# SET IDENTIFICATION IN MODELS WITH MULTIPLE EQUILIBRIA

ALFRED GALICHON† MARC HENRY§

Abstract. We propose a computationally feasible way of deriving the identified set of parameter values in models with multiple equilibria, with particular emphasis on oligopoly entry models. This is achieved through an equivalence result between the existence of an equilibrium selection mechanism compatible with the observed data and a set of inequalities, and through an appeal to efficient linear programming techniques.

**Keywords**: multiple equilibria, optimal transportation, identification regions, core determining classes.

**JEL subject classification**: C13, C72

---

*Date*: The present version is of October 29, 2008. This paper is a substantially revised version of the working paper Galichon and Henry (2006b), which was widely circulated since May 2006. Support from NSF grant SES 0350770 to Princeton University and from NSF grant SES 0532398 is gratefully acknowledged by both authors. Galichon s research is partly supported by Chaire X-Dauphine-EDF-Calyon"Finance et Développement Durable". We are grateful to Victor Chernozhukov, Pierre-André Chiappori, Ivar Ekeland, Guido Imbens and Bernard Salanié and Elie Tamer for encouragement, support and many helpful discussions and to Francesca Molinari and seminar participants at Berkeley, Chicago, Columbia, École polytechnique, Harvard-MIT, MIT Sloane OR, Northwestern, NYU, Princeton, SAMSI, Stanford, the Weierstrass Institut and Yale for helpful comments (with the usual disclaimer). Correspondence address: Département d'économie, École polytechnique, 91128 Palaiseau, France and Département de sciences économiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montral QC H3C 3J7, Canada. E-mail: alfred.galichon@polytechnique.edu and marc.henry@umontreal.ca.

## Introduction

The empirical study of imperfectly competitive markets based on game theoretic models is complicated by the presence of multiple equilibria. As noted in Jovanovic (1989), the existence of multiple equilibria generally leads to a failure of identification of the structural parameters governing the model. Berry and Tamer (2006) and Ackerberg, Benkard, Berry, and Pakes (2007) give an account of the various ways this identification issue was approached in the literature, where identification of structural parameters is achieved through equilibrium refinements, shape restrictions, informational assumptions or the specification of equilibrium selection mechanisms. An alternative approach is to eschew identification strategies and base inference purely on the identified features of the models with multiple equilibria, which are sets of values rather than a single value of the structural parameter vector. This approach is taken in the context of oligopoly entry models by Andrews, Berry, and Jia (2003) and Ciliberto and Tamer (2006). The inferential method they use, however, relies on a set of restrictions which is not guaranteed to exhaust all the restrictions embodied in the model, and hence leads to more conservative inference than could be achieved. This paper proposes a computationally feasible way of recovering the identified feature of a model with multiple equilibria, with a specific application to inference in oligopoly entry models.

We first note that the likelihood implied by a model with multiple equilibria can be represented by a non additive set function called a *Choquet capacity functional.* Seminal work on nonadditive likelihoods can be found in Manski (1990) and Heckman, Smith, and Clements (1997) among others. This nonadditive likelihood can be refined to a likelihood represented by a probability measure if there exists a mechanism that picks outcomes among the admissible equilibria in the region of multiplicity. We give a formal definition of an *equilibrium selection mechanism*, and call such a mechanism *compatible* with the data if the likelihood of the model augmented with such a mechanism is equal to the probabilities observed in the data. The identified feature of the model therefore is the set of parameter values such that there exists an equilibrium selection mechanism compatible with the data. The first main result of this paper is the equivalence between the latter condition and the actual probability of observed outcomes belonging to the core of the likelihood predicted by the model, where the *core* is a well known and studied notion in economics since the word was coined in Gillies (1953). This results allows the computation of the identified feature of models with multiple equilibria and a finite number of observable outcomes, as it reduces the problem to that of checking a finite number of moment inequalities. A related representation was developed independently by Beresteanu, Molchanov, and Molinari (2008) who emphasize the characterization of the identified set as an Aumann integral.

The computational burden remains high in situations with a large number of observable outcomes, since the number of inequalities to be checked is equal to the number of subsets of the set of observable outcomes. When the set of observable outcomes is infinite, the problem remains infinite dimensional. Galichon and

Henry (2006b) and Ekeland, Galichon, and Henry (2008) include results pertaining to that case. To lift the remaining computational burden, we propose two alternative strategies. First we propose to use efficient linear programming methods to determine whether the observed probabilities are in the core of the model likelihood. Second, we introduce the notion of *core determining classes*, which are suitably low cardinality classes of sets that are sufficient to characterize the core, and we provide some results to exhibit such core determining classes in practice. The method is illustrated on oligopoly entry examples proposed in Tamer (2003) and Berry and Tamer (2006).

Beyond the identification issue of computing the identified set given the knowledge of the true distribution of observable variables, the inference issue of constructing confidence regions for structural parameters in models with multiple equilibria is taken up in Galichon and Henry (2008) and Galichon and Henry (2006a) to complement the seminal contribution of Chernozhukov, Hong, and Tamer (2007). Related work on inference in partially identified models include Manski and Tamer (2002), Imbens and Manski (2004), Beresteanu and Molinari (2008), Romano and Shaikh (2008), Rosen (2008), Andrews and Soares (2007), Canay (2007), Fan (2008) among many others.

The remainder of the paper is organized as follows. Section 1.1 describes the framework and general results, while section 1.2 specializes and illustrates them on the example of oligopoly entry models. Section 2 introduces the notion of core determining classes and results necessary for their construction. Section 3 shows how to use efficient linear programming methods to compute the identified feature of a model with multiple equilibria. Section 4 illustrates the two methods on an oligopoly entry game with two types of players. The last section concludes, and proofs of the results are collected in an appendix.

## 1. Identified features of models with multiple equilibria

### 1.1. **Identified parameter sets in general models with multiple equilibria.**

The general framework is that of Jovanovic (1989). We consider three types of economic variables. Outcome variables $Y$, exogenous explanatory variables $X$, and latent variables, or random shocks, $\epsilon$. Outcome variables and latent variables are assumed to belong to complete and separable metric spaces, so that both outcomes and latent variables could be discrete, continuous, they could be probability distributions or stochastic processes. The economic model consists in a set of restrictions on the joint behaviour of the variables listed above. These restrictions may be induced by assumptions of rationality of agents, and they generally depend on a set of unknown structural parameters $\theta$. Without loss of generality, the model may be formalized as a measurable correspondence (defined in assumption 1 below) between the latent variables $\epsilon$ and the outcome variables $Y$ indexed by the exogenous variables $X$ and the vector of parameters $\theta$. We call this correspondence $G$, and write $Y \in G(\epsilon|X;\theta)$ to indicate admissible values of $Y$ given values of $\epsilon$, $X$ and $\theta$. The econometrician will be assumed to have access to a sample of independent and identically distributed vectors $(Y,X)$, and the problem considered is that of estimating the vector of parameters $\theta$. The latent variables $\epsilon$ is supposed to be distributed according to a parametric distribution $\nu_\theta$, where the indexing is meant to indicate that the unknown parameters that enter in the distribution of latent variables are contained in the vector $\theta$ of parameters to the estimated. We collect these assumptions next.

**Assumption 1.** *An independent and identically distributed sample of copies of the random vector $(Y, X)$ is available. The observable outcomes $Y$ conditionally distributed according to the probability distribution $P(\cdot|X)$ on $\mathcal{Y}$, a Polish space (i.e. a complete and separable metric space) endowed with its Borel $\sigma$-algebra of subsets $\mathcal{B}$) are related to unobservable variables $\epsilon$ according to the model $Y \in G(\epsilon|X;\theta)$, where $\theta$ belongs to an open subset $\Theta$ of $\mathbb{R}^{d_\theta}$, $\epsilon$ is distributed according to the probability measure $\nu_\theta$ on $\mathcal{U}$ (also Polish endowed with its Borel $\sigma$-algebra of subsets) independently of $X$, and $G$ is a measurable correspondence[1], i.e. such that for all open subsets $A$ of $\mathcal{Y}$, $G^{-1}(A|X;\theta) := \{\epsilon \in \mathcal{U} : \; G(\epsilon|X;\theta) \cap A \neq \varnothing\}$ is measurable (a measurable correspondence is also called* random correspondence *or* random set*, and the requirement is very mild) for almost all $X$ and for all $\theta \in \Theta$. Finally, the variables $(Y, X, \epsilon)$ are defined on the same underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

**Example 1.** *To illustrate assumption 1, we consider a simple game proposed by Jovanovic (1989). There are two firms with profit functions $\pi_1(Y_1, Y_2, \epsilon_1, \epsilon_2; \theta) = (\theta Y_2 - \epsilon_2)1_{\{Y_1=1\}}$ and $\pi_2(Y_1, Y_2, \epsilon_1, \epsilon_2; \theta) = (\theta Y_1 - \epsilon_1)1_{\{Y_2=1\}}$, where $Y_i \in \{0, 1\}$ is firm $i$'s action, and $\epsilon = (\epsilon_1, \epsilon_2)'$ are exogenous costs. The firms know their costs; the analyst, however, knows only that $\epsilon$ is uniformly distributed on $[0, 1]^2$, and that the structural parameter $\theta$ is in $(0, 1]$. There are two pure strategy Nash equilibria. The first is $Y_1 = Y_2 = 0$ for all $\epsilon \in [0, 1]^2$. The second is $Y_1 = Y_2 = 1$ for all $\epsilon \in [0, \theta]^2$ and $Y_1 = Y_2 = 0$ otherwise. Hence the model is described by the correspondence: $G(\epsilon; \theta) = \{(0, 0), (1, 1)\}$ for all $\epsilon \in [0, \theta]^2$, and $G(\epsilon; \theta) = \{(0, 0)\}$ otherwise.*

To conduct inference on the parameter vector $\theta$, one first needs to determine the identified features of the model. Because the correspondence $G$ may be multi-valued due to the presence of multiple equilibria, the outcomes may not be uniquely determined by the latent variable. In such cases, the likelihood of an outcome falling in the subset $A$ of $\mathcal{Y}$ predicted by the model is $\mathcal{L}(A|X;\theta) = \nu_\theta(G^{-1}(A|X;\theta)) = \mathbb{P}(G(\epsilon|X;\theta) \cap A \neq \varnothing|X)$ (We use both the notation used in our previous version, Galichon and Henry (2006b), and the notation used in Berry and Tamer (2006) to clarify the equivalence between concepts discussed in the former and the latter). Because of multiple equilibria, this likelihood may sum to more than one, as we may have $A_1 \cap A_2 = \varnothing$, and yet $G^{-1}(A|X;\theta) \cap G^{-1}(A|X;\theta) \neq \varnothing$, so that we may have $\mathcal{L}(A \cup B|X;\theta) < \mathcal{L}(A|X;\theta) + \mathcal{L}(B|X;\theta)$. The set function $A \mapsto \nu_\theta(G^{-1}(A|X;\theta))$ is generally not additive, and is called a *Choquet capacity functional* (see Choquet (1954)).

[1]In the previous version circulated, Galichon and Henry (2006b), we used the notation $U$ for $\epsilon$ and $\Gamma$ for $G^{-1}$.

**Definition 1** (Choquet capacity functional). *Let $G$ be a measurable correspondence from $\mathcal{U}$ to $\mathcal{Y}$, and let $\nu$ be a countably additive probability measure on $\mathcal{U}$. The set function $A \mapsto \nu(G^{-1}(A))$ (with $G^{-1}$ defined in assumption 1) for all $A$ measurable set in $\mathcal{Y}$ is called a Choquet capacity functional on $\mathcal{Y}$.*

**Example 1 continued** In example 1, $\nu_\theta$ is the uniform distribution on $[0,1]^2$ and the Choquet capacity functional $\nu_\theta G^{-1}$ gives value $\nu_\theta G^{-1}(\{(0,0)\}) = \nu_\theta([0,1]^2) = 1$ to the set $\{(0,0)\}$ and value $\nu_\theta G^{-1}(\{(1,1)\}) = \nu_\theta([0,\theta]^2) = \theta^2$ to the set $\{(1,1)\}$. Hence it is immediately apparent that the Choquet capacity functional $\nu_\theta G^{-1}$ is a set function that is not additive, as it sums to more than 1.

As discussed in Berry and Tamer (2006) and Ciliberto and Tamer (2006), the model with multiple equilibria can be completed with an equilibrium selection mechanism. Following Jovanovic (1989) and Berry and Tamer (2006) (See for instance the formulation (2.20) page 66 of Berry and Tamer (2006)), we define an equilibrium selection mechanism as a conditional distribution $\pi_{Y|\epsilon,X}$ over equilibrium outcomes $Y$ in the regions of multiplicity. By construction, an equilibrium selection is allowed to depend on the latent variables $\epsilon$ even after conditioning on $X$. This is summarized in the following definition.

**Definition 2** (Equilibrium selection mechanism). *An equilibrium selection mechanism is a conditional probability $\pi(Y|\epsilon,X)$ for $Y$ conditionally on $\epsilon$, such that the selected value of the outcome variable is actually an equilibrium, or more formally, such that $\pi(.|\epsilon,X)$ has support $G(\epsilon|X;\theta)$.*

As explained in Berry and Tamer (2006) and Ciliberto and Tamer (2006), "the identified features of the model is the set of parameters for which there exists a selection mechanism such that the probabilities of outcomes predicted by the model are equal to the probabilities obtained from the data."

