaster (1966), a large literature has aimed at providing a theoretical framework for pricing the attributes of highly differentiated goods. While this literature was initially mainly empirical in nature and early contributions lacked a proper theoretical setting, the first theoretical treatments of hedonic models appeared in Tinbergen (1956) and Rosen (1974). Tinbergen (1956) presents a stylized model in which preferences are quadratic and attributes normally distributed. Rosen (1974) showed the theoretical relation of hedonic prices to marginal willingness to produce and marginal willingness to consume. Hedonic models have also been used to study the pricing of highly differentiated products such as houses (Kain and Quigley, 1970), wine and champagne (Golan and Shalit, 1993), automobiles quality (Triplett, 1969) among others but also set forth a new literature on the Value of Statistical Life following Thaler and Rosen's (1976) original idea of seeing jobs attributes and in particular "risk taken on the job" as a vector of hedonic attributes valued on the labor market. More recently, significant progress on the understanding of the properties of hedonic models (properties of an equilibrium, identification of deep parameters etc.) has been achieved. These developments are to a large extent due to Ivar Ekeland's contributions, see e.g. Ekeland et al. (2004) and Ekeland (2010), and it is a pleasure to dedicate to him the present piece of work in recognition of our intellectual debt for him.

In this paper we contribute to the hedonic literature in two ways. First, we elaborate on an idea of Maurice Queyranne who reformulated the hedonic model in the discrete case as a network flow problem. This reformulation allows us to derive results on existence of a hedonic equilibrium in the discrete case, and it allows the use of powerful computational techniques to solve for the equilibrium. Second, building on recent development in the matching model literature and in particular the seminal contribution due to Choo and Siow (2006) generalized by Galichon and Salanié (2014), we introduce heterogeneities (unobserved by the econometrician) in producer and consumer types. This formalism has two advantages: (i) it allows for the incorporation of unobserved heterogeneity in the producers and consumers characteristics, and (ii) it provides straightforward identification results. Indeed, we follow Galichon and Salanié in making use of the convex duality in discrete choice problems to recover utilities from choice probabilities on both side of the market.

The remainder of the paper is organized as follows. Section 2 discusses the properties of an equilibrium in hedonic model and its reformulation as a network flow problem. Section 3 introduces a model with unobserved heterogeneities on both sides of the market and studies the identification of preference parameters. The discussion in Section 4 concludes the paper.

2 Equilibrium, existence and properties

2.1 Hedonic equilibrium

The model. Throughout this paper, $\mathcal{X}$ is the set of observable types of producers of a given good, and $\mathcal{Y}$ the set of observable types of consumers of that good. This good comes in various qualities; let $\mathcal{Z}$ be the set of the good's qualities. The sets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are assumed to be finite. It is assumed that there is a supply n_x (resp. m_y) of producers (resp. consumers) of type x (resp. y). It is assumed that producers (resp. consumers) can produce (consume) at most one unit of good. They have the option not to participate in the market, in which case they choose $z = 0$.

For example, hedonic models can be used to model the market for fine wines[1]. In that case, $\mathcal{X}$ may be the set of observable characteristics of wine producers (say, grapes used, average amount of sunshine, and harvesting technology), and $\mathcal{Y}$ may be the set of observable characteristics of wine consumers (say country and purchasing channel). $\mathcal{Z}$ will be the quality of the wine (say acidity, sugar content, expert rating).

Let p_z be the price of the good of quality z. If a producer of type x produces the good in quality z, the payoff to the producer is $\alpha_{xz} + p_z$, where $\alpha_{xz} \in \mathbb{R} \cup \{-\infty\}$ is the producer's productivity (the opposite of a production cost). Similarly, if the consumer of type y consumes the good in quality z, the payoff to the consumer is $\gamma_{yz} - p_z$, where $\gamma_{yz} \in \mathbb{R} \cup \{-\infty\}$ is the utility of the consumer[2]. Producers and consumers who do not participate in the market get a surplus of zero.

[1]We are confident Ivar will approve of this choice of example.

[2]Note that in this setup, the utility of agents on each side of the market does not depend directly on the type of the agent with whom they match, only through the type of the contract. A more general framework where α and γ depend simultaneously on x, y and z is investigated in Dupuy, Galichon and Zhao (2014).

