Set Identification in Models with Multiple Equilibria

ALFRED GALICHON

École polytechnique

and

MARC HENRY

Université de Montréal

We propose a computationally feasible way of deriving the identified features of models with multiple equilibria in pure or mixed strategies. It is shown that in the case of Shapley regular normal form games, the identified set is characterized by the inclusion of the true data distribution within the core of a Choquet capacity, which is interpreted as the generalized likelihood of the model. In turn, this inclusion is characterized by a finite set of inequalities and efficient and easily implementable combinatorial methods are described to check them. In all normal form games, the identified set is characterized in terms of the value of a submodular or convex optimization program. Efficient algorithms are then given and compared to check inclusion of a parameter in this identified set. The latter are illustrated with family bargaining games and oligopoly entry games.

Key words: Multiple equilibria, Optimal transport, Identified set, Core determining classes

JEL Codes: C13, C72

INTRODUCTION

The empirical study of 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 *et al.* (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 imperfectly competitive markets by Andrews, Berry and Jia (2003) and Ciliberto and Tamer (2009). The inferential method they use, however, relies on a set of restrictions that 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 specific applications to inference in participation games such as oligopoly entry models and group bargaining models.

We first note that the generalized likelihood implied by a model with multiple equilibria can be represented by a non-additive set function called a *Choquet capacity*. Seminal work on co-herency conditions and non-additive likelihoods can be found in Heckman (1978), Gouriéroux, Laffont and Monfort (1980), Manski (1990), and Heckman, Smith and Clements (1997) among others. In cases where identification holds, Dagsvik and Jovanovic (1994) and Bisin, Moro and Topa (2002) propose likelihood-based estimation methods that are robust to the lack of co-herency. Otherwise, the non-additive 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 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 generalized likelihood predicted by the model, where the *core* is a well known and well-studied notion in economics. This results allows the com-putation of the identified feature of models with multiple equilibria and a finite number of ob-servable outcomes, as it reduces the problem to that of checking a finite number of moment inequalities.

The computational burden remains high in situations with a large number of observable out-comes 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 (2006) and Ekeland, Galichon and Henry (2010) in-clude results pertaining to that case. To lift the remaining computational burden, we propose several alternative strategies and discuss their relative merits. First, in cases with only pure strat-egy equilibria, we show that the model generalized likelihood is a submodular function (the set function analogue of a convex function), and that checking that the data distribution belongs to the core of the generalized likelihood is equivalent to minimizing a submodular function, a well-studied problem (discrete analogue of minimizing a convex function) with efficient algorithms and easily available off the shelf implementations. Second, we show that if only pure strategy equilibria are considered, a special case of submodular function minimization applies, which relies on optimal transportation methods and provides more efficient algorithms for the problem of constructing the identified set. Finally, 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 when the model satisfies some monotonicity properties. The method is illustrated on the family bargaining game of Engers and Stern (2002) and the oligopoly entry game of Berry and Tamer (2006).

In cases where mixed strategies are included in the equilibrium concept, Beresteanu, Molchanov and Molinari (2009) (hereafter BMM) were the first to address the problem of con-structing the identified set (whereas, in the case of pure strategies, to the best our knowledge, our work had been the first to address this problem, in Galichon and Henry, 2006). Our incremental contribution in the case of mixed strategy equilibria is two-fold. When the game satisfies a reg-ularity condition defined in Shapley (1971), we show that the identified set is still characterized by the core of the model generalized likelihood, hence by a finite number of inequalities and a submodular optimization problem as before. In all other cases, we give a simple and efficient algorithm to construct the identified set based on convex optimization. This goes beyond the results in BMM that characterize the identified set only by a continuum of inequalities. Thus, our results are complementary to the results of BMM. The methodology proposed here applies to the empirical study of any normal form game, with both pure and mixed strategy equilibria.

In economics, coordination games and participation games are the most natural area of application, including models of labour force participation (including Bjorn and Vuong, 1985), models of discrete choice with social interactions (in Brock and Durlauf, 2001), oligopoly entry models (including Andrews, Berry and Jia, 2003; Ciliberto and Tamer, 2009), auction models (including Bajari, Hong and Ryan, 2010), bargaining models (including Engers and Stern, 2002), and network effects (including Sweeting, 2009). 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 (2009) to complement the seminal contribution of Chernozhukov, Hong and Tamer (2007).