**Definition 3** (Compatible equilibrium selection mechanism). *The equilibrium selection mechanism $\pi(.|\epsilon,X)$ is compatible with the data if the probabilities observed in the data are equal to the probabilities predicted by the equilibrium selection mechanism, or more formally (see for instance the formulation (3.24) page 72 of Berry and Tamer (2006)) if for all $A$ measurable subset of $\mathcal{Y}$, $P(A|X) = \int_{\mathcal{U}} \pi(A|\epsilon,X)\nu_\theta(d\epsilon)$.*

Hence the identified set is the set of parameters $\theta$ such that there exists an *equilibrium selection mechanism compatible with the data*.

**Definition 4** (Identified set). *We call* identified set *(sometimes called* sharp identified set*) the set $\Theta_I$ of $\theta \in \Theta$ such that there exists an equilibrium selection mechanism compatible with the data.*

The definition above is not an operational definition, in the sense that it does not alow the computation of the identified set based on the knowledge of the probabilities in the data because the conditional distribution $\pi$ is an infinite dimensional nuisance parameter. We now set out to show how to reduce the dimensionality of the problem with an appeal to mass transportation methods (see Rachev and Rüschendorf (1998) and

Villani (2003) for expositions of the theory). Our equivalent formulation of the identified set is based on an appeal to the notion of *core* of the Choquet capacity functional introduced in definition 1.

**Definition 5** (Core of a Choquet capacity functional)**.** *The* core *of a Choquet capacity functional $\rho$ on $\mathcal{Y}$ is the collection of probability distributions $Q$ on $\mathcal{Y}$ such that for all $A$ measurable subset of $\mathcal{Y}$, $Q(A) \leq \rho(A)$.*

**Example 1 continued** In example 1, the core of the Choquet capacity functional $\nu_\theta G^{-1}$ is the set of probabilities $P$ for the observed outcomes $(0,0)$ and $(1,1)$ such that $P(\{(0,0)\}) \leq \nu_\theta G^{-1}(\{(0,0)\}) = \nu_\theta([0,1]^2) = 1$ and $P(\{(1,1)\}) \leq \nu_\theta G^{-1}(\{(1,1)\}) = \nu_\theta([0,\theta]^2) = \theta^2$.

The result we propose next[2] shows the equivalence between the existence of a compatible equilibrium selection mechanism and the fact that the true distribution of the data belongs to the core of the Choquet capacity functional that characterizes the likelihood predicted by the model (which we shall call *core of the likelihood predicted by the model*).

**Theorem 1.** *The identified set $\Theta_I$ is equal to the set of parameters such that the true distribution of the observable variables lies in the core of the likelihood predicted by the model. Hence*

$$\begin{aligned}\Theta_I &= \{\theta \in \Theta : \forall A \in \mathcal{B},\, P(A|X) \leq \nu_\theta(G^{-1}(A|X;\theta)),\ X-\text{a.s.}\} \\ &= \{\theta \in \Theta : \forall A \in \mathcal{B},\, \mathbb{P}(Y \in A|X) \leq \mathbb{P}(G(\epsilon|X;\theta) \cap A \neq \varnothing|X),\ X-\text{a.s.}\}.\end{aligned}$$

**Example 1 continued** In example 1, the identified set is the set of values for $\theta$ such that $p \leq \theta^2$ and $1-p \leq 1$ where $p = \mathbb{P}((Y_1,Y_2)=(1,1))$ is the true probability that the observable variable takes the value $(1,1)$, i.e. that both firms enter the market. Hence, $\Theta_I = [\sqrt{p}, 1]$.

The first thing to note from this theorem is that the problem of computing the identified set has been transformed into a finite dimensional problem in the special case where $\mathcal{Y}$ is a finite set (or equivalently, the support of the distribution $P$ of observable outcomes has finite cardinality). Indeed, in the latter case, the problem of computing the identified set is reduced to the problem of computing a finite number of inequalities, i.e. $\mathbb{P}(Y \in A|X) \leq \mathbb{P}(G(\epsilon|X;\theta) \cap A \neq \varnothing)$ for each subset $A$ of $\mathcal{Y}$. However, in cases where the cardinality of $\mathcal{Y}$ is large, then the number of inequalities to be checked is $2^{\text{Card}(\mathcal{Y})}$, and the computational burden is only partially lifted, and the second section of the paper is devoted to the analysis of *core determining classes*, which are classes of test sets that allow a reduction in the number of inequalities to be checked in the computation of the identified set. First we turn to the specialization of our results and concepts to the case of oligopoly entry models, and illustrate them with a duopoly entry game extensively studied in the literature.

[2]This result appeared as equivalence between (ii') and (iv') in theorem 1' of the previous version circulated Galichon and Henry (2006b).

1.2. **Models of market entry.** A leading special example of the framework above is that of empirical models of oligopoly entry, proposed in Bresnahan and Reiss (1990) and Berry (1992), and considered in the framework of partial identification by Tamer (2003), Andrews, Berry, and Jia (2003), Berry and Tamer (2006), Ciliberto and Tamer (2006) and Pakes, Porter, Ho, and Ishii (2004) among others. In this setup, economic agents are firms who decide whether of not to enter a market. Markets are indexed by $m$, $m = 1, \ldots, M$ and firms that could potentially enter the market are indexed by $i$, $i = 1, \ldots, J$. $Y_{im}$ is firm $i$'s strategy in market $m$, and it is equal to 1 if firm $i$ enters market $m$, and zero otherwise. $Y_m$ denotes the vector $(Y_{1m}, \ldots, Y_{Jm})^t$ of strategies of all the firms. In standard notation, $Y_{-im}$ denotes the vector of strategies of all firms except firm $i$. In models of oligopoly entry, the profit $\pi_{im}$ of firm $i$ in market $m$ is allowed to depend on strategies $Y_{-im}$ of other firms, as well as on a set of profit shifters $X_{im}$ that are observed by all firms and the econometrician, a profit shifter $\epsilon_{im}$ that is observed by all the firms but not by the econometrician, and a vector of unknown structural parameters, so that it can be written $\pi_{im}(Y_m, X_{im}, \epsilon_{im}; \theta) = \pi_{im}(Y_{im}, Y_{-im}, X_{im}, \epsilon_{im}; \theta)$. If, for instance, firms are assumed to play Nash equilibria in pure strategies[3] in market $m$, their strategies $Y_{im}$ are such that they yield higher profits than $1 - Y_{im}$ given other firms' strategies $Y_{-im}$. So the restrictions induced on the strategies and latent profit shifters are $\pi_{im}(Y_{im}, Y_{-im}, X_{im}, \epsilon_{im}; \theta) \geq \pi_{im}(1 - Y_{im}, Y_{-im}, X_{im}, \epsilon_{im}; \theta)$ for all $i = 1, \ldots, J$. Hence the model can be written $Y_m \in G(\epsilon_m | X_m; \theta)$, where $X_m$ denotes the matrix of observed profit shifters for firms $i = 1, \ldots, J$, $\epsilon_m$ denotes the vector of latent profit shifters for firms $i = 1, \ldots, J$, and the correspondence $G$ is defined by $G(\epsilon | X; \theta) = \{Y : \ \pi_i(Y_i, Y_{-i}, X_i, \epsilon_i; \theta) \geq \pi_i(1 - Y_i, Y_{-i}, X_i, \epsilon_i; \theta); \text{all } i = 1, \ldots, J\}$, where the index $m$ is dropped when considering a generic market.

**Pilot 1.** *For illustration purposes, we describe the special case of this framework extensively studied in Tamer (2003), Berry and Tamer (2006) and Ciliberto and Tamer (2006). Two firms are present in the industry, so that $J = 2$, and a firm decides to enter the market $m$ if it makes a non negative profit in a pure strategy Nash equilibrium. Profit functions are supposed to have the following linear form $\pi_{im} = \alpha_i X_{im} + \delta_{-i} Y_{-im} + \epsilon_{im}$, so that $Y_{im} = 1$ if $\alpha_i X_{im} + \delta_{-i} Y_{-im} + \epsilon_{im} \geq 0$ and zero otherwise. As noted in Tamer (2003), if monopoly profits are larger than duopoly profits, i.e. $\delta_i < 0$, for $i = 1, 2$, and if $-\alpha_i X_{im} \leq \epsilon_{im} \leq -\alpha_i X_{im} - \delta_{-i}$, $i = 1, 2$, then there are multiple equilibria, since the model predicts either $Y_{1m} = 1$ and $Y_{2m} = 0$ or $Y_{1m} = 0$ and $Y_{2m} = 1$. The correspondence $G(\epsilon_m | X_m; \theta)$, where $\theta$ contains $\alpha_i, \delta_i$, $i = 1, 2$ is represented in figure 1. Note that $G$ is multi-valued in the rectangle with lower left corner $(-\alpha_1 X_{1m}, -\alpha_2 X_{2m})$ and upper-right corner $(-\alpha_1 X_{1m} - \delta_2, -\alpha_2 X_{2m} - \delta_1)$.*

The model thus described is incomplete in the sense that more information is required in the regions of multiplicity to determine which equilibrium will be selected. Without knowledge of such an equilibrium

[3] As noted by Francesca Molinari, equilibria in mixed strategies can be handled identically. However, we concentrate here on pure strategies for illustration purposes.

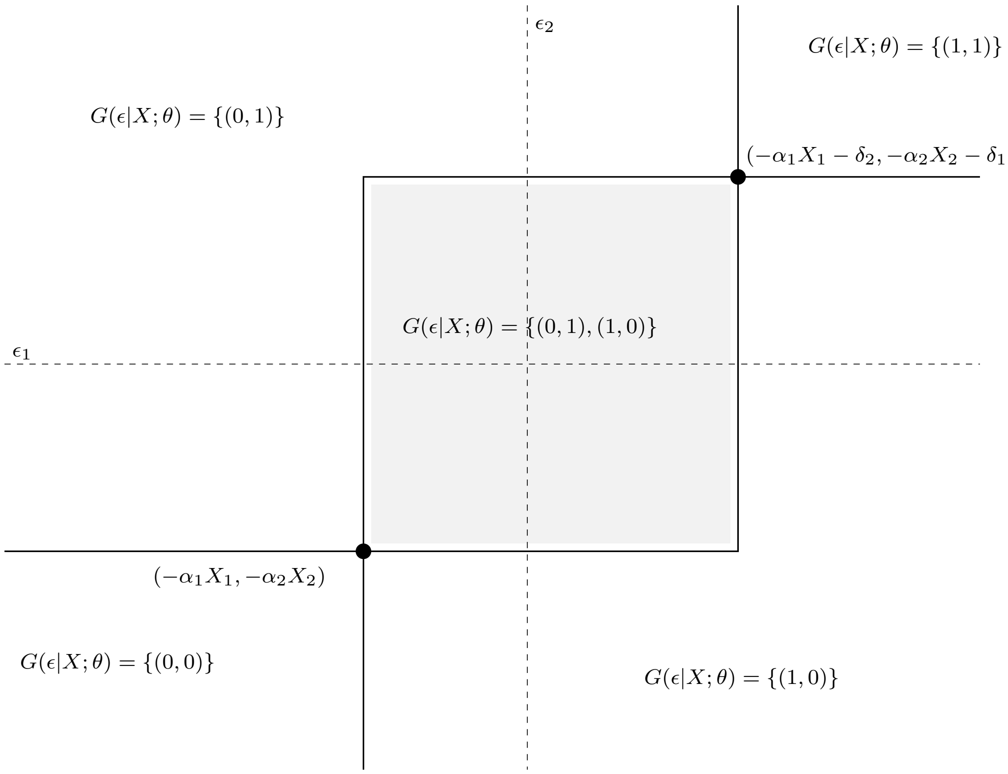


FIGURE 1. Representation of the model correspondence $G(\epsilon|X;\theta)$ as a function of $\epsilon$ in the $2\times 2$ entry game with duopoly profits lower than monopoly profits. The dotted lines represents the axes in the $\epsilon$ space, and the full lines represent the frontiers of the regions defining the correspondence $G$. The shaded area is the area of multiplicity, where $G(\epsilon|X;\theta)$ contains two values $(0,1)$ and $(1,0)$.

selection mechanism, the likelihood predicted by the model can be written as follows. Call $\mathcal{Y}$ the set of possible outcomes in a generic market. The likelihood of observation $y$ is $\mathcal{L}(y|X;\theta) = \mathbb{P}(y \in G(\epsilon|X;\theta)|X) = \nu_\theta(G^{-1}(y|X;\theta))$, for all $y \in \mathcal{Y}$ and $\sum_{y\in\mathcal{Y}} \mathcal{L}(y|X;\theta) \geq 1$, where the inequality may be strict if there are regions of multiplicity.