Supply and demand. Let μ_{xz} be the supply function, that is the number of producers of type x offering quality z; similarly, μ_{zy} is the demand function, the number of consumers of type y demanding quality z. One has

$$\sum_{z \in \mathcal{Z}} \mu_{xz} \leq n_x \ , \ \sum_{z \in \mathcal{Z}} \mu_{zy} \leq m_y$$

where the difference between the right-hand side and the left-hand side of these inequalities is the number of producers of type x (resp. consumers of type y) deciding to opt out of the market. The market clearing condition for quality z expresses that the total quantity of good of quality z produced is equal to the total quantity consumed, that is

$$\sum_{x \in \mathcal{X}} \mu_{xz} = \sum_{y \in \mathcal{Y}} \mu_{zy}$$

(it is assumed that there is no free disposal; if free disposal is assumed the equality is replaced by $\geq$ in the expression).

Equilibrium prices. At equilibrium, each producer x will optimize its production behavior given the price vector (p_z); hence if producing quality z' yields strictly more profit than producing quality z, then quality z will not be produced at all; that is $\alpha_{xz} + p_z < \alpha_{xz'} + p_{z'}$ for some z' implies $\mu_{xz} = 0$. A similar condition holds for consumers.

One can now state a formal definition.

Definition 2.1 (Hedonic equilibrium). *Let $(p_z)_{z \in \mathcal{Z}}$ be a price vector, μ_{xz} a supply function, and μ_{yz} a demand function. Then:*

(a) (p, μ) is called a hedonic equilibrium *whenever the following three conditions are all verified:*

(i) People counting: the number of producers of type x actually participating in the market does not exceed the total number of agents of type x, and similarly for consumers of type y. That is, for any x and y,

$$\sum_{z} \mu_{xz} \leq n_x \ , \ \sum_{z} \mu_{zy} \leq m_y. \tag{2.1}$$

(ii) Market clearing: for any z, supply for quality z will equate demand, that is

$$\sum_{x \in \mathcal{X}} \mu_{xz} = \sum_{y \in \mathcal{Y}} \mu_{zy}. \tag{2.2}$$

(iii) Rationality: no producer or consumer chooses a quality that is sub-optimal. That is, given (x, y, z, z'), then

$$\begin{aligned} \alpha_{xz} + p_z &< \alpha_{xz'} + p_{z'} \text{ implies } \mu_{xz} = 0 \\ \gamma_{yz} - p_z &< \gamma_{yz'} - p_{z'} \text{ implies } \mu_{yz} = 0. \end{aligned}$$

(b) If n_x and m_y are integer, (p, μ) is called an integral equilibrium *whenever (p, μ) is a hedonic equilibrium and all the entries μ are integers.*

The indirect utility u_x of a producer of type x and the indirect utility v_y of a consumer of type y are given by $u_x = G_x(\alpha_{x.} + p_.)$ and $v_y = H_y(\gamma_{.y} - p_.)$, where G and H are respectively the indirect surpluses of producers and consumers, defined by

$$G_x(U_{x.}) = \max_z U_{xz}^+ \text{ and } H_y(V_{.y}) = \max_z V_{yz}^+ \tag{2.3}$$

where a^+ denotes the positive part of a.

As a result, if p_z is an equilibrium price, then for all x, y and z, $u_x \geq \alpha_{xz} + p_z$ and $v_y \geq \gamma_{zy} - p_z$, thus $\gamma_{yz} - v_y \leq p_z \leq u_x - \alpha_{xz}$. Therefore:

Proposition 2.1. *For a given optimal solution u and v, the set of equilibrium prices are the prices p_z such that*

$$p_z^{\max} \geq p_z \geq p_z^{\min}. \tag{2.4}$$

where

$$p_z^{\min} = \max_y(\gamma_{yz} - v_y) \text{ and } p_z^{\max} = \min_x(u_x - \alpha_{xz}). \tag{2.5}$$

As a result, $u_x + v_y \geq \alpha_{xz} + \gamma_{yz}$, hence

$$u_x + v_y \geq \max_z(\alpha_{xz} + \gamma_{yz}), \tag{2.6}$$

thus, as observed by Chiappori, McCann and Nesheim (2010), u and v are the stable payoffs of the assignment game in transferable utility with surplus $\Phi_{xy} = \max_z(\alpha_{xz} + \gamma_{yz})$. In the next paragraph, we shall go beyond this equivalence by seeing a reformulation of the hedonic model as a network flow problem.

2.2 Network flow formulation

Interestingly, as understood by Maurice Queyranne, the hedonic equilibrium problem can be reformulated as a network flow problem. This reformulation will be of particular interest since, as we show below, it help us establish the existence of a hedonic equilibrium and provides the building blocks to compute an equilibrium. While the present exposition is as self-contained as possible, a good reference for network flow problems is Ahuja, Magnanti and Orlin (1993).