The remainder of the paper is organized as follows. Section 1.1 describes the framework and general results in the case where only equilibria in pure strategies are considered, while Section 1.2 specializes and illustrates them on leading examples of participation games. Section 2 describes three related methods to efficiently compute the identified feature of the model based on the characterization from Section 1.1 and discusses their relative merits. Section 3 illustrates the three methods on an oligopoly entry game with two types of players, Section 4 shows how the results extend to the case where mixed strategy equilibria are also considered. Section 4 also proposes extensive simulation results and Section 5 applies the methodology to the study of the determinants of long-term care option choices for elderly parents in American families. 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). It applies to the empirical analysis of normal form games, where only equilibria in pure strategies are considered (the framework applies also to the analysis of matching models with non-transferable utility, where stability is the relevant equilibrium concept). We defer the extension of our results to mixed and correlated strategies to Section 4. We consider three types of economic variables. Outcome variables Y, exogenous explanatory variables X, and latent variables, or random shocks, ϵ. 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 θ. Without loss of generality, the model may be formalized as a measurable correspondence (defined in Assumption 1 below) between the latent variables ϵ and the outcome variables Y indexed by the exogenous variables X and the vector of parameters θ. We call this correspondence G, and write $Y \in G(\varepsilon|X; \theta)$ to indicate admissible values of Y given values of ε, X, and θ. 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 θ. The latent variables ε is supposed to be distributed according to a parametric distribution $\nu(\cdot|X; \theta)$, where the indexing is meant to indicate that the unknown parameters that enter in the distribution of latent variables are contained in the vector θ 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 σ-algebra of subsets $\mathcal{B}$) are related to unobservable variables ε according to the model $Y \in G(\varepsilon|X;\theta)$, where θ belongs to an open subset Θ of $\mathbb{R}^{d_\theta}$, ε is distributed according to the probability measure $\nu(\cdot|X;\theta)$ on $\mathcal{U}$ (also Polish endowed with its Borel σ-algebra of subsets), and G is a measurable correspondence, i.e. such that for all open subsets A of $\mathcal{Y}$, $G^{-1}(A|X;\theta) := \{\varepsilon \in \mathcal{U}: G(\varepsilon|X;\theta) \cap A \neq \varnothing\}$ 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, ε) 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 $\Pi_1(Y_1, Y_2, \varepsilon_1, \varepsilon_2; \theta) = (\theta Y_2 - \varepsilon_2)Y_1$ and $\Pi_2(Y_1, Y_2, \varepsilon_1, \varepsilon_2; \theta) = (\theta Y_1 - \varepsilon_1)Y_2$, where $Y_i \in \{0, 1\}$ is firm i's action, and $\varepsilon = (\varepsilon_1, \varepsilon_2)'$ are exogenous costs. The firms know their costs; the analyst, however, knows only that ε is uniformly distributed on $[0, 1]^2$, and that the structural parameter θ is in $(0, 1]$. There are two pure strategy Nash equilibria. The first is $Y_1 = Y_2 = 0$ for all $\varepsilon \in [0, 1]^2$. The second is $Y_1 = Y_2 = 1$ for all $\varepsilon \in [0, \theta]^2$ and $Y_1 = Y_2 = 0$ otherwise. Hence, the model is described by the correspondence: $G(\varepsilon; \theta) = \{(0, 0), (1, 1)\}$ for all $\varepsilon \in [0, \theta]^2$, and $G(\varepsilon; \theta) = \{(0, 0)\}$ otherwise.*

To conduct inference on the parameter vector θ, 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 generalized likelihood of an outcome falling in the subset A of $\mathcal{Y}$ predicted by the model is $\mathcal{L}(A|X;\theta) = \nu(G^{-1}(A|X;\theta)|X;\theta)$. Because of multiple equilibria, this generalized likelihood may sum to more than one, as we may have $A \cap B = \varnothing$, and yet $G^{-1}(A|X;\theta) \cap G^{-1}(B|X;\theta) \neq \varnothing$, 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 $A \mapsto \mathcal{L}(A|X;\theta) = \nu(G^{-1}(A|X;\theta)|X;\theta)$ is generally not additive and is called a *Choquet capacity* (see Choquet, 1954). This non-additivity of the model likelihood is referred to as "lack of coherency" in Gouriéroux, Laffont and Monfort (1980) and Tamer (2003).

Definition 1 (Choquet capacity). *A Choquet capacity $\mathcal{L}$ on a finite-set $\mathcal{Y}$ is a set function $\mathcal{L}: A \subseteq \mathcal{Y} \mapsto [0, 1]$ which is*

- *normalized, i.e. $\mathcal{L}(\varnothing) = 0$ and $\mathcal{L}(\mathcal{Y}) = 1$,*
- *monotone, i.e. $\mathcal{L}(A) \leq \mathcal{L}(B)$, for any $A \subseteq B \subseteq \mathcal{Y}$.*

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

As discussed in Jovanovic (1989) and Berry 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) p. 66 of Berry and Tamer, 2006), we define an equilibrium selection mechanism as a conditional distribution $\pi_{Y|\varepsilon, X;\theta}$ over equilibrium outcomes Y in the regions of multiplicity. By construction, an equilibrium selection is allowed to depend on the latent variables ε 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 $\pi(\cdot|\varepsilon, X; \theta)$ for Y conditionally on ε and X, such that the selected value of the outcome variable is actually an equilibrium, or more formally, such that $\pi(\cdot|\varepsilon, X; \theta)$ has support contained in $G(\varepsilon|X; \theta)$.*

The identified feature of the model is the smallest set of parameters that cannot be rejected by the data. Hence, it is the set of parameters for which one can find an equilibrium selection mechanism that completes the model and equates probabilities of outcomes predicted by the model with the probabilities obtained from the data.