**Pilot example 1 continued** In the case of the duopoly entry game, the set of possible outcomes is $\mathcal{Y} = \{(0,0),\ (0,1),\ (1,0),\ (1,1)\}$. The likelihood of each individual outcome predicted by the model can be written as follows.

$$\mathcal{L}((0,0)|X;\theta) = \mathbb{P}(\epsilon_1 \leq -\alpha_1 X_1,\ \epsilon_2 \leq -\alpha_2 X_2|X) = \nu_\theta(G^{-1}((0,0)|X;\theta))$$
$$\mathcal{L}((0,1)|X;\theta) = \mathbb{P}(\epsilon_1 \leq -\alpha_1 X_1 - \delta_2,\ \epsilon_2 \geq -\alpha_2 X_2|X) = \nu_\theta(G^{-1}((0,1)|X;\theta))$$
$$\mathcal{L}((1,0)|X;\theta) = \mathbb{P}(\epsilon_1 \geq -\alpha_1 X_1,\ \epsilon_2 \leq -\alpha_2 X_2 - \delta_1|X) = \nu_\theta(G^{-1}((1,0)|X;\theta))$$
$$\mathcal{L}((1,1)|X;\theta) = \mathbb{P}(\epsilon_1 \geq -\alpha_1 X_1 - \delta_2,\ \epsilon_2 \geq -\alpha_2 X_2 - \delta_1|X) = \nu_\theta(G^{-1}((1,1)|X;\theta))$$

The likelihood predicted by the model is the set function $A \mapsto \nu_\theta G^{-1}(A|X;\theta)$ for $A$ subset of $\mathcal{Y} = \{(0,1),$ $(0,1)$, $(1,0)$, $(1,1)\}$. This set function is the Choquet capacity functional associated with the correspondence $G(\epsilon|X;\theta)$ and the distribution $\nu_\theta$ of $\epsilon$. If the support of $\nu_\theta$ is sufficiently large, the likelihood sums to more than one, because the region of multiple equilibria is "counted twice". This is related to the non-additive feature of Choquet capacity functionals, as seen here with the inequality $\nu_\theta(G^{-1}(\{(0,1)\} \cup \{(1,0)\}|X;\theta)) < \nu_\theta(G^{-1}(\{(0,1)\}|X;\theta)) + \nu_\theta(G^{-1}(\{(1,0)\}|X;\theta))$, since the latter is equal to the former plus $\mathbb{P}(-\alpha_1 X_1 \leq \epsilon_1 \leq -\alpha_1 X_1 - \delta_2, -\alpha_2 X_2 \leq \epsilon_2 \leq -\alpha_2 X_2 - \delta_1)$.

The model can be completed by adding an equilibrium selection mechanism which will pick out a single equilibrium for each value of the latent variable $\epsilon$ in the region of multiplicity. As formally defined in the previous section, an equilibrium selection mechanism is a conditional probability $\pi(.|\epsilon, X)$ supported on $G(\epsilon|X;\theta)$. It is compatible with the data if the probabilities it predicts are equal to the true probabilities of the observable variables.

**Pilot example 1 continued** As in (2.20) page 66 of Berry and Tamer (2006), in the duopoly example, we have for $i,j = 0,1$:

$$P((i,j)|X) = \int_{\mathcal{U}} \pi((i,j)|\epsilon, X)\nu_\theta(d\epsilon)$$

For purposes of illustration and additional interpretation, we now introduce an additional result that links the existence of a compatible equilibrium selection mechanism to measurable selections of the model correspondence $G$.

**Definition 6.** *A measurable selection of the measurable correspondence $G(\epsilon|X;\theta)$ is a measurable function $\gamma(\epsilon|X)$ such that $\gamma(\epsilon|X) \in G(\epsilon|X;\theta)$ for almost all $\epsilon$ and $X$, and for all $\theta$. The set of measurable selections is denoted* $\mathrm{Sel}(G(.|X;\theta))$.

**Pilot example 1 continued** A measurable selection $\gamma$ picks out either $(0,1)$ or $(1,0)$ for each $\epsilon$ such that $-\alpha_1 X_1 \leq \epsilon_1 \leq -\alpha_1 X_1 - \delta_2, -\alpha_2 X_2 \leq \epsilon_2 \leq -\alpha_2 X_2 - \delta_1$. An example of such a selection is represented in figure 2.

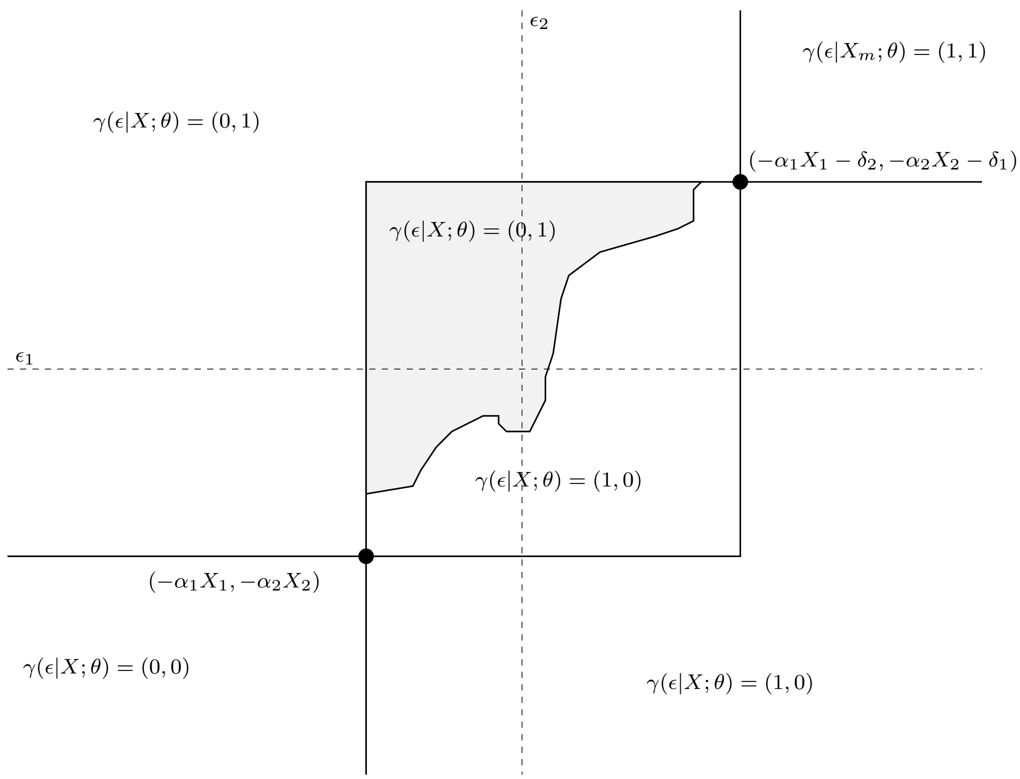


FIGURE 2. Representation of a selection $\gamma(\epsilon|X;\theta)$ from the model correspondence $G(\epsilon|X;\theta)$. Note that $\gamma$ is identical to $G$ except in the region of multiplicity.

The likelihood of the model given measurable selection $\gamma$ is $\nu_\theta\gamma^{-1}$. Since $\gamma$ is a measurable function, the probability image $\nu_\theta\gamma^{-1}$ of $\nu_\theta$ by $\gamma$ is a countably additive probability measure, unlike the Choquet capacity $\nu_\theta G^{-1}$. Heuristically, an equilibrium selection mechanism is a distribution over selections, so that the likelihood predicted by the model augmented with the equilibrium selection mechanism is a mixture of $\nu_\theta\gamma^{-1}$ where $\gamma$ ranges over the set $\mathrm{Sel}(G(.|X;\theta))$ of measurable selections of $G(.|X;\theta)$. The following proposition[4] shows that the heuristics above are correct.

**Proposition 1.** *The existence of a compatible equilibrium selection mechanism is equivalent to the fact that the observed probability distribution of the outcome variable $Y$ is a mixture of likelihoods induced by selections. Formally, $P(.|X)$ is in the closure (in the topology of convergence in distribution) of the convex hull of the set of images of $\nu_\theta$ by selections of the model correspondence, or* $P \in \mathrm{WCCH}\{\nu_\theta\gamma^{-1} : \gamma \in \mathrm{Sel}(G(.|X;\theta))\, X\text{a.s.}\}$.

[4]This result appeared as equivalence between (i') and (ii') in theorem 1' of the previous version circulated Galichon and Henry (2006b).

As was noted in Berry and Tamer (2006) and Ciliberto and Tamer (2006), since the model contains no prior information about which outcome is selected in the regions of multiplicity, the identified set $\Theta_I$ for the parameter vector $\theta$ is "the set of parameters for which there exists a proper selection function such that the choice probabilities predicted by the model are equal to the choice probabilities obtained from the data". The definition of the identification region using a semiparametric likelihood representation, where the selection mechanism is included as the infinite dimensional nuisance parameter $\pi$ is impractical, so we use theorem 1 to provide an operational method to compute $\Theta_I$. The existence of a compatible selection mechanism is equivalent to the fact that the true distribution $P$ of observed outcomes lies in the core of the Choquet capacity functional $\mathcal{L} = \nu_\theta G^{-1}$ defined by the model. Hence, we have

$$\begin{aligned}\Theta_I &= \{\theta \in \Theta : \ \forall A \in 2^{\mathcal{Y}};\ P(A|X) \leq \nu_\theta(G^{-1}(A|X;\theta));\ X \text{ a.s.}\} \\ &= \{\theta \in \Theta : \ \forall A \in 2^{\mathcal{Y}};\ \mathbb{P}(Y \in A|X) \leq \mathbb{P}(G(\epsilon|X;\theta) \cap A \neq \varnothing|X);\ X \text{ a.s.}\}\end{aligned}$$

where $2^S$ denotes the set of all subsets of a set $S$, and where the last equality stems from the definition of the pre-image of the correspondence $G$, and the fact that $\nu_\theta$ is the distribution of the latent variable $\epsilon$.

**Pilot example 1 continued** In the case of the duopoly entry game, the identified region is the set of parameter vectors that satisfy the 16 inequalities $P(A|X) \leq \nu_\theta(G^{-1}(A|X;\theta))$, or in a different notation $\mathbb{P}(Y \in A|X) \leq \mathbb{P}(G(\epsilon|X;\theta) \cap A \neq \varnothing|X)$ for all sets $A$ in $2^{\mathcal{Y}}$, $X$-almost surely. An illustration of this procedure is given in figure 3.

## 2. Core determining classes, or which inequalities to check

As we have seen in the first section, theorem 1 allows to reduce the problem of computing the identified set to that of checking a set of inequalities. However, the computational burden is only partially lifted, as the number of inequalities to check can be very large if the cardinality of the outcome space is large. In this section, we shall analyze ways of reducing this remaining computational burden, by eliminating redundant inequalities in the computation of the identified set. This is formalized with the concept of *core determining classes*, which was first introduced in section 3.2.2 page 27 of Galichon and Henry (2006b).

**Definition 7.** *A class $\mathcal{A}$ of measurable subsets of $\mathcal{Y}$ is called* core determining *for the Choquet capacity functional $\rho$ on $\mathcal{Y}$ if it is sufficient to characterize the core of $\rho$, i.e. if a probability $Q$ is in* core($\rho$) *when $Q(A) \leq \rho(A)$ for all $A \in \mathcal{A}$. In other words, $Q(A) \leq \rho(A)$ for all $A \in \mathcal{A}$ implies $Q(A) \leq \rho(A)$ for every measurable set $A$.*

A core determining class $\mathcal{A}$ allows the elimination of all the inequalities $Q(A) \leq \rho(A)$ for $A \notin \mathcal{A}$ when checking whether a probability $Q$ belongs to the core of a Choquet capacity functional $\rho$. Since the likelihood predicted by the model $Y \in G(\epsilon|X;\theta)$ was characterized by the Choquet capacity functional

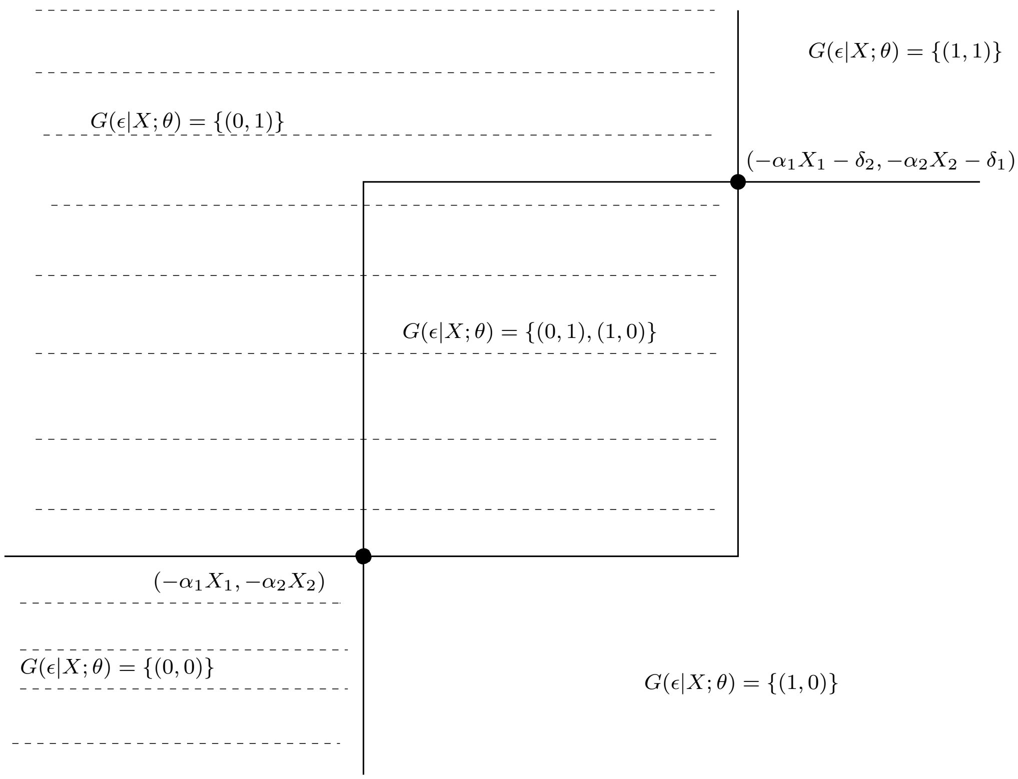


FIGURE 3. Representation of one of the inequalities to be checked. The probability of the outcome being either $(0,0)$ or $(0,1)$ needs to be no larger than the probability of the latent variable lying in the set covered with horizontal dashed lines.