The network. Define a set of *nodes* by $\mathcal{N} = \mathcal{X} \cup \mathcal{Z} \cup \mathcal{Y}$, and a set of *arcs* $\mathcal{A}$ which is a subset of $\mathcal{N} \times \mathcal{N}$ and is such that if $ww' \in \mathcal{A}$, then $w'w \notin \mathcal{A}$. Here, the set of arcs is $\mathcal{A} = (\mathcal{X} \times \mathcal{Z}) \cup (\mathcal{Y} \times \mathcal{Z})$.

A *vector* is defined as an element of $\mathbb{R}^{\mathcal{A}}$. Here, we introduce the following *direct surplus vector*

$$\phi_{ww'} \quad : \quad = \alpha_{xz} \text{ if } w = x \text{ and } w' = z \tag{2.7a}$$
$$\phi_{ww'} \quad : \quad = \gamma_{yz} \text{ if } w = z \text{ and } w' = y. \tag{2.7b}$$

For two nodes w and w', a *path* from w to w' is a chain

$$(w_0 w_1), (w_1 w_2), ..., (w_{T-2} w_{T-1}), (w_{T-1} w_T)$$

such that $w_i w_{i+1} \in \mathcal{A}$ for each i. T is the *length* of the path. Here, the only nontrivial paths are of length 2 and are of the form $(xz), (zy)$ where $x \in \mathcal{X}$, $z \in \mathcal{Z}$ and $y \in \mathcal{Y}$.

For two nodes w and w', we define the *reduced surplus*, or *indirect surplus* as the surplus associated to the optimal path from w to w'. Here, for $x \in \mathcal{X}$, $y \in \mathcal{Y}$, the indirect suplus Φ_{xy} of producer x and consumer y is

$$\Phi_{xy} := \max_{z \in \mathcal{Z}} (\alpha_{xz} + \gamma_{yz}). \tag{2.8}$$

For $w \in \mathcal{N}$, we let N_w be the algebraic quantity of mass leaving the network at w. Hence N_w is the flow of mass being consumed ($N_w > 0$) or produced ($N_w < 0$) at w. The nodes such that $N_w < 0$ (resp. $N_w = 0$ and $N_w > 0$) are called the source nodes, whose set is denoted $\mathcal{S}$ (resp.

intermediate nodes $\mathcal{I}$ and target nodes $\mathcal{T}$). Here, for $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $z \in \mathcal{Z}$, we set

$$N_x := -n_x \ , \ N_y := m_y \ , \ N_z := 0 \tag{2.9}$$

so that the set of source nodes is $\mathcal{S} := \mathcal{X}$, the set of intermediate nodes is $\mathcal{I} := \mathcal{Z}$, and the set of target nodes is $\mathcal{T} := \mathcal{Y}$.

Gradient, flows. We define a *potential* as an element of $\mathbb{R}^{\mathcal{N}}$. We define the *gradient matrix* as the matrix of general term ∇_{aw}, $a \in \mathcal{A}$, $w \in \mathcal{N}$ such that

$$\nabla_{aw} = -1 \text{ if } a \text{ is out of } w,\ \nabla_{aw} = 1 \text{ if } a \text{ is into } w,\ \nabla_{aw} = 0 \text{ else,}$$

so that, for a potential $f \in \mathbb{R}^{\mathcal{N}}$, ∇f is the vector such that for $a = ww' \in \mathcal{A}$, one has $(\nabla f)_{ww'} = f_{w'} - f_w$. Here, set the potential of surpluses U as

$$U_x := -u_x \ , \ U_z := -p_z \ , \ U_y := v_y, \tag{2.10}$$

and

$$(\nabla U)_{xz} = u_x - p_z \text{ and } (\nabla U)_{zy} = v_y + p_z. \tag{2.11}$$

We define the *divergence* matrix ∇^* (sometimes also called *node-edge*, or *incidence matrix*[3]) as the transpose of the gradient matrix: $\nabla^*_{xa} := \nabla_{ax}$. As a result, for a vector v,

$$(\nabla^* v)_{ww'} = \sum_z v_{zw'} - \sum_z v_{wz}.$$

A *flow* is a nonnegative vector $\mu \in \mathbb{R}^{\mathcal{A}}_+$ that satisfies the *balance of mass equation*[4], that is

$$\begin{aligned} (N - \nabla^*\mu)_w &\geq 0,\ w \in \mathcal{S} & (2.12)\\ (N - \nabla^*\mu)_w &= 0,\ w \in \mathcal{I} & (2.13)\\ (N - \nabla^*\mu)_w &\leq 0,\ w \in \mathcal{T} & (2.14) \end{aligned}$$

[3]The node-edge matrix is usually denoted A; our notations ∇^* and terminology are chosen to stress the analogy with the corresponding differential operators in the continuous case.