Definition 3 (Compatible equilibrium selection mechanism). *The equilibrium selection mechanism $\pi(\cdot|\varepsilon, X; \theta)$ 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) p. 72 of Berry and Tamer, 2006) if for all A measurable subset of $\mathcal{Y}$, $P(A|X) = \int_{\mathcal{U}} \pi(A|\varepsilon, X; \theta)\nu(d\varepsilon|X; \theta)$.*

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

Definition 4 (Identified set). *We call identified set (sometimes somewhat redundantly called sharp identified set), the set Θ_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 allow the computation of the identified set based on the knowledge of the probabilities in the data because the conditional distribution π is an infinite-dimensional nuisance parameter. We now set out to show how to reduce the dimensionality of the problem. Our equivalent formulation of the identified set is based on an appeal to the notion of *core* of the Choquet capacity introduced in Definition 1.

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

In cooperative game theory, a Choquet capacity on a set $\mathcal{Y}$ is interpreted as a game, where $\mathcal{Y}$ is the set of players and $\mathcal{L}$ is the utility value or worth of coalition $A \subseteq \mathcal{Y}$ and the core of the game $\mathcal{L}$ is the collection of allocations that cannot be improved upon by any coalition of players (see Moulin, 1995).

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

The result we propose next 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 that characterizes the generalized likelihood predicted by the model (which we shall call *core of the generalized likelihood predicted by the model*).

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

$$\Theta_I = \{\theta \in \Theta \colon (\forall A \in \mathcal{B}, P(A|X) \le \mathcal{L}(A|X;\theta)); X - \text{a.s.}\}.$$

Remark 1.   Theorem 5 below extends the validity of the result in Theorem 1 to the case of equilibria in mixed strategies with a regular core.

Example 3 (Example 1 continued).   *In Example 1, the identified set is the set of values for θ such that $p \le \theta^2$ and $1 - p \le 1$ where $p = P((Y_1, Y_2) = (1,1))$ 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 checking a finite number of inequalities, *i.e.* $P(A|X) \le \mathcal{L}(A|X;\theta) = \nu(\varepsilon \colon G(\varepsilon|X;\theta) \cap A \neq \varnothing|X;\theta)$ 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 $2^{\text{Card}(\mathcal{Y})} - 2$, and the computational burden is only partially lifted, and the second section of the paper is devoted to efficient methods of computation of the identified set based on the characterization of Theorem 1. First, we turn to the specialization of our results and concepts to some leading examples and illustrate them with a simplified version of a family bargaining game studied in the literature.

1.2. *Some illustrative examples*

1.2.1. 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), and Ciliberto and Tamer (2009). In this set-up, economic agents are firms who decide whether or 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 Π_{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 ε_{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 $\Pi_{im}(Y_m, X_{im}, \varepsilon_{im}; \theta) = \Pi_{im}(Y_{im}, Y_{-im}, X_{im}, \varepsilon_{im}; \theta)$. If, for instance, firms are assumed to play Nash equilibria in pure strategies 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 $\Pi_{im}(Y_{im}, Y_{-im}, X_{im}, \varepsilon_{im}; \theta) \ge \Pi_{im}(1 - Y_{im}, Y_{-im}, X_{im}, \varepsilon_{im}; \theta)$ for all $i = 1, \ldots, J$. Hence, the model can be written $Y_m \in G(\varepsilon_m|X_m; \theta)$, where X_m denotes the matrix of observed profit shifters for firms $i = 1, \ldots, J$, ε_m denotes the vector of latent profit shifters for firms $i = 1, \ldots, J$, and the correspondence G is defined by $G(\varepsilon|X; \theta) = \{Y \colon \Pi_i(Y_i, Y_{-i}, X_i, \varepsilon_i; \theta) \ge \Pi_i(1 - Y_i, Y_{-i}, X_i, \varepsilon_i; \theta); \text{all } i = 1, \ldots, J\}$, where the index m is dropped when considering a generic market.