$A\mapsto\mathbb{P}(G(\epsilon|X;\theta)\cap A\neq\varnothing|X)$, a core determining class of sets is sufficient to characterize the identified region $\Theta_I$ as summarized in the following proposition.

**Proposition 2.** *If $\mathcal{A}(\theta)$ is core determining for the Choquet capacity functional $A\mapsto\mathbb{P}(G(\epsilon|X;\theta)\cap A\neq\varnothing|X)$, then $\Theta_I=\{\theta\in\Theta:\forall A\in\mathcal{A}(\theta),\,\mathbb{P}(Y\in A|X)\leq\mathbb{P}(G(\epsilon|X;\theta)\cap A\neq\varnothing|X),\ X-\text{a.s.}\}$.*

The challenge therefore becomes that of finding a core determining class $\mathcal{A}$ in order to reduce the number of inequalities to be checked to the cardinality of $\mathcal{A}$. We first consider the case of our pilot example, before turning to a criterion that will prove useful in exhibiting core determining classes in many important cases.

**Pilot example 1 continued** We return to the duopoly entry game and consider some proposals for the computation of the identified set proposed in the literature. We call *ABJ class* the class of singleton sets $(\{(0,0)\},\{(0,1)\},\{(1,0)\},\{(1,1)\})$, since it corresponds to the class of sets proposed in Andrews, Berry, and Jia (2003) specialized to this simple case. It is immediate to see that the *ABJ* class is not core determining in general. Indeed, if $\epsilon$ has large enough support, the two equalities $\mathbb{P}(Y\in\{(0,1)\}|X)=\mathbb{P}(G(\epsilon|X;\theta)\cap\{(0,1)\}\neq$

$\varnothing|X) = \mathbb{P}(\epsilon_1 \leq -\alpha_1 X_1 - \delta_2,\ \epsilon_2 \geq -\alpha X_2|X)$ and $\mathbb{P}(Y \in \{(1,0)\}|X) = \mathbb{P}(G(\epsilon|X;\theta) \cap \{(1,0)\} \neq \varnothing|X) = \mathbb{P}(\epsilon_1 \geq -\alpha_1 X_1,\ \epsilon_2 \leq -\alpha X_2 - \delta_1|X)$ jointly imply $\mathbb{P}(Y \in \{(0,1),(1,0)\}|X) > \mathbb{P}(G(\epsilon|X;\theta) \cap \{(0,1),(1,0)\} \neq \varnothing|X) = \mathbb{P}([\epsilon_1 \geq -\alpha_1 X_1 \,\text{or}\, \epsilon_2 \leq -\alpha X_2 - \delta_1]\,\text{and}\,[\epsilon_1 \leq -\alpha X_1 - \delta_2 \,\text{or}\, \epsilon_2 \leq -\alpha X_2 - \delta_1]|X)$.

We now state a general criterion for the core determining property.

**Proposition 3.** *A class $\mathcal{A}(\theta)$ of subsets of $\mathcal{Y}$ is core determining for the Choquet capacity $A \mapsto \mathbb{P}(G(\epsilon|X;\theta) \cap A \neq \varnothing|X)$ if for every measurable subset $A$ of $\mathcal{Y}$, there exists nonnegative integers $K, L, N$, $\alpha_k$, $k = 1, \ldots, K$, and elements $A_1, \ldots, A_K$ of $\mathcal{A}(\theta)$ such that for almost all $y \in \mathcal{Y}$ and almost all $\epsilon \in \mathcal{U}$,*

$$1_A(y) \leq \frac{1}{N}\left(\sum_{k=1}^{K} \alpha_k 1_{A_k}(y) - L\right) \quad \text{and} \quad 1_{\{G(\epsilon|X;\theta)\cap A \neq \varnothing\}}(\epsilon) \geq \frac{1}{N}\left(\sum_{k=1}^{K} \alpha_k 1_{\{G(\epsilon|X;\theta)\cap A_k \neq \varnothing\}}(\epsilon) - L\right). \tag{2.1}$$

We illustrate an immediate application of this proposition in our pilot example.

**Pilot example 1 continued** In the duopoly entry game, we can use proposition (2.1) directly to show that the class of sets $\mathcal{A} = \{\{(0,0)\},\{(0,1)\},\{(1,0)\},\{(1,1)\},\{(0,1),(1,0)\}\}$ is a core determining class. Note that this class has cardinality 5, which is very low compared to $2^4 = 16$. Take any subset $A$ of $\mathcal{Y}$. Write $A = A_1 \cup A_2$, where $A_1$ is a subset of $\{(0,1),(1,0)\}$, hence an element of $\mathcal{A}$, and $A_2$ is a subset of $\{(0,0),(1,1)\}$. Then we have $1_A = 1_{A_1} + 1_{A_2} = 1_{A_1} + 1_{\{(0,0)\in A_2\}} * 1_{\{(0,0)\}} + 1_{\{(1,1)\in A_2\}} * 1_{\{(1,1)\}}$ and $1_{\{\epsilon:\, G(\epsilon|X;\theta)\cap A \neq \varnothing\}} = 1_{\{\epsilon:\, G(\epsilon|X;\theta)\cap A_1 \neq \varnothing\}} + 1_{\{\epsilon:\, G(\epsilon|X;\theta)\cap A_2 \neq \varnothing\}} = 1_{\{\epsilon:\, G(\epsilon|X;\theta)\cap A_1 \neq \varnothing\}} + 1_{\{(0,0)\in A_2\}} * 1_{\{\epsilon:\, G(\epsilon|X;\theta)\cap \{(0,0)\} \neq \varnothing\}} + 1_{\{(1,1)\in A_2\}} * 1_{\{\epsilon:\, G(\epsilon|X;\theta)\cap \{(1,1)\} \neq \varnothing\}}$. Then it follows from proposition 3 that $\mathcal{A}$ is a core determining class.

We now show how to use proposition 3 to identify core determining classes more generally to avoid painstaking case-by-case elimination of redundant inequalities. The main tool in the computation of identified sets in entry games is the following corollary of proposition 3. It gives general conditions under which one can find a core determining class of low cardinality. Recall that a subset $A$ of an ordered set (with ordering $\preceq$) is said to be *connected* if any $a$ such that $\inf A \preceq a \preceq \sup A$ belongs to $A$.

**Assumption 2** (Monotonicity)**.** *There exists an ordering $\precsim_{\mathcal{Y}}$ of the set of outcomes $\mathcal{Y}$ and an ordering $\precsim_{\mathcal{U}}$ of the set of latent variables $\mathcal{U}$ such that $G(\epsilon|X;\theta)$ is connected for all $\epsilon \in \mathcal{U}$, $X$-a.s., all $\theta$, and $\sup G(\epsilon_2|X;\theta) \succsim_{\mathcal{Y}} \sup G(\epsilon_1|X;\theta)$ and $\inf G(\epsilon_2|X;\theta) \succsim_{\mathcal{Y}} \inf G(\epsilon_1|X;\theta)$ when $\epsilon_1 \precsim_{\mathcal{U}} \epsilon_2$. Both ordering are allowed to depend on the exogenous variables $X$, but the dependence is suppressed in the notation for clarity.*

**Remark 1.** *This assumption is related to monotone comparative statics in supermodular games (see Topkis (1998), Vives (1990) and Milgrom and Roberts (1990)). Testing monotone comparative statics is considered in Echenique and Komunjer (2008).*

We illustrate this assumption on our pilot duopoly entry game before stating the corollary and applying it to the more interesting case of an oligopoly entry game with two types of players presented in Berry and Tamer (2006).

**Pilot example 1 continued** In the duopoly entry game, the orderings are very simple to construct. A lexicographic order on $\mathcal{Y}$ is suitable, with $(0,0) \precsim_{\mathcal{Y}} (0,1) \precsim_{\mathcal{Y}} (1,0) \precsim_{\mathcal{Y}} (1,1)$. On $\mathcal{U}$ the ordering is related to predicted outcomes. All $\epsilon$ producing the same predicted outcomes will be equivalent, and the ordering on predicted outcomes is $\{(0,0)\} \precsim_{\mathcal{U}} \{(0,1)\} \precsim_{\mathcal{U}} \{(0,1),(1,0)\} \precsim_{\mathcal{U}} \{1,0\} \precsim_{\mathcal{U}} \{1,1\}$, where $A_1 \precsim_{\mathcal{U}} A_2$ is a short-hand notation for $\epsilon_1 \precsim_{\mathcal{U}} \epsilon_2$ if $G(\epsilon_1|X;\theta) = A_1$ and $G(\epsilon_2|X;\theta) = A_2$. It is straightforward to check assumption 2. For instance, taking $\epsilon_1$ such that $G(\epsilon_1|X;\theta) = \{(0,0)\}$ and $\epsilon_2$ such that $G(\epsilon_1|X;\theta) = \{(0,1),(1,0)\}$ we have $\sup G(\epsilon_1|X;\theta) = (0,1) \precsim_{\mathcal{Y}} (1,0) = \sup G(\epsilon_1|X;\theta)$ and $\inf G(\epsilon_1|X;\theta) = (0,1) = \inf G(\epsilon_1|X;\theta)$. This is illustrated in figure 4.

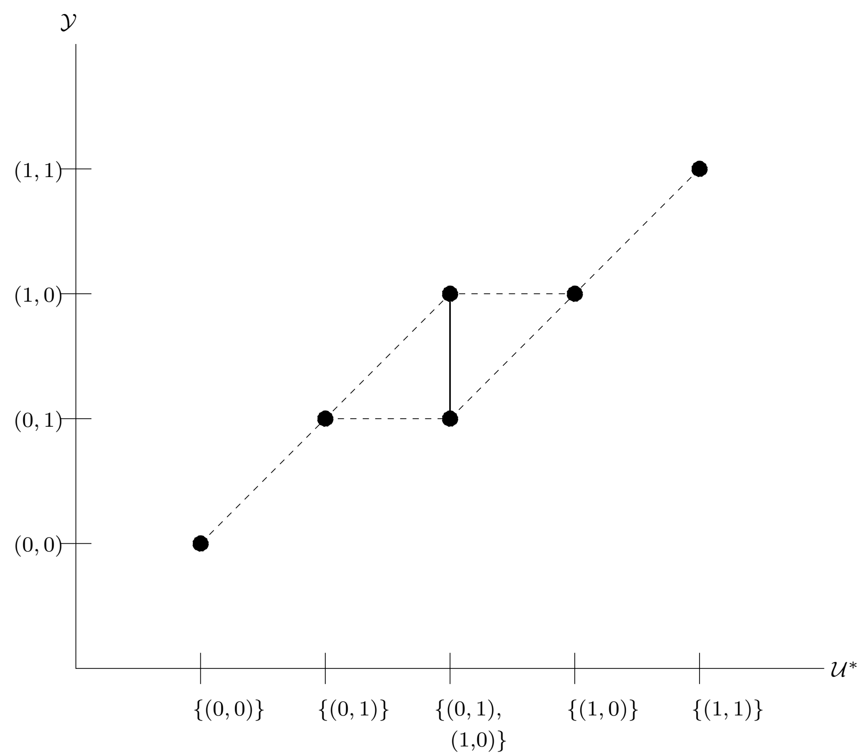


FIGURE 4. The monotonicity requirement in assumption 2 is satisfied for this choice of orderings in the duopoly entry example. (The thick dots represent the correspondence $G(.|X;\theta)$). $\mathcal{U}^*$ denotes the ordered set of combinations of equilibria.

We are now in a position to state the corollary[5], which is the main tool in the construction of core determining classes, and hence in the computation of the identified set.

**Corollary 1.** *Suppose assumption 2 is satisfied with orderings* $\precsim_{\mathcal{Y}}$ *and* $\precsim_{\mathcal{U}}$*. Call* $I$ *the cardinality of* $\mathcal{Y}$*, and list outcomes (elements of* $\mathcal{Y}$*) in increasing order (with respect to ordering* $\precsim_{\mathcal{Y}}$*) as* $y_1, \ldots, y_I$*. Then* $\mathcal{A}(\theta) = (\{y_1, \ldots, y_i\}, \{y_i, \ldots, y_I\} : \ i = 1 \ldots, I)$ *is core determining.*

Corollary allows to reduce the cardinality of the power set $2^{\mathcal{Y}}$ to twice the cardinality of $\mathcal{Y}$ minus 2 (since the inequality needn't be checked on the whole set $\mathcal{Y}$), as we illustrate in our pilot example.

**Pilot example 1 continued** In the duopoly example, assumption 2 is satisfied, as seen on figure 4, with the ordering described above. Hence the class $(\{\{(0,0)\}, \{(0,0),(0,1)\}, \{(0,0),(0,1),(1,0)\}, \{(0,1),(1,0),(1,1)\}, \{(1,0),(1,1)\}, \ \{(1,1)\}\})$ is core determining.