[4]In most physical systems, mass is conserved and the balance equation has the more usual form of *Kirchoff's law* $\nabla^*\mu = N$. However, in the present setting, producers and consumers have an option not to participate in the market, hence $\nabla^*\mu = N$ is replaced by Eqs. (2.12)-(2.14).

Here, $\mu : (\mu_{xz}, \mu_{zy})$ is a flow if and only if μ_{xz} and μ_{zy} satisfy the people counting and market clearing equations, that is

$$\sum_z \mu_{xz} \leq n_x \ , \ \sum_z \mu_{zy} \leq m_y \text{ and } \sum_{x \in \mathcal{X}} \mu_{xz} = \sum_{y \in \mathcal{Y}} \mu_{zy}.$$

Maximum surplus flow. We now consider the *maximum surplus flow problem*, that is

$$\begin{aligned} & \max_{\mu \in \mathbb{R}_+^{\mathcal{A}}} \sum_{a \in A} \mu_a \phi_a \\ s.t. \qquad & \mu \text{ satisfies (2.12), (2.13), (2.14),} \end{aligned} \tag{2.15}$$

whose value coincides with the value of its dual version, that is

$$\begin{aligned} & \min_{U \in \mathbb{R}^{\mathcal{N}}} \sum_{w \in \mathcal{N}} U_w N_w \\ s.t. \qquad & U_w \geq 0, \ \forall w \in \mathcal{S} \cup \mathcal{T} \\ & \nabla U \geq \phi, \end{aligned} \tag{2.16}$$

and by complementary slackness, for $w \in \mathcal{S} \cup \mathcal{T}$, $U_w > 0$ implies $N_w = (\nabla^* \mu)_w$. A standard result is that if N has only integral entries, then (2.15) has an integral solution μ.

Here the solution U of (2.16) is related to the solution to the hedonic model by Equations (2.10), that is $u_x = -U_x \ , \ p_z = -U_z \ , \ v_y = U_y$. Using (2.11) and (2.7), $\nabla U \geq \phi$ implies $u_x - p_z = U_z - U_x \geq \phi_{xz} = \alpha_{xz}$ and $v_y + p_z = U_y - U_z \geq \phi_{zy} = \gamma_{zy}$, thus, using complementary slackness one recovers

$$u_x = \max_z (\alpha_{xz} + p_z)^+ \text{ and } v_y = \max_z (\gamma_{zy} - p_z)^+ .$$

Further, if n and m have only integral entries, then there is an integral solution μ to (2.15). Therefore:

Theorem 2.1 (Queyranne)**.** *The hedonic equilibrium problem of Theorem 2.2 can be reformulated as a matching flow problem as described above.*

As announced above, this reformulation has several advantages. First, it establishes the existence of a hedonic equilibrium, and its integrality.

Theorem 2.2 (Existence)**.** *Consider a market given by n_x producers of type x, m_y consumers of type y, and where productivity of producer x is given by α_{xz}, and utility of consumer y is γ_{yz}. Then:*

(i) There exists a hedonic equilibrium $(p_z, \mu_{xz}, \mu_{yz})$;

(ii) (μ_{xz}, μ_{yz}) are solution to the primal problem of the expression of the social welfare

$$\max_{\mu_{xz},\mu_{yz}\geq 0} \sum_{xz} \mu_{xz}\alpha_{xz} + \sum_{yz} \mu_{yz}\gamma_{yz} \tag{2.17}$$

$$\sum_z \mu_{xz} \leq n_x \text{ and } \sum_z \mu_{yz} \leq m_y \text{ and } \sum_x \mu_{xz} = \sum_y \mu_{yz},$$

while (p_z) is obtained from the solution of the dual expression of the social welfare

$$\min_{u_x,v_y\geq 0;p_z} \sum_x n_x u_x + \sum_y m_y v_y \tag{2.18}$$

$$u_x \geq \alpha_{xz} + p_z \text{ and } v_y \geq \gamma_{yz} - p_z.$$

expressed equivalently as $\min_{p_z} \sum_x n_x G_x (\alpha_{x.} + p_.) + \sum_y m_y H_y (\gamma_{.y} - p_.)$, where the indirect surpluses G_x and H_y are defined in (2.3).

(iii) If n_x and m_y are integral for each x and y, then μ_{xz} and μ_{yz} can be taken integral.

Second, on the practical side, Theorem 2.2 also has a useful consequence in terms of computation of the equilibrium, as shown in the following corollary.