1.2.2. Family bargaining. For illustration purposes, we consider a simplified version of the bargaining model of decision regarding the long-term care of an elderly parent in Engers and Stern (2002). Consider a family with two children. The issue is which of the children will become the primary care giver of an elderly and disabled parent or whether the parent moves to a nursing home. The pay-off to family member i, $i = 1, 2$ is represented as the sum of three terms. The first term V_{ij} represents the value to child i of care option j, where $j > 0$ means child j becomes the primary care giver and $j = 0$ means the parent is moved to a nursing home. The matrix $V = (V_{ij})_{ij}$ is known to both children. We suppose it takes the form

$$V = \begin{pmatrix} 0 & 2\theta & 4\theta \\ 0 & 4\theta & 2\theta \end{pmatrix},$$

where θ is a non-negative parameter, unknown to the analyst. Both children simultaneously decide whether or not to take part in the long-term care decision. Suppose M is the set of children who participate. The option chosen is option j that maximizes the sum $\sum_{i \in M} V_{ij}$ among the available options (only participating children can become primary care givers). It is assumed that participants abide with the decision and that benefits are then shared equally among children participating in the decision through a monetary transfer s_i, which is the second term in the children's pay-off. The third term ε_i in the pay-off is a random benefit from participation, which is 0 for children who decide not to participate and distributed according to $\nu(\cdot|\theta)$ for children who participate. All children observe the realizations of ε, whereas the analyst only knows its distribution. The matrix of pay-offs of the participation game is given in Table 1.

We derive the equilibrium correspondence, restricting the analysis to pure strategy Nash equilibria for now. We shall later add equilibria in mixed strategies to illustrate results of Section 4.

$\{(0, 0)\}$ is a Nash equilibrium in pure strategies if and only if $\varepsilon_2 < -2\theta$ and $\varepsilon_1 < -2\theta$.
$\{(1, 1)\}$ is a Nash equilibrium in pure strategies if and only if $\varepsilon_2 > \theta$ and $\varepsilon_1 > \theta$.
$\{(0, 1)\}$ is a Nash equilibrium in pure strategies if and only if $\varepsilon_2 > -2\theta$ and $\varepsilon_1 < \theta$.
$\{(1, 0)\}$ is a Nash equilibrium in pure strategies if and only if $\varepsilon_2 < \theta$ and $\varepsilon_1 > -2\theta$.

The equilibrium correspondence $G(\varepsilon|\theta)$ is represented in Figure 1(a) as a function of ε. The dotted lines represents the axes in the ε space, and the full lines represent the frontiers of the regions defining the correspondence G. The shaded area is the area of multiplicity, where $G(\varepsilon|X; \theta)$ contains two values $(0, 1)$ and $(1, 0)$.

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 selection mechanism, the generalized likelihood predicted by the model can be written as follows. As before $\mathcal{Y}$ denotes the set of possible outcomes. The generalized likelihood of observation y is $\mathcal{L}(y|\theta) = \nu(\varepsilon: y \in G(\varepsilon|\theta)|\theta) = \nu(G^{-1}(y|\theta)|\theta)$, for all $y \in \mathcal{Y}$ and $\sum_{y \in \mathcal{Y}} \mathcal{L}(y|\theta) \geq 1$, where the inequality may be strict if there are regions of multiplicity.

TABLE 1
Family game pay-off matrix

Child 1 \ Child 2	N	P
N	0, 0	$4\theta, 2\theta + \varepsilon_2$
P	$2\theta + \varepsilon_1, 4\theta$	$3\theta + \varepsilon_1, 3\theta + \varepsilon_2$

In the case of the family bargaining game, the set of possible outcomes is $\mathcal{Y} = \{(0,0), (0,1), (1,0), (1,1)\}$. The generalized likelihood of outcomes predicted by the model can be written as follows:

$$\mathcal{L}(\{(0,0)\}|\theta) = \nu(\varepsilon: \varepsilon_1 \leq -2\theta, \ \varepsilon_2 \leq -2\theta|\theta) = \nu(G^{-1}((0,0)|\theta)|\theta),$$

$$\mathcal{L}(\{(0,1)\}|\theta) = \nu(\varepsilon: \varepsilon_1 \leq \theta, \ \varepsilon_2 \geq -2\theta|\theta) = \nu(G^{-1}((0,1)|\theta)|\theta),$$

$$\mathcal{L}(\{(1,0)\}|\theta) = \nu(\varepsilon: \varepsilon_1 \geq -2\theta, \ \varepsilon_2 \leq \theta|\theta) = \nu(G^{-1}((1,0)|\theta)|\theta),$$

$$\mathcal{L}(\{(1,1)\}|\theta) = \nu(\varepsilon: \ \varepsilon_1 \geq \theta, \ \varepsilon_2 \geq \theta|\theta) = \nu(G^{-1}((1,1)|\theta)|\theta)$$

and the generalized likelihood of the remaining events can be derived as follows

$$\mathcal{L}(\{(0,0)\} \cup A|\theta) = \mathcal{L}(\{(0,0)\}|\theta) + \mathcal{L}(A|\theta), \ \text{for all } A \subseteq \mathcal{Y}\setminus\{(0,0)\},$$

$$\mathcal{L}(\{(1,1)\} \cup A|\theta) = \mathcal{L}(\{(1,1)\}|\theta) + \mathcal{L}(A|\theta), \ \text{for all } A \subseteq \mathcal{Y}\setminus\{(1,1)\},$$

$$\mathcal{L}(\{(0,1), (1,0)\}|\theta) = 1 - \mathcal{L}(\{(0,0), (1,1)\}|\theta)$$