## 3. Alternative characterization of the core: linear programming approach

As the construction of sufficiently small core determining classes may be arduous in certain cases, we propose an alternative for the computation of the identified features of models with a finite set of observable outcomes, based on a linear programming approach. To set up the method, we need the following notations and definitions. For standard definitions in graph theory, we refer the reader to Papadimitriou and Steiglitz (1998).

Call $\mathcal{U}^*$ the set of predicted combinations of equilibria, formally $\mathcal{U}^* = \{G(\epsilon|X;\theta); \epsilon \in \mathcal{U}\}$. Hence $\mathcal{U}^*$ contains subsets of $\mathcal{Y}$, but is typically of much lower cardinality than the power set $2^{\mathcal{Y}}$. Further consider the bi-partite graph $\mathcal{G}(\theta, X)$ in $\mathcal{Y} \times \mathcal{U}^*$. The latter is defined as the set of pairs $(y,u) \in \mathcal{Y} \times \mathcal{U}^*$ such that $y \in u$. Each vertex $y$ in $\mathcal{Y}$ has weight $\mathbb{P}(Y = y|X)$ and each vertex $u \in \mathcal{U}^*$ has weight $\mathbb{P}(G(\epsilon|X;\theta) = u|X)$. The graph contains edges $(y,u)$ linking an element $y \in \mathcal{Y}$ to an element $u \in \mathcal{U}^*$ if the former is an element of the latter (i.e. $y \in u$). Finally, call $P(y|X) = \mathbb{P}(Y = y|X)$ the actual probabilities of observable variables $y \in \mathcal{Y}$, and call $Q(.|X;\theta)$ the probabilities $Q(u|X;\theta) = \mathbb{P}(G(\epsilon|X;\theta) = u|X)$. If we consider $G$ (keeping the same notation for simplicity) as a correspondence from $\mathcal{U}^*$ to $\mathcal{Y}$, then, formally $G(u) = u$, and we have shown in theorem 1 that $\theta$ belongs to the identified set if and only if for any subset $A$ of $\mathcal{Y}$, $P(A|X) \leq Q(G^{-1}(A)|X;\theta)$. Galichon and Henry (2008) show that it is equivalent to the existence of a joint probability $\pi$ on $\mathcal{Y} \times \mathcal{U}^*$ with marginal distributions $P(.|X)$ and $Q(.|X;\theta)$ and such that all the restrictions embodied in the model hold almost surely, or formally, such that $\pi\{(y,u) \in \mathcal{Y} \times \mathcal{U}^* : y \in u\} = 1$. This is summarized in the following proposition[6].

---

[5]This result is a reformulation of theorem 3d of the previous version circulated Galichon and Henry (2006b).

[6]This result is a special case of equivalence between (ii') and (iii') in theorem 1' of the previous version circulated Galichon and Henry (2006b).

**Proposition 4.** *The parameter value $\theta$ belongs to the identified set if and only if there exists a probability on $\mathcal{Y} \times \mathcal{U}^*$ with domain $\mathcal{G}(X;\theta)$ and with marginal probabilities $P(.|X)$ and $Q(.|X;\theta)$.*

Note that one implication in proposition 4 is very easy to prove. Call $U$ the random element with distribution $Q$. If a joint probability exists with the required properties, then $Y \in A \Rightarrow U \in G^{-1}(A)$, so that $1_{\{Y\in A\}} \leq 1_{\{U\in G^{-1}(A)\}}$, $\pi$-almost surely. Taking expectation, we have $\mathbb{E}_\pi(1_{\{Y\in A\}}) \leq \mathbb{E}_\pi(1_{\{U\in G^{-1}(A)\}})$, or equivalently $P(A|X) \leq Q(G^{-1}(A))|X;\theta)$. The converse is much more involved and relies on the optimal transportation theory.

We illustrate this requirement on our pilot example.

**Pilot example 1 continued** For the case of the duopoly entry model, $\mathcal{U}^* = \{\{(0,0)\}, \{(0,1)\}, \{(1,0)\}, \{(1,1)\}, \{(0,1),(1,0)\}\}$. The bi-partite graph is represented in figure 5, where $p_y$ denotes $\mathbb{P}(Y=y|X)$ and $q_u = \mathbb{P}(G(\epsilon|X;\theta)=u|X)$. The existence of a joint probability on $\mathcal{Y}\times\mathcal{U}^*$ supported on $\mathcal{G}(X;\theta)$ with marginal probabilities $p_y$, $y \in \mathcal{Y}$ and $p_u$, $u \in \mathcal{U}^*$ can be represented graphically by a set of non negative numbers attached to each edge of the graph, that sum to 1, and such that the weight of each vertex is equal to the sum of the weights on the edges that reach it. For instance, a joint probability is denoted $\alpha_1,\dots,\alpha_6$ and must satisfy $\alpha_i \geq 0$ for $i=1,\dots,6$, $\sum_{i=1}^6 \alpha_i = 1$ and equalities such as $p_{01} = \alpha_2 + \alpha_3$ and $q_{01,10} = \alpha_3 + \alpha_4$.

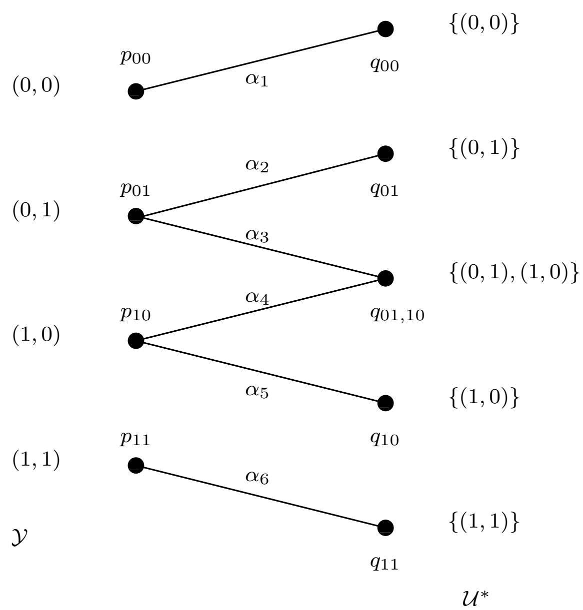


FIGURE 5. Bi-partite graph representing the admissible connections between observable outcomes and combinations of equilibria.

Since we have now formulated the problem of computing the identified set as a problem involving the existence of a probability measure with given marginal distributions, we can appeal to efficient computational methods in the optimal transportation literature. The problem of sending $p_y$, $y \in \mathcal{Y}$ *units* of a *good* from sources $y \in \mathcal{Y}$ to $p_u$, $u \in \mathcal{U}^*$ *units* to *terminals* at minimum cost of transportation, where costs are attached to each pair $(y, u) \in \mathcal{Y} \times \mathcal{U}^*$ is called the Monge-Kantorovich problem (a variant of Monge (1781) formulated by Hitchcock (1941), Kantorovich (1942) and Koopmans (1949)), and many efficient algorithms exist for this problem (see for instance page 143 of Papadimitriou and Steiglitz (1998)). Our problem can be reduced to a Monge-Kantorovich problem with 0-1 cost of transportation, where a pair $(y, u)$ is assigned cost zero if it belongs to $\mathcal{G}(X; \theta)$, and 1 otherwise, and there exists a joint law on $\mathcal{G}(X; \theta)$ with marginals $P$ and $Q$ (in other words, $\theta$ is in the identified set) if and only if there is a zero cost solution to the Monge-Kantorovich problem thus defined.

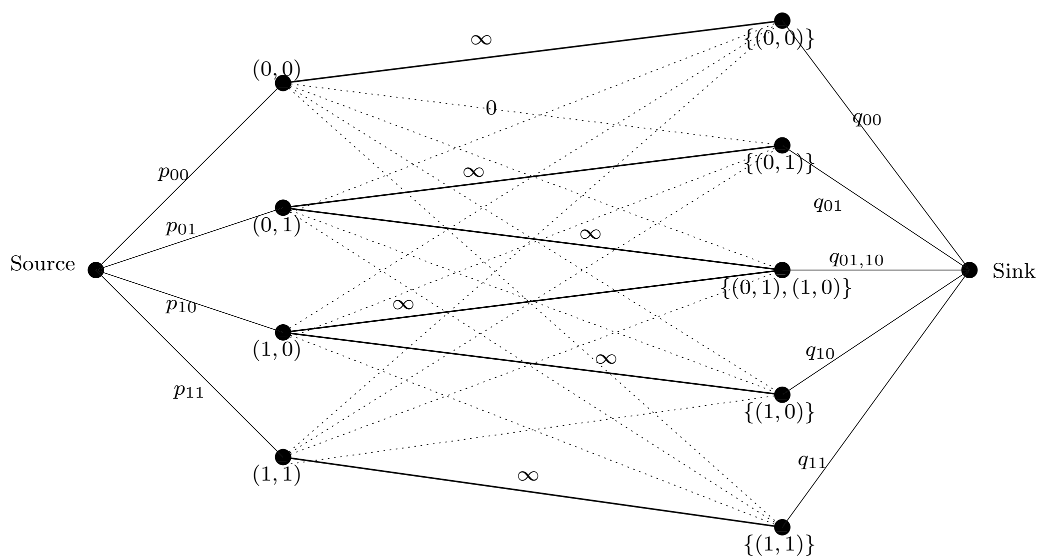


FIGURE 6. Maximum flow formulation of the Monge-Kantorovich problem for the duopoly example. Mass flows from the source to the sink through the network. The number on each edge is the maximum mass that can flow through that edge. The source sends $p_y$ mass exactly to each node corresponding to elements of $\mathcal{Y}$, and the sink receives $q_u$ mass from each node corresponding to an element of $\mathcal{U}^*$. Between edges in $\mathcal{Y}$ and $\mathcal{U}^*$, mass can flow freely through pairs $(y, u)$ such that $y \in u$ (full lines with infinite capacity), and not at all through pairs $(y, u)$ such that $y \notin u$ (dotted lines, with zero capacity). $\theta$ is in the identified set if and only if the maximum flow through this network is exactly 1.

As explained in Ford and Fulkerson (1957) (see also Papadimitriou and Steiglitz (1998) section 7.4 page 143), there is an equivalent dual formulation of this Monge-Kantorovich minimum cost of transportation problem as a maximum flow problem described in figure 6. The edges in the graph with zero cost in the minimum cost of transportation problem have infinite capacity (not to be confused with Choquet capacity functional) in the dual maximal flow problem. Hence efficient maximum flow programs (such as maxflow.m in the Matlab BGL library) can be applied directly to the network described in figure 6, and the parameter value $\theta$ is in the identified set if and only if the maximum flow program returns a maximum flow of exactly 1 (note that the network capacities depend on $\theta$ through the probabilities $q_u$). Classical algorithms exist for this problem, first and foremost, the Ford-Fulkerson algorithm (see Ford and Fulkerson (1957)). The algorithm implemented in the Matlab BGL library involves an order of $\text{Card}(\mathcal{U}^*)^3$ arithmetic operations (see page 217 of Papadimitriou and Steiglitz (1998)), which is the best known order of complexity for dense networks. A standard laptop computer requires only a couple of minutes to test $10^6$ values of the parameter vector in section 4.

## 4. Illustration: oligopoly entry with two types of players

We now turn to a more substantive illustration of our methods to compute the identified set, first, to show the operational usefulness of corollary 1, and second, to illustrate the power of the linear programming approach. To do so, we consider the oligopoly entry game with two types of players presented in appendix A of Berry and Tamer (2006). The profit function of type 1 firms depends on the total number of firms in the market, but not on the type of those firms, whereas profits of type 2 firms depend both on the number and on the type of firms present in the market. The latent variable is the fixed cost $f_1$ for firms of type 1 and $f_2$ for firms of type 2. The model is simplified by assuming linearity of profits in firm number as follows.

$$\pi_1(Y_1, Y_2, X, f; \theta) = \alpha_0 + \alpha_1(Y_1 + Y_2) + \alpha_2 X - f_1$$
$$\pi_2(Y_1, Y_2, X, f; \theta) = \beta_0 + \beta_1 Y_1 + \beta_2 Y_2 + \beta_3 X - f_2,$$

with $\alpha_1, \beta_1, \beta_2$ strictly negative and $\beta_2 > \beta_1$ to fix ideas (profit of type 2 firms will decrease by a larger amount if a type 1 firm enters the market than if a type 2 firm does). The set of observable outcomes is $\mathcal{Y} = \{(i, j) : \ i, j = 0, 1, 2\}$, where $i$ denotes the number of type 1 firms and $j$ the number of type 2 firms present in the market. $\mathcal{Y}$ can be ordered lexicographically, where the number of firms present in the market is considered first, and then the identity of firms (type 1 dominating type 2)[7]. Hence $(0,0) \precsim_{\mathcal{Y}} (0,1) \precsim_{\mathcal{Y}} (1,0) \precsim_{\mathcal{Y}} (0,2) \precsim_{\mathcal{Y}} (1,1) \precsim_{\mathcal{Y}} (2,0) \precsim_{\mathcal{Y}} (1,2) \precsim_{\mathcal{Y}} (2,1) \precsim_{\mathcal{Y}} (2,2)$. The model correspondence is represented in figure 7, which is taken from Berry and Tamer (2006).