Corollary 2.1. *The equilibrium prices (p_z) as well as the quantities μ_{xz}, μ_{yz} supplied at equilibrium can be determined using one of the many maximum flows algorithms, see for instance Ahuja, Magnanti and Orlin (1993).*

Example 2.1. *Assume that there are four sellers and three buyers, each of whom is unique among her type, and three qualities. Participation is endogenous but there is no free disposal. Assume that the technology and preference parameters are given by*

$$(\alpha_{xz}) = \begin{pmatrix} 2 & 5 & 3 \\ 2 & 1 & 4 \\ 1 & 5 & 8 \\ 4 & 2 & 4 \end{pmatrix} \text{ and } (\gamma_{yz}) = \begin{pmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 1 & 2 & 6 \end{pmatrix}.$$

The indirect utilities of the buyers and the sellers are determined by linear programming. One finds $u_x^{\min} = (0\ 0\ 4\ 0)$ *and* $v_y^{\max} = (8\ 9\ 10)$*, and* $u_x^{\max} = (3\ 0\ 4\ 0)$*, and* $v_y^{\min} = (8\ 6\ 10)$*, and the optimal matching will consist in matching* x_1 *with* y_2*, which produce together quality 2, and any other two remaining producers with the two other remaining consumers, producing two units of quality of quality 3. Hence the optimal production of quality is* $l_{x_1} = 0$*,* $l_{x_2} = 1$ *and* $l_{x_3} = 2$*. Making use of* $p_z^{\min} = \max_y \left(\gamma_{yz} - v_y^{\max}\right)$ *and* $p_z^{\max} = \min_x \left(u_x^{\max} - \alpha_{xz}\right)$*,one finds that if* $u = (0\ 0\ 4\ 0)$ *and* $v = (8\ 9\ 10)$*, then* $p \in [-7, -4] \times [-5, -2] \times \{-4\}$.

3 Introducing heterogeneities

In the spirit of Galichon and Salanié (2014), who extended the model of Choo and Siow (2006), we are now going to introduce heterogeneities in producers' and consumers' characteristics. As before, we consider the set $\mathcal{X}$ of observable types of producers, the set $\mathcal{Y}$ of observable types of consumers, and $\mathcal{Z}$ be the set of qualities, and the sets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are finite[5]. In the sequel, i will denote an individual producer, and j will denote an individual consumer. The analyst observes the "observable type" $x_i \in \mathcal{X}$ of producer i, and the "observable type" $y_j \in \mathcal{Y}$ of consumer j. Two producers (resp. consumers) sharing the same observable type may differ in some additional heterogeneity term that will affect their profitability (resp. utility) function. This heterogeneity is observed by the consumers but not by the analyst. It is assumed that the quality $z \in \mathcal{Z}$ is fully observable by all parties and the analyst.

If the price of quality z is p_z, then the profit of an individual producer i selling quality z is defined as $\tilde{\alpha}_{iz} + p_z \in \mathbb{R} \cup \{-\infty\}$, and the utility of an individual consumer j purchasing z is defined as $\tilde{\gamma}_{jz} - p_z \in \mathbb{R} \cup \{-\infty\}$. If producer i (resp. consumer j) does not participate in the market, she gets a surplus of $\tilde{\alpha}_{i0}$ (resp. $\tilde{\gamma}_{j0}$). The tilde notation in $\tilde{\alpha}$ and $\tilde{\gamma}$ indicates that these terms characterize the invididual level, which will be random from the point of view of the observer. Note that the utility of agents on each side of the

[5]However, the ideas presented here extend to the continuous case, see Dupuy and Galichon (2014) for a continuous logit approach and Chernozhukov, Galichon and Henry (2014) for an approach based on multivariate quantile maps.

market still does not depend directly on the type of the agent with whom they match, but only indirectly via the type of the contract.

3.1 Structure of the heterogeneity

We introduce an structural assumption regarding the structure of unobserved heterogeneity.

Assumption 3.1. *Assume that the pre-transfer profitability and utility terms have structure*

$$\begin{aligned}\tilde{\alpha}_{iz} &= \alpha_{x_i z} + \varepsilon_{iz} \text{ and } \tilde{\gamma}_{jz} = \gamma_{y_j z} + \eta_{jz} \\ \tilde{\alpha}_{i0} &= \varepsilon_{i0} \text{ and } \tilde{\gamma}_{j0} = \eta_{j0}\end{aligned}$$

where:

a) The surplus shock, or unobserved heterogeneity component ε_i of all producers of observable characteristics x_i are drawn from the same distribution $\mathbf{P}_{x_i}$.

b) The surplus shock, or unobserved heterogeneity component η_j of all consumers of observable characteristics y_j are drawn from the same distribution $\mathbf{Q}_{y_j}$.

c) The distributions $\mathbf{P}$ and $\mathbf{Q}$ have full support.