The generalized likelihood predicted by the model is the set function $A \mapsto \mathcal{L}(A|\theta) = \nu(G^{-1}(A|\theta)|\theta)$ for A subset of $\mathcal{Y} = \{(0,0), (0,1), (1,0), (1,1)\}$. This set function is a Choquet capacity and if the support of ν is sufficiently large, the generalized 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 capacities, as seen here with the inequality $\nu(G^{-1}(\{(0,1)\} \cup \{(1,0)\}|\theta)|\theta) < \nu(G^{-1}(\{(0,1)\}|\theta)|\theta) + \nu(G^{-1}(\{(1,0)\}|\theta)|\theta)$ since the latter is equal to the former plus $\nu(\varepsilon: -2\theta \leq \varepsilon_1 \leq \theta, -2\theta \leq \varepsilon_2 \leq \theta|\theta)$.

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

In the family bargaining example, we have for $i, j = 0, 1$:

$$P((i,j)|X) = \int_{\mathcal{U}} \pi((i,j)|\varepsilon, X; \theta)\nu(d\varepsilon|X; \theta).$$

As noted above, since the model contains no prior information about which outcome is selected in the regions of multiplicity, the identified set Θ_I for the parameter vector θ is the set of parameters for which one can find an equilibrium selection mechanism which completes the model and equates probabilities of outcomes predicted by the model with true outcome frequencies. The definition of the identification region using a semi-parametric likelihood representation, where the equilibrium selection mechanism is included as the infinite-dimensional nuisance parameter π is impractical, so we use Theorem 1 to provide an operational method to compute Θ_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 $\mathcal{L} = \nu G^{-1}$ defined by the model. Hence, we have $\Theta_I = \{\theta \in \Theta: (\forall A \in 2^{\mathcal{Y}}; P(A|X) \leq \mathcal{L}(A|X; \theta);)X\text{a.s.}\}$, where 2^S denotes the set of all subsets of a set S.

In the case of the family bargaining game, the identified region is the set of parameter vectors that satisfy the 16 inequalities $P(A) \leq \nu(G^{-1}(A|\theta)|\theta)$ for all sets A in $2^{\mathcal{Y}}$. Figure 1(b) gives a 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) Model Correspondence   (b) Inequality to check
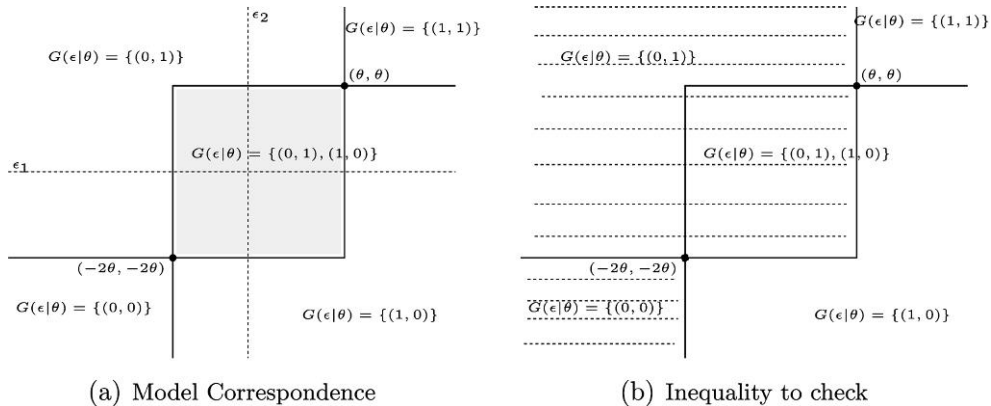
FIGURE 1
Family game

2. EFFICIENT COMPUTATION OF THE IDENTIFIED SET: WHICH INEQUALITIES TO CHECK, AND HOW TO CHECK THEM?

We now describe three approaches to the effective computation of the identified set based on our characterization of Theorem 1. The first approach, described in Section 2.1, is based on submodular optimization and extends readily to the case with mixed strategies, entertained in Section 4. The second approach, described in Section 2.2, relies on the highly efficient algorithms (and easily available packaged implementations) for optimal transportation problems. The third approach, based on the notion of *core-determining* classes of sets and providing a dramatic reduction in the computational complexity under specific assumptions on the game under study, is described in Section 2.3.

2.1. Submodular optimization

The first proposal to deal with the complexity of the problem of checking inequalities in theorem 1 is a method of general validity based on the minimization of a submodular function, the discrete equivalent of a convex function, which is a well-known problem in combinatorial optimization, and for which efficient algorithms and easily available off the shelf implementations exist.