### 4.1. Core determining class approach.

**Core determining class approach.** We first illustrate the usefulness of corollary 1 for the determination of a core determining class in this example. Figure 8 graphs the orderings that satisfy assumption 2 up to the fact that the set of equilibria is not always connected. Indeed, in the ordering of outcomes, $(1,1)$

[7] The order could be rationalized by total profit in the industry, but it is not necessary for the construction of a core determining class nor the computation of the identified set.

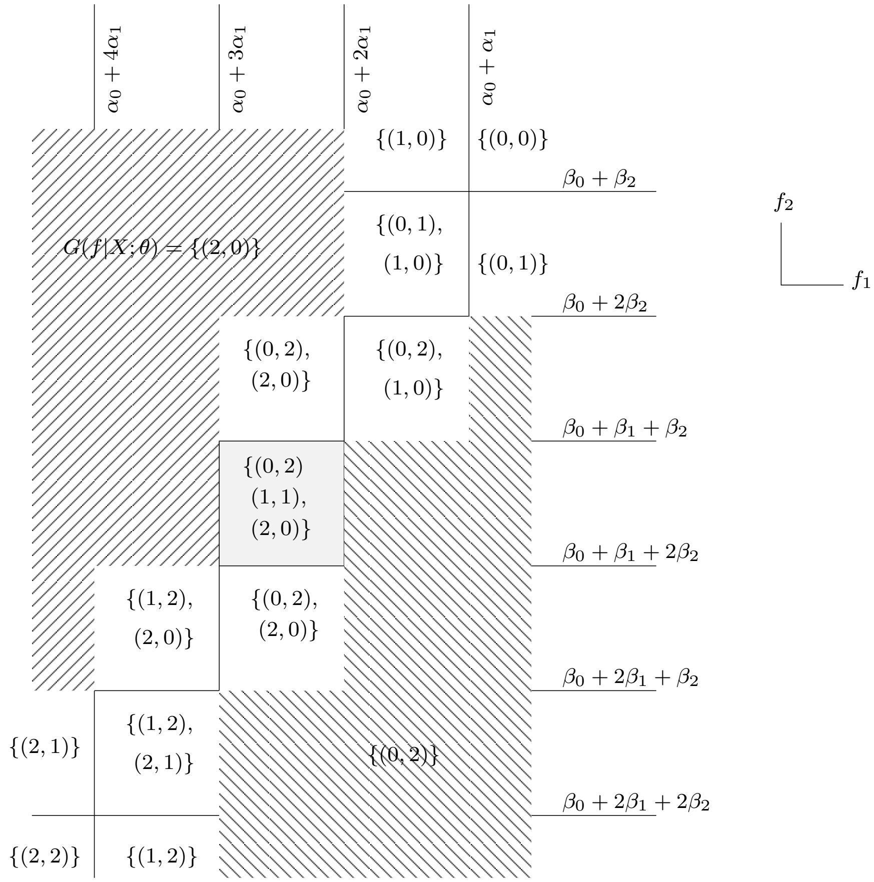


FIGURE 7. Model correspondence in the oligopoly entry game with two types of firms, two of each type.

comes between $(0,2)$ and $(2,0)$, or more precisely, $(0,2) \precsim_{\mathcal{Y}} (1,1) \precsim_{\mathcal{Y}} (2,0)$. Now $(1,1)$ is not an equilibrium when $\alpha_0 + 3\alpha_1 < f_1 \leq \alpha_0 + 2\alpha_1$ and $\beta_0 + \beta_1 + \beta_2 < f_2 \leq \beta_0 + 2\beta_2$, so the set of equilibria $\{(0,2),(2,0)\}$ is disconnected in that case. However, since $(1,1)$ is observed only when $\epsilon \in \{(0,2),(1,1),(2,0)\}$, the mass $p_{11}$ can be removed from $q_{02,11,20}$ and after re-normalization corollary 1 can be applied directly to $\mathcal{Y}\backslash\{(1,1)\}$ and $\mathcal{U}^*$, yielding the class $\mathcal{A}(\theta) = (\{(0,0)\}, \{(0,0), (0,1)\}, \{(0,0), (0,1), (1,0)\}, \{(0,0), (0,1), (1,0), (0,2)\}, \{(0,0), (0,1), (1,0), (0,2), (2,0)\}, \{(0,0), (0,1), (1,0), (0,2), (2,0), (1,2)\}, \{(0,0), (0,1), (1,0), (0,2), (2,0), (1,2), (2,1)\}, \{(0,1), (1,0), (0,2), (2,0), (1,2), (2,1), (2,2)\}, \{(1,0), (0,2), (2,0), (1,2), (2,1), (2,2)\}, \{(0,2), (2,0), (1,2), (2,1), (2,2)\}, \{(2,0), (1,2), (2,1), (2,2)\}, \{(1,2), (2,1), (2,2)\}, \{(2,1), (2,2)\}, \{(2,2)\})$. Note that its cardinality is $2 \times 7 = 14$, as opposed to the cardinality of the power set of $\mathcal{Y}$ which is $2^9 = 512$. Note in addition that this class is in fact independent of $\theta$, as long as the conditions stated in Ciliberto and Tamer (2006) that $0 > \beta_2 > \beta_1$ and $\alpha_1 < 0$ are satisfied.

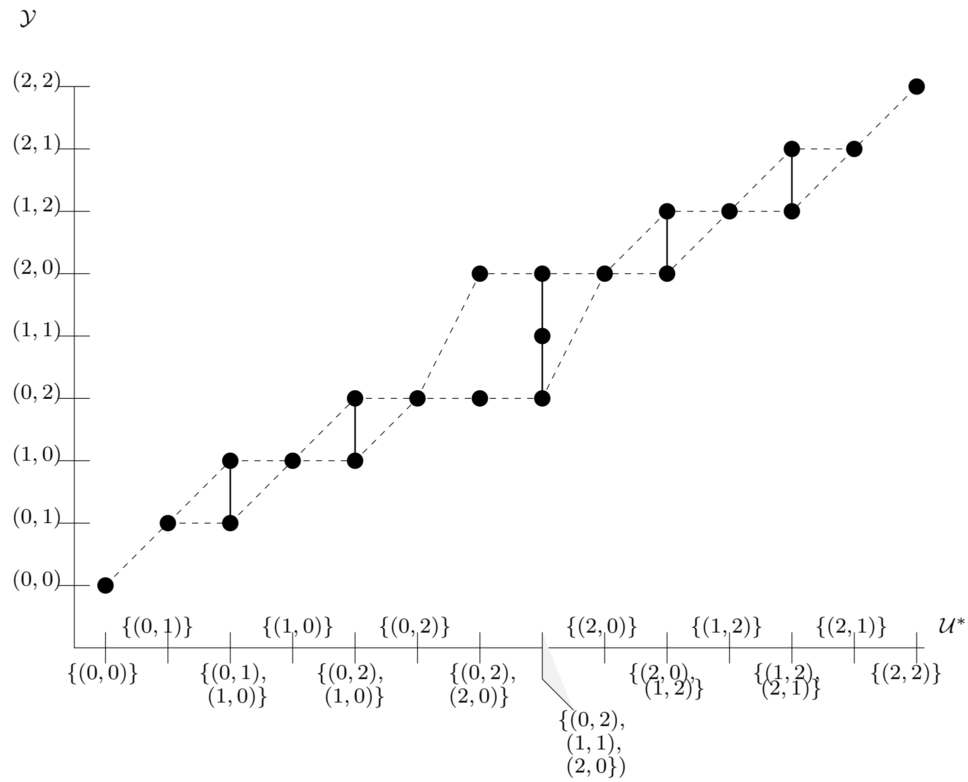


FIGURE 8. Assumption 2 is satisfied in the oligopoly entry game with two types of firms, two of each type, up to the fact that the image of $G(\{(0,2),(2,0)\}|X;\theta)$ is not connected.

4.2. **Linear programming approach.** Consider now the linear programming strategy for computing the identified set. The bipartite graph corresponding to this example is represented in figure 9. As shown in proposition 4, a value of the parameter vector is in the identified set if and only if there exists a zero cost transportation plan for the transfer of masses $p_y$ on the elements of $\mathcal{Y}$ to masses $q_u$ on the elements of $\mathcal{U}^*$. A transportation plan is a set of nonnegative numbers attached to all pairs $(y,u) \in \mathcal{Y} \times \mathcal{U}^*$ (which represents the amount of mass from $y$ that is transferred to $u$ via the edge $(y,u)$). In our application, the transportation cost from $y$ to $u$ is zero if $y$ and $u$ are connected by an edge in the graph of figure 9, and 1 otherwise. If the algorithm returns a zero cost transportation plan, it means that mass is transferred through edges of the graph only, and for instance the pair $((1,1),\{(0,2),(1,1),(2,0)\})$ is assigned a non negative number (i.e. some mass is transported there), but the pair $((1,1),\{(2,0),(0,2)\})$ is assigned zero (i.e. no mass is transported there). The existence of a zero cost transportation plan is equivalent to the existence of a joint distribution on $\mathcal{Y}\times\mathcal{U}^*$ which is concentrated on the graph of figure 9 and has the correct marginal distributions, hence, it is equivalent to the fact that $\theta$ is in the identified set, as we showed in proposition 4.

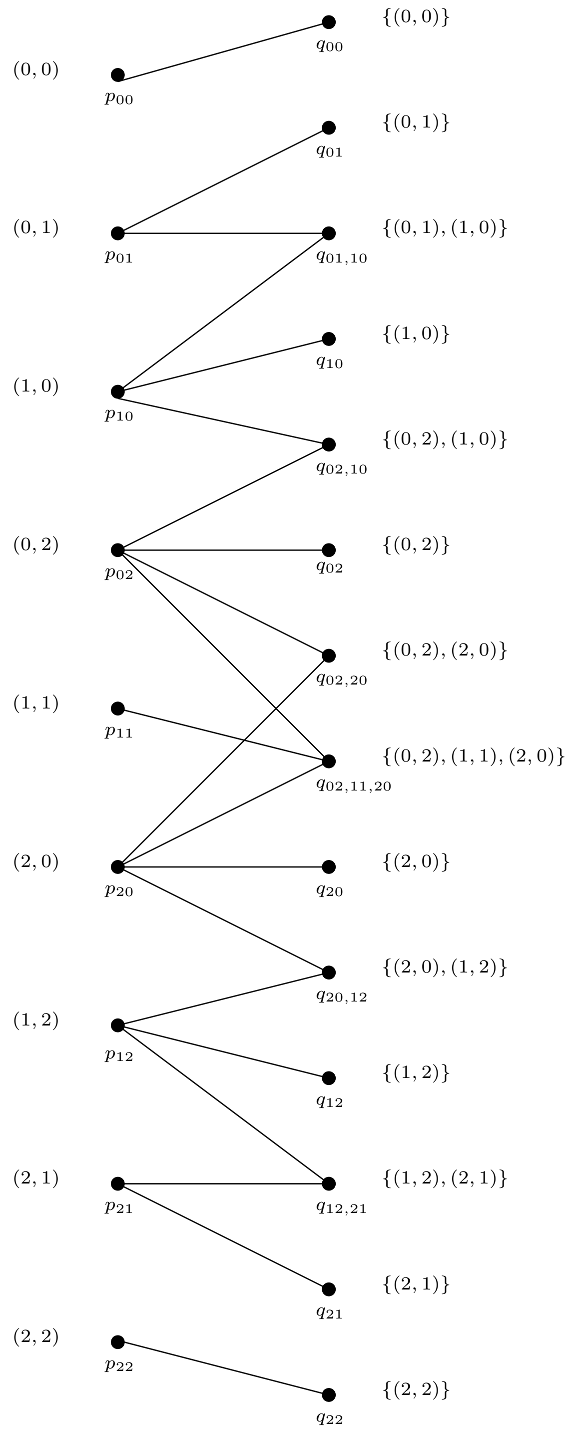


FIGURE 9. Bi-partite graph representing the admissible connections between observable outcomes and combinations of equilibria in the two-type oligopoly entry model.