Part a) and b) of this assumption are not very restrictive. They essentially express that the quality z is fully observed. Part c) is more restrictive. It implies that for each type of producer or consumer, and for any quality, some individual of this type will produce or consume this quality. This assumption does not hold if, say, some technological constraint prevents some producers to produce a given quality. Although this assumption is not required, and is not needed in Galichon and Salanié (2014), it greatly simplifies the results on identification and we will maintain it for the purposes of this paper.

We will also assume that:

Assumption 3.2. *There is a large number of producers and consumers of each given observable type, and each of them are price takers.*

This assumption has two virtues. First, it implies that we can have a statistical description of the producers and the consumer of a given type and we do not need to worry about sample variations. Second, it rules out any strategic behaviour by agents: the market here is assumed perfectly competitive.

3.2 Social welfare

We now investigate the social welfare, understood as the sum of the producers' and consumers' surpluses. We first focus on the side of producers. At equilibrium, producer i will get utility

$$U_{x_i z} + \varepsilon_{iz}$$

from producing quality z, where

$$U_{xz} = \alpha_{xz} + p_z.$$

The sum of the ex-ante indirect surpluses of the producers of observable type x is $n_x G_x(U_{x\cdot})$, where $G_x(U_{x\cdot})$ is the expected indirect utility of a consumer of type x, that is

$$G_x(U_{x\cdot}) = \mathbb{E}_{\mathbf{P}_x}\left[\max_{z\in\mathcal{Z}}(U_{xz} + \varepsilon_{iz}, \varepsilon_{i0})\right] \tag{3.1}$$

where $U_{x\cdot}$ denotes the vector of $(U_{xz})_{z\in\mathcal{Z}}$, and where the expectation is taken with respect to the distribution $\mathbf{P}_x$ of unobserved heterogeneity component ε_i. By the Envelope theorem, the number of producers of type x choosing quality z, denoted $\mu_{z|x}$, is given by

$$\begin{aligned} \mu_{z|x} = \frac{\mu_{xz}}{n_x} = & \quad \mathbf{P}_x\,(x \text{ chooses } z) \\ = & \quad \frac{\partial G_x(U_{x.})}{\partial U_{xz}}. \end{aligned} \tag{3.2}$$

This result sheds light on the *equilibrium characterization problem*: based on the vector of producer surpluses U, this allows to deduce the production patterns μ, and a similar picture holds on the consumers' side. However, the *identification problem* consists in recovering utility parameters, here $U_{x\cdot}$ based on the observation of producer' choices, here summarized by μ_{xz}, the number of producers of observable type x who choose to sell quality z. This requires inverting relation (3.2). To do this, still following Galichon and Salanié (2014), introduce the Legendre-Fenchel transform G_x^* of G_x as

$$\begin{aligned} G_x^*(\mu_{.|x}) &= \max_{U_{xz}} \left(\sum_{z \in \mathcal{Z}} \mu_{z|x} U_{xz} - G_x(U_{x.}) \right) \text{ if } \sum_{z \in \mathcal{Z}} \mu_{z|x} \leq 1 \qquad (3.3) \\ &= +\infty \text{ otherwise.} \end{aligned}$$

where $\mu_{.|x}$ is the vector of choice probabilities $\left(\mu_{z|x}\right)_{z \in \mathcal{Z}}$. By the Envelope theorem, one has

$$U_{xz} = \frac{\partial G_x^*(\mu_{z|x})}{\partial \mu_{z|x}}. \qquad (3.4)$$

Hence U_{xz} is identified from $\mu_{x.}$by equation (3.4). Galichon and Salanié (2014) have shown that G^* can be very efficiently computed as the solution to an optimal matching problem.

Similarly to the producers' side of the market, denote $V_{yz} = \gamma_{yz} - p_z$ the deterministic part of the consumer's payoff from buying good quality z, and write $V_{y.}$ for the $|\mathcal{Z}|$-dimensional vector with z-th component V_{yz}. The sum of expected utilities of consumers with observable characteristics y is given by $m_y H_y(V_{y.})$, where $H_y(V_{y.})$ is the expected indirect utility of a consumer of type y, that is

$$H_y(V_{y.}) = \mathbb{E}_{\mathbf{Q}_y} \left[\max_{z \in \mathcal{Z}} (V_{yz} + \eta_{yz}, \eta_{y0}) \right],$$

and $\mathbf{Q}_y$ is the distribution of the unobserved heterogeneity component η_j for a consumer indexed by j, with observable characteristics $y = y_j$. Hence, as in the producer's case, we obtain identification of V_{yz} through the following relation.

$$V_{yz} = \frac{\partial H_y^*(\mu_{.|y})}{\partial \mu_{z|y}}, \qquad (3.5)$$

where H_y^* is the convex conjugate of H_y, defined by a formula similar to (3.3).