Definition 6 (Submodular function). *A set function* $\mathcal{L}: \mathcal{Y} \to \mathbb{R}$ *is called submodular if for each* $A, B \subseteq \mathcal{Y}$, *we have* $\mathcal{L}(A \cup B) + \mathcal{L}(A \cap B) \leq \mathcal{L}(A) + \mathcal{L}(B)$. *Note that in the special case where* $\mathcal{L}$ *is a probability measure, the latter inequality holds as an equality.*

Submodularity for set functions is the analogue of convexity, and the problem of minimizing a submodular function is well studied (see, for instance, Topkis, 1998, Chapter 2). An up-to-date account of the theory can be found in Fujishige (2005) and off the shelf matlab implementation of the most efficient known algorithms can be found in Andreas Krause's SFO toolbox.

We now show that checking inequalities involved in the characterization of the identified set in Theorem 1 is indeed equivalent to the minimization of a submodular function. Theorem 1 shows that the identified set is the set of values of θ such that X-almost surely, we have the

domination $\forall A \subseteq \mathcal{Y}$, $P(A|X) \leq \mathcal{L}(A|X;\theta)$, or equivalently $\min_{A \subseteq \mathcal{Y}}(\mathcal{L}(A|X;\theta) - P(A|X)) \geq 0$. We first note that the function in the minimization above is indeed submodular.

Lemma 1 (Submodularity of the generalized likelihood).   *For all $\theta \in \Theta$ and all X, the set function on $\mathcal{Y}$ defined for all $A \subseteq \mathcal{Y}$ by $A \mapsto \mathcal{L}(A|X;\theta) - P(A|X)$ is submodular.*

The most efficient generic way to check that a convex function is everywhere non-negative is to minimize it and to verify that the minimum is non-negative. In the same way, we propose to check inequalities of Theorem 1 by minimizing the submodular function defined for all $A \subseteq \mathcal{Y}$ by $A \mapsto \mathcal{L}(A|X;\theta) - P(A|X)$ and checking that the minimum is indeed non-negative. Of course, we can speed up the algorithm by stopping short of the minimum as soon as a negative value is found.

Theorem 2 (Computation of the identified set).   *The identified set is obtained by minimization of a submodular set function: $\Theta_I = \{\theta \in \Theta\colon \min_{B \subseteq \mathcal{Y}}(\mathcal{L}(B|X;\theta) - P(B|X)) = 0, X - a.s.\}$.*

More details on the procedure are given in Section 4, as this method is one of the recommended methods of construction of the identified set in the case where equilibria in mixed strategies are considered. Below is a description of a special case of submodular optimization, which is more efficient and applies to the case where only equilibria in pure strategies are considered.

2.2. *Optimal transportation approach*

In the special case with only pure strategy equilibria that we are considering until Section 4, the model generalized likelihood $\mathcal{L}$ is a very special case of submodular function since it is derived as the distribution function of a random set or random correspondence $\mathcal{L}(A|X;\theta) = \nu(\varepsilon\colon G(\varepsilon|X;\theta) \cap A \neq \varnothing|X;\theta)$. It is that property that allows the following refinement of the method, and the simpler and more efficient technique proposed below. When mixed equilibria are considered, this improvement in efficiency is no longer available, precisely because in general, the model generalized likelihood is no longer the distribution of a random set. To describe the method, we need the following notations and definitions. The terminology used is intended to be self-explanatory. Call $\mathcal{U}^*$ the set of predicted combinations of equilibria, formally $\mathcal{U}^* = \{G(\varepsilon|X;\theta); \varepsilon \in \mathcal{U}\}$ (we suppress reference to the dependence of $\mathcal{U}^*$ on θ for notational convenience). Note that $\mathcal{U}^*$ is a quotient space for the correspondence G. 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 $P(y|X)$ and each vertex $u \in \mathcal{U}^*$ has weight $\nu(\{\varepsilon\colon G(\varepsilon|X;\theta) = u|X\})$. 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 $Q(\cdot|X;\theta)$ the probabilities $Q(u|X;\theta) = \nu(\{\varepsilon\colon G(\varepsilon|X;\theta) = u|X\})$. 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 θ belongs to the identified set if and only if for any subset A of $\mathcal{Y}$, $P(A|X) \leq Q(G^{-1}(A)|X;\theta)$, where $G^{-1}(A) = \{u \in \mathcal{U}^*|A \cap u \neq \varnothing\}$. Galichon and Henry (2009) show that it is equivalent to the existence of a joint probability Λ on $\mathcal{G}(\theta, X)$ with marginal distributions $P(\cdot|X)$ and $Q(\cdot|X;\theta)$. This is summarized in the following proposition.