Table 1. Adjacency matrix for the two-type oligopoly model.

| | (0,0) | (0,1) | (1,0) | (0,2) | (1,1) | (2,0) | (1,2) | (2,1) | (2,2) | *Sink* |
|---|---|---|---|---|---|---|---|---|---|---|
| *Source* | $p_{00}$ | $p_{01}$ | $p_{10}$ | $p_{02}$ | $p_{11}$ | $p_{20}$ | $p_{12}$ | $p_{21}$ | $p_{22}$ | |
| $\{(0,0)\}$ | $\infty$ | | | | | | | | | $q_{00}$ |
| $\{(0,1)\}$ | | $\infty$ | | | | | | | | $q_{01}$ |
| $\{(0,1),(1,0)\}$ | | $\infty$ | $\infty$ | | | | | | | $q_{01,10}$ |
| $\{(1,0)\}$ | | | $\infty$ | | | | | | | $q_{10}$ |
| $\{(1,0),(0,2)\}$ | | | $\infty$ | $\infty$ | | | | | | $q_{02,10}$ |
| $\{(0,2)\}$ | | | | $\infty$ | | | | | | $q_{02}$ |
| $\{(0,2),(2,0)\}$ | | | | $\infty$ | | $\infty$ | | | | $q_{02,20}$ |
| $\{(0,2),(1,1),(2,0)\}$ | | | | $\infty$ | $\infty$ | $\infty$ | | | | $q_{02,11,20}$ |
| $\{(2,0)\}$ | | | | | | $\infty$ | | | | $q_{20}$ |
| $\{(2,0),(1,2)\}$ | | | | | | $\infty$ | $\infty$ | | | $q_{20,12}$ |
| $\{1,2\}$ | | | | | | | $\infty$ | | | $q_{12}$ |
| $\{(1,2),(2,1)\}$ | | | | | | | $\infty$ | $\infty$ | | $q_{12,21}$ |
| $\{(2,1)\}$ | | | | | | | | $\infty$ | | $q_{21}$ |
| $\{(2,2)\}$ | | | | | | | | | $\infty$ | $q_{22}$ |

The minimum cost transportation problem is equivalent to the dual maximum flow problem, as described in the previous section. Mass flows through the network with 25 nodes, which include the *source*, the 9 elements of $\mathcal{Y}$, the 14 elements of $\mathcal{U}^*$ and the *sink* (mass flows in the direction Source $\rightarrow \mathcal{Y} \rightarrow \mathcal{U}^* \rightarrow$ Sink). A network is characterized by its *adjacency matrix*, which gives all the links between nodes with their capacity. In the case of interest here, the adjacency matrix is given in table 1. Maximum flow programs take this adjacency matrix as an input, and return the maximum flow through the network it characterizes. This maximum flow cannot be larger than $\sum_{y\in\mathcal{Y}} p_y = \sum_{u\in\mathcal{U}^*} p_u = 1$, and it is equal to 1 if and only if $\theta$ is in the identified set.

As an illustration of the procedure, we compute the identified set for the two-type oligopoly model with the following distributional hypotheses and normalization restrictions. The fixed cost vector $(f_1, f_2)$ is assumed to be uniformly distributed on $[0,1]^2$. $\alpha_0$ and $\beta_0$ are set equal to 1. As previously noted, we assume that monopoly profits are larger than oligopoly profits, and that a type-two firm's profit decreases more if a type one firm enters than a type two firm, hence $0 > \alpha_0$ and $0 > \beta_2 > \beta_1$. We can therefore calculate the probabilities of each combination of equilibria $u \in \mathcal{U}^*$. These probabilities are computed in a Matlab program file available on request. These implied probabilities are entered together with the true frequencies of observable outcomes into the adjacency matrix of table 1 and a maximum flow algorithm returns a flow of 1 if the value of $\theta = (\alpha_1, \beta_1, \beta_2)'$ (used to compute the predicted probabilities) belongs to the identified set, and a flow strictly smaller than 1 if it doesn't. To give an idea of the efficiency of the method, we can test $10^5$ values of $\theta = (\alpha_1, \beta_1, \beta_2)'$ in less than a second on a standard portable computer.

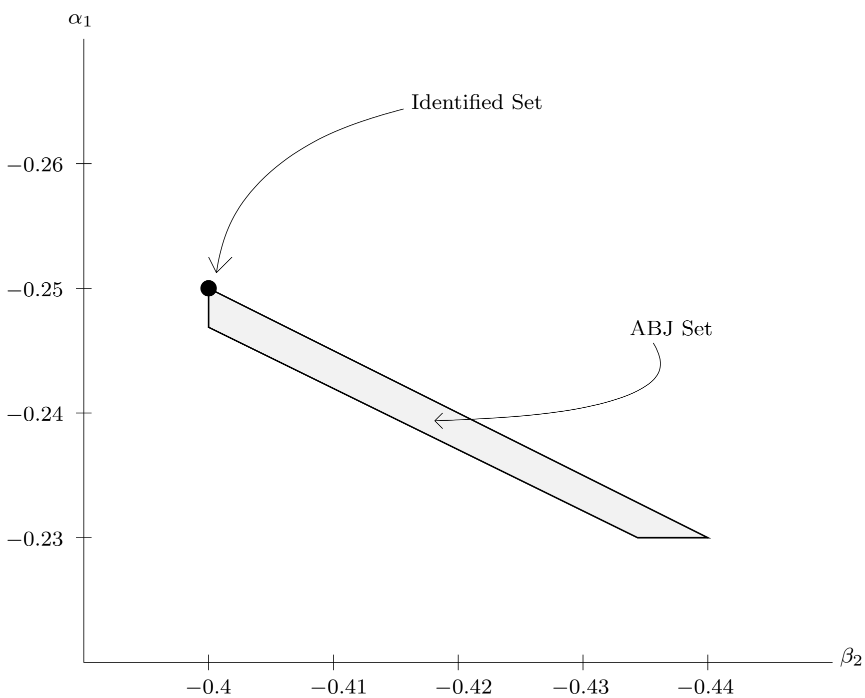


FIGURE 10. Projection of the identified set, and of the set characterized by the ABJ class on the $(\beta_2, \alpha_1)$ space. The projection of the identified set is a single point $(\beta_2, \alpha_1) = (-0.4, -0.25)$. The frequencies of observable variables are $(p_{00}, p_{01}, p_{10}, p_{02}, p_{11}, p_{20}, p_{12}, p_{21}, p_{22}) = (0.1, 0.15, 0.15, 0.1, 0, 0.5, 0, 0, 0)$.

For illustration purposes, we compute the identified set for a given choice of the observable frequencies, namely $(p_{00}, p_{01}, p_{10}, p_{02}, p_{11}, p_{20}, p_{12}, p_{21}, p_{22}) = (0.1, 0.15, 0.15, 0.1, 0, 0.5, 0, 0, 0)$, and compare it to the set obtained by imposing the inequality restrictions on the *ABJ class* only. The latter corresponds to the set of values of the parameters such that $p_{00} \leq q_{00}$, $p_{01} \leq q_{01} + q_{01,10}$, $p_{10} \leq q_{01,10} + q_{10} + q_{02,10}$, $p_{02} \leq q_{02,10} + q_{02} + q_{02,20} + q_{02,11,20}$, $p_{11} \leq q_{02,11,20}$, $p_{20} \leq q_{02,20} + q_{02,11,20} + q_{20} + q_{20,12}$, $p_{12} \leq q_{20,12} + q_{12} + q_{12,21}$, $p_{21} \leq q_{12,21} + q_{21}$ and $p_{22} \leq q_{22}$. It turns out the values of $\alpha_1$ and $\beta_2$ are identified (under these specific values for the true probabilities of observable variables, which were chosen for the simulation purposes from the parameter values), and all values of $\beta_1 < \beta_2$ are compatible with the given frequencies. The set defined by the *ABJ class* restrictions, however, is much larger, as shown by its projection on the $(\beta_2, \alpha_1)$ space in figure 10. There are also many values of the observed frequencies, for which the identified set is empty, so that the model is rejected, but the set defined by the *ABJ class* restrictions is non-empty, so that it fails to reject the model.

## Conclusion

In the context of models with multiple equilibria, we have proposed an equivalence result between the existence of an equilibrium selection mechanism compatible with the data and a set of inequalities characterizing the core of the model likelihood, and provided methods to reduce this number of inequalities to be checked with an appeal to the notion of core determining families and to efficient linear programming techniques. The issue of statistical inference on the identified feature thus characterized is taken up in Galichon and Henry (2008) and Galichon and Henry (2006a), which complement the seminal work of Chernozhukov, Hong, and Tamer (2007).

## Appendix A. Proofs of results in the main text

A.1. **Proof of theorem 1.** It suffices to show that for all $\theta \in \Theta$, statement 1 and statement 2 are equivalent, where statement 1 and statement 2 are defined as the following. Statement 1: $\mathbb{P}(Y \in A|X) \leq \mathbb{P}(G(\epsilon|X;\theta) \cap \varnothing|X)$ for all $A$ measurable subset of $\mathcal{Y}$, $X$-almost surely. Statement 2: For almost all $\epsilon$, there exists a probability measure $\pi(.|\epsilon, X)$ with support $G(\epsilon|X;\theta)$ such that $P(A|X) = \int_{\mathcal{U}} \pi(A|\epsilon, X)\nu_\theta(d\epsilon)$ for all $A$ measurable subset of $\mathcal{Y}$, $X$-almost surely. We proceed in six steps. *Step 1:* Since $G(\epsilon|X;\theta)$ is nonempty and closed by assumption, the set $\Delta(G(\epsilon|X;\theta))$ of probability measures on $\mathcal{Y}$ with support $G(\epsilon|X;\theta)$ is convex and closed in the topology of convergence in distribution. *Step 2:* Since $G(.|X;\theta)$ is a measurable correspondence, for any $f \in \mathcal{C}_b(\mathcal{Y})$, the set of all continuous and bounded real functions on $\mathcal{Y}$, the map $\epsilon \mapsto \sup\{\int f d\mu : \mu \in \Delta(G(\epsilon|X;\theta))\}$ is measurable. *Step 3:* By step 1 and step 2, we can apply theorem 3 of Strassen (1965) to conclude that statement 2 is equivalent to $\int_{\mathcal{Y}} f(y)P(dy) \leq \int_{\mathcal{U}} \sup\{\int f d\mu : \mu \in \Delta(G(\epsilon|X;\theta))\}\nu_\theta(d\epsilon)$ for all $f \in \mathcal{C}_b(\mathcal{Y})$. *Step 4:* For any bounded continuous function $f$, we have $\sup\{\int f d\mu : \mu \in \Delta(G(\epsilon|X;\theta))\} = \max\{f(y) : y \in G(\epsilon|X;\theta)\}$. *Step 5:* Call $\rho$ the Choquet capacity functional defined for all measurable subset $A$ of $\mathcal{Y}$ by $\rho(A) = \mathbb{P}(G(\epsilon|X;\theta) \cap A \neq \varnothing|X) = \nu_\theta(G^{-1}(A|X;\theta))$. We show that $\int_{\mathcal{U}} \max\{f(y) : y \in G(\epsilon|X;\theta)\}\nu_\theta(d\epsilon) = \int_{\text{Choquet}} f d\rho$, where the latter is the Choquet integral with respect to the Choquet capacity functional $\rho$, which is defined by $\int_{\text{Choquet}} f d\rho = \int_0^\infty \rho(\{f \geq v\})\,\mathrm{d}v + \int_{-\infty}^0 (\rho(\{f \geq v\}) - 1)\,\mathrm{d}v$. The latter can be rewritten $\int_0^\infty \mathbb{P}\big(G(\epsilon|X;\theta) \cap \{f \geq v\} \neq \varnothing|X\big)\,\mathrm{d}v + \int_{-\infty}^0 (\mathbb{P}\big(G(\epsilon|X;\theta) \cap \{f \geq v\} \neq \varnothing|X\big) - 1)\,\mathrm{d}v$, which is equal to $\int_0^\infty \mathbb{P}\big(\max_{y \in G(\epsilon|X;\theta)} f(y) \geq v|X\big)\,\mathrm{d}v + \int_{-\infty}^0 (\mathbb{P}\big(\max_{y \in G(\epsilon|X;\theta)} f(y) \geq v|X\big) - 1)\,\mathrm{d}v = \int_{\mathcal{U}} \max\{f(y) : y \in G(\epsilon|X;\theta)\}\nu_\theta(d\epsilon)$, as we set out to show. *Step 6:* Finally, by monotone continuity, we have that $\int_{\mathcal{Y}} f(y)P(dy|X) \leq \int_{\text{Choquet}} f d\rho$ for all $f \in \mathcal{C}_b(\mathcal{Y})$ is equivalent to the fact that $P(.|X)$ is in the core of the Choquet capacity functional $\rho$, which is statement 1.

A.2. **Proof of proposition 1.** By theorem 1, the existence of a compatible equilibrium selection mechanism is equivalent to the fact that $P$ is in the core of $\nu_\theta G^{-1}(.|X;\theta)$. By corollary 1 of Castaldo, Maccheroni, and Marinacci (2004), the core of the Choquet capacity functional $\nu_\theta G^{-1}(.|X;\theta)$ is equal to the closed convex hull of the set of images of measurable selections of $G$. Hence, the result follows.

A.3. **Proof of proposition 3.** Subtracting the second inequality in equation 2.1 yields $1_A - 1_{\{\epsilon:\, G(\epsilon|X;\theta)\cap A\neq\varnothing\}} \leq \frac{1}{N}\left(\sum_{k=1}^K \alpha_k(1_{A_k} - 1_{\{\epsilon:\, G(\epsilon|X;\theta)\cap A_k\neq\varnothing\}})\right)$. Taking expectations (conditionally on $X$) on both sides of the previous equation yields

$$\mathbb{P}(Y\in A|X) - \mathbb{P}(G(\epsilon|X;\theta)\cap A\neq\varnothing) \leq \frac{1}{N}\left(\sum_{k=1}^K \alpha_k(\mathbb{P}(Y\in A_k|X) - \mathbb{P}(G(\epsilon|X;\theta)\cap A_k\neq\varnothing))\right).$$

This in turn implies that $\mathbb{P}(Y\in A|X)\leq\mathbb{P}(G(\epsilon|X;\theta)\cap A\neq\varnothing)$ if $\mathbb{P}(Y\in A_k|X)\leq\mathbb{P}(G(\epsilon|X;\theta)\cap A_k\neq\varnothing)$ for each $k$, which means that $\mathcal{A}(\theta)$ is indeed core determining, which completes the proof.