Recall that the social welfare $\mathcal{W}$ is the sum of the producers and consumers surpluses. We are now able to state the following result.

Theorem 3.1. *(i) The optimal social welfare in this economy is given by*

$$\mathcal{W} = \min_{p_z} \sum_{x \in \mathcal{X}} n_x G_x \left(\alpha_{x.} + p_. \right) + \sum_{y \in \mathcal{Y}} m_y H_y \left(\gamma_{y.} - p_. \right). \tag{3.6}$$

(ii) Alternatively, $\mathcal{W}$ can be expressed as

$$\begin{aligned} \mathcal{W} &= \max_{\mu \geq 0} \sum_{x \in \mathcal{X}, z \in \mathcal{Z}} \mu_{xz} \alpha_{xz} + \sum_{y \in \mathcal{Y}, z \in \mathcal{Z}} \mu_{yz} \gamma_{yz} - \mathcal{E}(\mu) \qquad (3.7) \\ s.t. & \quad \mu \text{ satisfies (2.1) and (2.2),} \end{aligned}$$

where $\mathcal{E}(\mu)$ is a generalized entropy function, defined by

$$\mathcal{E}(\mu) = \sum_{x \in \mathcal{X}} n_x G_x^* \left(\mu_{x.} \right) + \sum_{y \in \mathcal{Y}} m_y H_y^* \left(\mu_{y.} \right).$$

(iii) Further the equilibrium $(p_z, \mu_{xz}, \mu_{yz})$ is unique and is such that (p_z) is a minimizer for (3.6) and (μ_{xz}, μ_{yz}) is a maximizer for (3.7).

The terminology "generalized entropy" comes from the fact, that in the Logit case where the utility shocks ε and η are i.i.d. and have a Gumbel distribution, then $\mathcal{E}(\mu)$ is a regular entropy function, namely

$$\mathcal{E}(\mu) = \sum_{x \in \mathcal{X},\ y \in \mathcal{Y}} \mu_{xy} \log \frac{\mu_{xy}^2}{n_x m_y} + \sum_{x \in \mathcal{X}} \mu_{xy} \log \frac{\mu_{x0}}{n_x} + \sum_{y \in \mathcal{Y}} \mu_{xy} \log \frac{\mu_{0y}}{m_y}$$

where $\mu_{x0} = n_x - \sum_{y \in \mathcal{Y}} \mu_{xy}$ and $\mu_{0y} = m_y - \sum_{x \in \mathcal{X}} \mu_{xy}$.

3.3 Identification

As a result of the first order conditions in the previous theorem, the model is exactly identified from the observation of the hedonic prices p_z, along with the production and consumption patterns μ_{xz} and μ_{yz}.

Theorem 3.2. *The producers and consumers systematic surpluses at equilibrium U and V are identified from μ_{xz} and μ_{yz} by*

$$U_{xz} = \frac{\partial G_x^*(\mu)}{\partial \mu_{x|z}} \text{ and } V_{yz} = \frac{\partial H_y^*(\mu)}{\partial \mu_{y|z}}.$$

Hence α and γ are identified from μ_{xz}, μ_{yz} and p_z by

$$\alpha_{xz} = \frac{\partial G_x^*(\mu)}{\partial \mu_{x|z}} - p_z \text{ and } \gamma_{yz} = \frac{\partial H_y^*(\mu)}{\partial \mu_{y|z}} + p_z.$$

4 Discussion

The results presented in this paper are applicable to many different empirical settings. Returning to the market for fine wines for example, the analyst will typically have access to data about the share of consumers with observable characteristics y purchasing wine of quality z and the share of producers of type x selling wine of quality z. Our methodology allows to identify the surpluses of consumers and producers from these data. If in addition, the price of wine of various qualities are observed, then the utility α of consumers and technology γ of producers are identified as well.

Next, consider the marriage market example. In classical models of sorting on the marriage market, following Becker (1973) and Shapley and Shubik (1972), the matching surplus between a man of type x and a woman of type y is

$$\Phi_{xy} = \alpha_{xy} + \gamma_{xy}$$

where α and γ are the man and the woman's surplus for being married to eachother. However, this analysis misses the fact that the partners in the marriage market also need to make a number of joint decisions, such as whether/when/how to raise children, where to live, how to spend their spare time together, etc. This has the flavour of a hedonic model. For the sake of discussion, consider (on the other extreme) a framework where the observed characteristics is, say, the date of birth of each agent, and where the only variable agents care about is, say, the date of birth of their first child. In this context, the matching surplus is now