Theorem 3.   *The parameter value θ belongs to the identified set if and only if there exists a probability on $\mathcal{Y} \times \mathcal{U}^*$ with support contained in $\mathcal{G}(X;\theta)$ and with marginal probabilities $P(\cdot|X)$ and $Q(\cdot|X;\theta)$.*

Note that one implication in Theorem 3 is very easy to prove. Call U the random element with distribution Q. If a joint probability Λ exists with the required properties, then $Y \in A \Rightarrow U \in G^{-1}(A)$, so that $1_{\{Y \in A\}} \leq 1_{\{U \in G^{-1}(A)\}}$, Λ-almost surely. Taking expectation, we have $\mathbb{E}_\Lambda(1_{\{Y \in A\}}) \leq \mathbb{E}_\Lambda(1_{\{U \in G^{-1}(A)\}})$ or equivalently $P(A|X) \leq Q(G^{-1}(A))|X;\theta)$. The converse is much more involved and relies on optimal transportation theory. A similar result is proved in Theorem 3 of Artstein (1983), based on an extension of the marriage lemma.

(Example 1.2.2 continued). For the case of the family bargaining game, $\mathcal{U}^* = \{\{(0,0)\}, \{(0,1)\}, \{(1,0)\}, \{(1,1)\}, \{(0,1), (1,0)\}\}$. The bi-partite graph representing the admissible connections between observable outcomes and combinations of equilibria is shown in Figure 2(a), where p_y denotes $P(y|X)$ and $q_u = \nu(\{\varepsilon: G(\varepsilon|X;\theta) = u|X\})$. 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 q_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 α_y^u, $(y, u) \in \mathcal{Y} \times \mathcal{U}^*$ and must satisfy $\alpha_y^u \geq 0$ for all $(y, u) \in \mathcal{Y} \times \mathcal{U}^*$, $\alpha_y^u = 0$ if $y \notin u$, $\sum_{y \in u} \alpha_y^u = 1$ and equalities such as $p_{01} = \alpha_{01}^{01} + \alpha_{01}^{01,10}$ and $q_{01,10} = \alpha_{01}^{01,10} + \alpha_{10}^{01,10}$.

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 q_u, $u \in \mathcal{U}^*$ units in *terminals* $u \in \mathcal{U}^*$ at minimum cost of transportation, where costs are attached to each pair $(y, u) \in \mathcal{Y} \times \mathcal{U}^*$ is called an optimal transportation problem. Our problem can be reduced to an optimal transportation 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 one otherwise, and there exists a joint law on $\mathcal{G}(X;\theta)$ with marginals P and Q (in other words, θ is in the identified set) if and only if there is a zero-cost solution to the optimization problem thus defined. This minimum cost of transportation problem has an equivalent dual formulation as a maximum flow problem described in Figure 2(b). The edges in the graph with zero cost in the minimum cost of transportation problem have infinite *carrying*
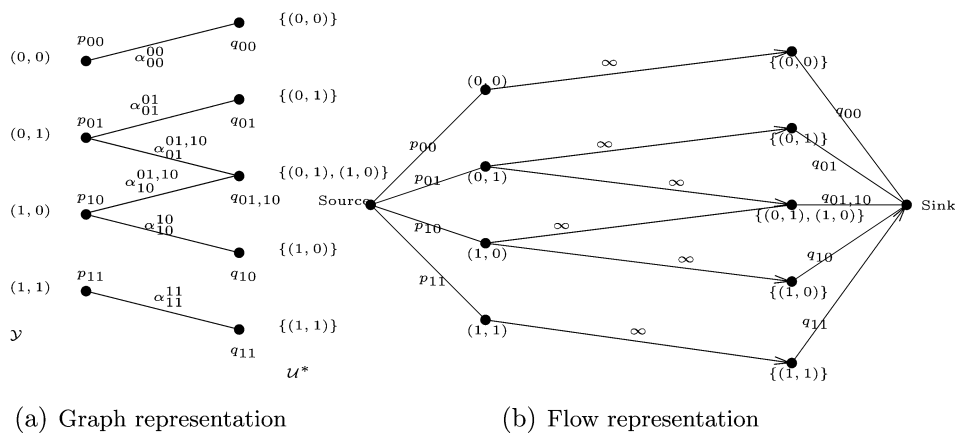

(a) Graph representation (b) Flow representation

FIGURE 2
Family game (continued)

capacity (not to be confused with Choquet capacity) 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 2(b): 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 from y to u such that $y \in u$ (full lines with infinite carrying capacity), and not at all through pairs (y, u) such that $y \notin u$. The parameter value θ is in the identified set if and only if the maximum flow program returns a maximum flow of exactly 1 (note that the network carrying capacities depend on θ through the probabilities q_u). The full procedure is described in detail on the oligopoly entry example of Section 3.

2.3. *Core-determining classes*

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 analyse 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 of Galichon and Henry (2006).

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

A core-determining class of sets is sufficient to characterize the identified region Θ_I as summarized in the following proposition.