A.4. **Proof of corollary 1.** We consider the equivalent problem where the set of latent variables $\mathcal{U}$ is replaced by the set of predicted combinations of equilibria $\mathcal{U}^*$. We keep the same notation for the equilibrium correspondence $G$, and call $u$ the elements of $\mathcal{U}^*$. For simplicity, we also drop the dependence on $X$ and $\theta$ in the notation, so that $G: u\mapsto G(u)$ is a correspondence between $\mathcal{U}^*$ and $\mathcal{Y}$. Note that by construction $G(u) = u\in 2^{\mathcal{Y}}$, but it is not the identity when considered as a correspondence. Let $A$ be a subset of $\mathcal{Y}$. Call $K_y$ the cardinality of $\mathcal{Y}$ and $K_u$ the cardinality of $\mathcal{U}^*$. List all elements of $\mathcal{Y}$ as $y_k$, $k=1\ldots,K_y$ and all elements of $\mathcal{U}^*$ as $u_k$, $k=1,\ldots,K_u$. For any $u\in\mathcal{U}^*$, define $k_u$ by $u=u_{k_u}$ and for any $y\in\mathcal{Y}$, define $k_y$ by $y=y_{k_y}$. As usual, denote $G^{-1}(A)$ the set $\{u\in\mathcal{U}^*: G(u)\cap A\neq\varnothing\}$. Call $\Delta 1_{G^{-1}(A)}(u_1) = 1_{G^{-1}(A)}(u_1)$ and for $k\geq 2$, $\Delta 1_{G^{-1}(A)}(u_k) = 1_{G^{-1}(A)}(u_k) - 1_{G^{-1}(A)}(u_{k-1})$. Call $\Delta 1^+_{G^{-1}(A)}$ and $\Delta 1^-_{G^{-1}(A)}$ the positive and negative parts of $\Delta 1_{G^{-1}(A)}$. By construction, we have for any $u\in\mathcal{U}^*$,

$$\begin{aligned}
1_{G^{-1}(A)}(u) &= \sum_{k=1}^{k_u}\Delta 1_{G^{-1}(A)}(u_k)\\
&= \sum_{k=1}^{K_u} 1_{\{u_k,\ldots,u_{K_u}\}}(u)\Delta 1_{G^{-1}(A)}(u_k)\\
&= \sum_{k=1}^{K_u} 1_{\{u_k,\ldots,u_{K_u}\}}(u)\Delta 1^+_{G^{-1}(A)}(u_k) - \sum_{k=1}^{K_u} 1_{\{u_k,\ldots,u_{K_u}\}}(u)\Delta 1^-_{G^{-1}(A)}(u_k)\\
&= \sum_{k=1}^{K_u} 1_{\{u_k,\ldots,u_{K_u}\}}(u)\Delta 1^+_{G^{-1}(A)}(u_k) + \sum_{k=1}^{K_u} 1_{\{u_1,\ldots,u_{k-1}\}}(u)\Delta 1^-_{G^{-1}(A)}(u_k) - \sum_{k=1}^{K_u}\Delta 1^-_{G^{-1}(A)}(u_k).
\end{aligned}$$

We then apply proposition 3 with the following choice of parameters. $K=2K_u$, $\alpha_j/N = \Delta 1^+_{G^{-1}(A)}(u_j)$, $j=1\ldots,K_u$, $\alpha_j/N = \Delta 1^-_{G^{-1}(A)}(u_{j-K_u})$, $j=K_u+1,\ldots,2K_u$, and $L/N = \sum_{k=1}^{K_u}\Delta 1^-_{G^{-1}(A)}(u_k)$. There remains to show that

$$1_A(y) \leq \sum_{k=1}^{K_u} 1_{\{\inf G(u_k),\ldots,y_{K_y}\}}(y)\Delta 1^+_{G^{-1}(A)}(u_k) + \sum_{k=1}^{K_u} 1_{\{y_1,\ldots,\sup G(u_{k-1})\}}(y)\Delta 1^-_{G^{-1}(A)}(u_k) - \sum_{k=1}^{K_u}\Delta 1^-_{G^{-1}(A)}(u_k)$$

to complete the proof. The latter follows from

$$\begin{aligned}
& \sum_{k=1}^{K_u} 1_{\{\inf G(u_k),\dots,y_{K_y}\}}(y)\Delta 1^{+}_{G^{-1}(A)}(u_k) + \sum_{k=1}^{K_u} 1_{\{y_1,\dots,\sup G(u_{k-1})\}}(y)\Delta 1^{-}_{G^{-1}(A)}(u_k) - \sum_{k=1}^{K_u} \Delta 1^{-}_{G^{-1}(A)}(u_k) \\
= & \sum_{k=1}^{K_u} 1_{\{\inf G(u_k),\dots,y_{K_y}\}}(y)\Delta 1^{+}_{G^{-1}(A)}(u_k) - \sum_{k=1}^{K_u} 1_{\{y_{k_{\sup G(u_{k-1})}+1},\dots,y_{K_u}\}}(y)\Delta 1^{-}_{G^{-1}(A)}(u_k). \qquad \text{(A.1)}
\end{aligned}$$

We have $k_{\inf G(u_k)} < k_{\sup G(u_{k-1})} + 1$ (otherwise, there would be a $y$ that belongs to none of the $u$'s, i.e. that is never an equilibrium outcome, and it could be eliminated from the analysis). Hence (A.1) is equal to

$$\begin{aligned}
& \sum_{k=1}^{K_u} 1_{\{\inf G(u_k),\dots,y_{k_{\sup G(u_{k-1})}+1}\}}(y)\Delta 1^{+}_{G^{-1}(A)}(u_k) \\
& \qquad + \sum_{k=1}^{K_u} 1_{\{y_{k_{\sup G(u_{k-1})}+1},\dots,y_{K_u}\}}(y)(\Delta 1^{+}_{G^{-1}(A)}(u_k) - \Delta 1^{-}_{G^{-1}(A)}(u_k)) \\
= & \sum_{k=1}^{K_u} 1_{\{\inf G(u_k),\dots,y_{k_{\sup G(u_{k-1})}+1}\}}(y)\Delta 1^{+}_{G^{-1}(A)}(u_k) + \sum_{k=1}^{K_u} 1_{\{y_{k_{\sup G(u_{k-1})}+1},\dots,y_{K_u}\}}(y)\Delta 1_{G^{-1}(A)}(u_k) \\
\geq & \sum_{k=1}^{K_u} 1_{\{\inf G(u_k),\dots,y_{k_{\sup G(u_{k-1})}+1}\}}(y)\Delta 1_{G^{-1}(A)}(u_k) + \sum_{k=1}^{K_u} 1_{\{y_{k_{\sup G(u_{k-1})}+1},\dots,y_{K_u}\}}(y)\Delta 1_{G^{-1}(A)}(u_k) \\
= & \sum_{k=1}^{K_u} 1_{\{\inf G(u_k),\dots,y_{K_u}\}}(y)\Delta 1_{G^{-1}(A)}(u_k) = 1_A(y),
\end{aligned}$$

Which completes the proof.

A.5. **Proof of Proposition 4.** First note that any specification of the latent variable $\epsilon$ that produces the same combinations of equilibria listed in $\mathcal{U}^*$ with the same probabilities are observationally equivalent. We can therefore replace $\mathcal{U}$ by $\mathcal{U}^*$, where each $u \in \mathcal{U}^*$ has probability $Q(u|X;\theta) = \mathbb{P}(G(\epsilon|X;\theta) = u|X)$, and redefine $G$ as the correspondence from $\mathcal{U}^*$ to $\mathcal{Y}$ defined by $G(u) = u$. By theorem 1, $\theta$ belongs to the identified set if and only if for any subset $A$ of $\mathcal{Y}$, $\mathbb{P}(Y \in A|X) \leq \mathbb{P}(G(\epsilon|X;\theta) \cap A \neq \varnothing|X)$ or equivalently $P(A|X) \leq Q(G^{-1}(A)|X;\theta)$. By proposition 1 of Galichon and Henry (2008), this is equivalent to the existence of a probability $\pi$ on $\mathcal{Y} \times \mathcal{U}^*$ with marginal distributions $P(.|X)$ and $Q(.|X;\theta)$ and such that $\pi\{(y,u) \in \mathcal{Y} \times \mathcal{U}^* : y \in u\} = 1$, in other words such that it is supported on the subset of pairs $(y,u)$ such that $y \in u$. This completes the proof.

## References

Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2007): "Econometric tools for analyzing market outcomes," *Handbook of Econometrics*, Volume 6A.

Andrews, D., S. Berry, and P. Jia (2003): "Placing bounds on parameters of entry games in the presence of multiple equilibria," unpublished manuscript.

Andrews, D., and G. Soares (2007): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," unpublished manuscript.

Beresteanu, A., I. Molchanov, and F. Molinari (2008): "Sharp identification regions in games," cemmap working paper CWP15/08.

Beresteanu, A., and F. Molinari (2008): "Asymptotic properties for a class of partially identified models," *Econometrica*, 76, 763–814.

Berry, S. (1992): "Estimation of a model of entry in the airline industry," *Econometrica*, 60, 889–917.

Berry, S., and E. Tamer (2006): "Identification in models of oligopoly entry," in *Advances in Economics and Econometrics*, pp. 46–85. Cambridge University Press.

Bresnahan, T., and P. Reiss (1990): "Entry in monopoly markets," *Review of Economic Studies*, 57, 531–553.

Canay, I. (2007): "Empirical Likelihood Inference for Partially Identified Models: Large Deviation Optimality and Bootstrap Validity," unpublished manuscript.

Castaldo, A., F. Maccheroni, and M. Marinacci (2004): "Random sets and their distributions," *Sankhya (Series A)*, 66, 409–427.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1285.

Choquet, G. (1954): "Theory of capacities," *Annales de l'Institut Fourier*, 5, 131–295.

Ciliberto, F., and E. Tamer (2006): "Market structure and multiple equilibria in airline markets," unpublished manuscript.

Echenique, F., and I. Komunjer (2008): "A test for monotone comparative statics.," unpublished manuscript.

Ekeland, I., A. Galichon, and M. Henry (2008): "Optimal transportation and the falsifiability of incompletely specified economic models.," accepted by *Economic Theory*.

Fan, Y. (2008): "Confidence Sets for Distributions of Treatment Effects With Covariates," unpublished manuscript.

Ford, L., and D. Fulkerson (1957): "A simple algorithm for finding maximal network flows and an application to the Hitchcock problem," *Canadian Journal of Mathematics*, 9, 210–218.

Galichon, A., and M. Henry (2006a): "Dilation Bootstrap. A methodology for constructing confidence regions with partially identified models," unpublished manuscript.

Galichon, A., and M. Henry (2006b): "Inference in incomplete models," Columbia University Discussion Paper 0506-28 available at http://www.columbia.edu/cu/economics/discpapr/DP0506-28.pdf.

Galichon, A., and M. Henry (2008): "A test of non-identifying restrictions and confidence regions for partially identified parameters," forthcoming *Journal of Econometrics*.

Gillies, D. (1953): "Some theorems on n-person games," Princeton Ph.D.

Heckman, J., J. Smith, and N. Clements (1997): "Making the most out of programme evaluation and social experiments: accounting for heterogeneity in programme impacts," *Review of Economic Studies*, 64, 487–535.

Hitchcock, F. (1941): "The distribution of a product from several sources to numerous localities," *Journal of Mathematics and Physics*, 20, 224–230.

Imbens, G., and C. Manski (2004): "Confidence intervals for partially identified parameters," *Econometrica*, 72, 1845–1859.

Jovanovic, B. (1989): "Observable implications of models with multiple equilibria," *Econometrica*, 57, 1431–1437.

Kantorovich, L. (1942): "On the translocation of mass," *Doklady Academii Nauk SSSR*, 37, 199–201.

Koopmans, T. (1949): "Optimum utilization of the transportation system," *Econometrica*, 17, 136–146.

Manski, C. (1990): "Nonparametric bounds on treatment effects," *American Economic Review*, 80, 319–323.

Manski, C., and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546.

Milgrom, P., and J. Roberts (1990): "Rationalizability, learning and equilibrium in games with strategic complementarities," *Econometrica*, 58, 1255–1277.

Monge, G. (1781): *Mémoire sur la théorie des déblais et des remblais*. Académie Royale des Sciences de Paris.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2004): "Moment inequalities and their application," unpublished manuscript.

Papadimitriou, C., and K. Steiglitz (1998): *Combinatorial Optimization: Algorithms and Complexity*. Dover.

Rachev, S., and L. Rüschendorf (1998): *Mass Transportation Problems. Volume I: Theory*. Springer.

Romano, J., and A. Shaikh (2008): "Inference for identifiable parameters in partially identified econometric models," forthcoming in the *Journal of Statistical Planning and Inference*.

Rosen, A. (2008): "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," forthcoming in the *Journal of Econometrics*.

Strassen, V. (1965): "The existence of probability measures with given marginals," *Annals of Mathematical Statistics*, 36, 423–439.

Tamer, E. (2003): "Incomplete simultaneous discrete response model with multiple equilibria," *Review of Economic Studies*, 70, 147–165.

Topkis, D. (1998): *Supermodularity and complementarity*. Princeton University Press.

Villani, C. (2003): *Topics in Optimal Transportation*. Providence: American Mathematical Society.

Vives, X. (1990): "Nash equilibrium with strategic complementarities," *Journal of Mathematical Economics*, 19, 305–321.