$$\Phi_{xy} = \sup_{z}(\alpha_{xz} + \gamma_{yz})$$

and the methodology developed in this paper can identify the surplus of a man born in $x = 1985$ to have his first child in say $z = 2012$ and the surplus of a woman born in $y = 1986$ to have her first child in $z = 2013$. The required data are the shares of men and women born in a given year who had their first child in a given year. This example, however, is peculiar as men and women are likely to form preferences not only over the hedonic attribute z, i.e. the year of birth of first child, but also over their spouse's attributes x and y. One therefore needs to consider a model encompassing the hedonic model a la Rosen (1974) with the sorting model à la Becker (1973). In this model, developed and studied in Dupuy and Galichon and Zhao (2014) who apply it to the study of migration in China, the matching surplus is

$$\Phi_{xy} = \sup_z (\alpha_{xyz} + \gamma_{xyz})$$

and this model embeds both the classical sorting model ($\alpha_{xyz} = \alpha_{xy}$ and $\gamma_{xyz} = \gamma_{xy}$) and the hedonic model ($\alpha_{xyz} = \alpha_{xz}$ and $\gamma_{xyz} = \gamma_{yz}$). The empirically interesting question there is to assess which of the "sorting effect" or "hedonic effect" is strongest.

References

[1] Ahuja, R., T. Magnanti, and J. Orlin (1993): *Network Flows: Theory, Algorithms and Applications*, Prentice-Hall.

[2] Brown, J., and H. Rosen (1982): "On the estimation of structural hedonic price models," *Econometrica*, 50, 765–769.

[3] Chernozhukov, V., A. Galichon and M. Henry (2014): "Identifying multi-attribute hedonic models," working paper.

[4] Chiappori, P.-A., R. McCann, and L. Nesheim (2010): "Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness," *Economic Theory*, 42, 317–354.

[5] Chiong, K., A. Galichon, and M. Shum (2013): "Estimating dynamic discrete choice models via convex analysis," working paper.

[6] Choo, E., and A. Siow (2006): "Who marries whom and why?," *Journal of Political Economy*, 114, 175–201.

[7] Court, L. 1941: "Entrepreneurial and consumer demand theories for commodity spectra," *Econometrica*, 9, 135-162, 241-297.

[8] Dupuy, A. and A. Galichon (2014): "Personality traits and the marriage market," forthcoming in the *Journal of Political Economy.*

[9] Dupuy, A., A. Galichon and L. Zhao (2014): "Migration in China: to Work or to Wed?," working paper.

[10] Ekeland, I. (2010): "Existence, uniqueness and efficiency of equilibrium in hedonic markets with multidimensional types," *Economic Theory*, 42, 275–315.

[11] Ekeland, I., J. Heckman, and L. Nesheim (2004): "Identification and estimation of hedonic models," *Journal of Political Economy*, 112, 60–109.

[12] Epple, D. (1987): "Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products," *Journal of Political Economy*, 95, 59–80.

[13] Galichon, A., and B. Salanié (2014): "Cupid's invisible hand: social surplus and identification in matching models," working paper.

[14] Griliches, Z. 1961, "Hedonic price indexes for automobiles: an econometric analysis of quality change. The price statistics of the federal government," National Bureau of Economic Research, New York. p. 173-196.

[15] Golan, A. and Shalit, H. (1993) "Wine Quality Differentials in Hedonic Grape Pricing," *Journal of Agricultural Economics*, 44(2), pp. 311–321.

[16] Heckman, J., R. Matzkin, and L. Nesheim (2010): "Nonparametric identification and estimation of nonadditive hedonic models," *Econometrica*, 78(5), 1569–1591.

[17] Kain, J. F. and J. M. Quigley. 1970. "Measuring the value of house quality," *Journal of the American Statistical Association*, 65(330): 532-548.

[18] Lancaster, K. J. 1966. "A new approach to consumer theory," *Journal of Political Economy*, 74 (2): 132-157.

[19] Matzkin, R. (2003): "Nonparametric estimation of nonadditive random functions," *Econometrica*, 71, 1339–1375.

[20] Matzkin, R. (2013): "Nonparametric identification in structural econometric models," *Annual Review of Economics*, 5, 457–486.

[21] Nesheim, L. (2014): "Identification of multidimensional hedonic models," working paper.

[22] Queyranne, M. (2011). Personal communication.

[23] Rosen, S. (1974): "Hedonic prices and implicit markets: product differentiation in pure competition," *Journal of Political Economy*, 82, 34–55.

[24] Thaler, R. and S. Rosen (1976): "The Value of Saving a Life: Evidence from the Labor Market," *NBER Chapters, in: Household Production and Consumption*, pp. 265-302.

[25] Tinbergen, Jan (1956): "On the Theory of Income Distribution," *Weltwirtschaftliches Archiv*, 77(2), pp. 155–173.