Proposition 1. *If $\mathcal{A}$ is core determining for the Choquet capacity $A \mapsto \mathcal{L}(A|X;\theta) = \nu(\varepsilon\colon G(\varepsilon|X;\theta) \cap A \neq \varnothing|X;\theta)$ X-almost surely, and for all θ, then $\Theta_I = \{\theta \in \Theta\colon \forall A \in \mathcal{A}, P(A|X) \leq \mathcal{L}(A|X), X - \text{a.s.}\}$.*

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 Example 1.2.2, before turning to a criterion that will prove useful in exhibiting core-determining classes in many important cases.

(Example 1.2.2 continued).   We return to the family bargaining game and consider some proposals for the computation of the identified set proposed in the literature. We call *Singleton 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 *Singleton class* is not core determining in general. Indeed, if ε has large enough support, the two equalities $P(\{(0,1)\}) = \nu(G(\varepsilon|\theta) \cap \{(0,1)\} \neq \varnothing|\theta) = \nu(\varepsilon\colon \varepsilon_1 \leq \theta, \varepsilon_2 \geq -2\theta|\theta)$ and $P(\{(1,0)\}) = \nu(G(\varepsilon|\theta) \cap \{(1,0)\} \neq \varnothing|\theta) = \nu(\varepsilon\colon \varepsilon_1 \geq -2\theta, \varepsilon_2 \leq \theta|\theta)$ jointly imply $P(\{(0,1),(1,0)\}) > \nu(G(\varepsilon|\theta) \cap \{(0,1),(1,0)\} \neq \varnothing|\theta) = \nu(\varepsilon\colon [\varepsilon_1 \geq -2\theta \text{ or } \varepsilon_2 \geq -2\theta] \text{ and } [\varepsilon_1 \leq \theta \text{ or } \varepsilon_2 \leq \theta]|\theta)$.

We now show how to identify core-determining classes more generally to avoid painstaking case-by-case elimination of redundant inequalities. To that end, we give 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 $\preceq_{\mathcal{Y}}$ of the set of outcomes $\mathcal{Y}$ and an ordering $\preceq_{\mathcal{U}}$ of the set of latent variables $\mathcal{U}$ such that $G(\varepsilon|X;\theta)$ is connected for all $\varepsilon \in \mathcal{U}$, X-a.s., all θ, and $\sup G(\varepsilon_2|X;\theta) \succeq_{\mathcal{Y}} \sup G(\varepsilon_1|X;\theta)$ and $\inf G(\varepsilon_2|X;\theta) \succeq_{\mathcal{Y}} \inf G(\varepsilon_1|X;\theta)$ when $\varepsilon_1 \preceq_{\mathcal{U}} \varepsilon_2$. Both ordering are allowed to depend on the exogenous variables X, but the dependence is suppressed in the notation for clarity.*

Remark 2. This assumption is related to monotone comparative statics in supermodular games (see Milgrom and Roberts, 1990; Vives, 1990; Topkis, 1998).

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

(Example 1.2.2 continued). In the family bargaining game, the orderings are very simple to construct. A lexicographic order on $\mathcal{Y}$ is suitable, with $(0,0) \preceq_{\mathcal{Y}} (0,1) \preceq_{\mathcal{Y}} (1,0) \preceq_{\mathcal{Y}} (1,1)$. On $\mathcal{U}$, the ordering is related to predicted outcomes. All ε producing the same predicted outcomes will be equivalent, and the ordering on predicted outcomes is $\{(0,0)\} \preceq_{\mathcal{U}} \{(0,1)\} \preceq_{\mathcal{U}} \{(0,1),(1,0)\} \preceq_{\mathcal{U}} \{1,0\} \preceq_{\mathcal{U}} \{1,1\}$, where $A_1 \preceq_{\mathcal{U}} A_2$ is a short-hand notation for $\varepsilon_1 \preceq_{\mathcal{U}} \varepsilon_2$ if $G(\varepsilon_1|X;\theta) = A_1$ and $G(\varepsilon_2|X;\theta) = A_2$. It is straightforward to check Assumption 2. For instance, taking ε_1 such that $G(\varepsilon_1|\theta) = \{(0,1)\}$ and ε_2 such that $G(\varepsilon_2|\theta) = \{(0,1),(1,0)\}$ we have $\sup G(\varepsilon_1|\theta) = (0,1) \preceq_{\mathcal{Y}} (1,0) = \sup G(\varepsilon_2|\theta)$ and $\inf G(\varepsilon_1|\theta) = (0,1) = \inf G(\varepsilon_2|\theta)$. This is illustrated in Figure 3.

We are now in a position to state the theorem, which is the main tool in the construction of core-determining classes, and hence in the computation of the identified set under the monotonicity assumption.

Theorem 4. *Suppose Assumption 2 is satisfied with orderings $\preceq_{\mathcal{Y}}$ and $\preceq_{\mathcal{U}}$. Call I the cardinality of $\mathcal{Y}$ and list outcomes (elements of $\mathcal{Y}$) in increasing order (with respect to ordering $\preceq_{\mathcal{Y}}$) as $y_1, \ldots, y_I$. Then $\mathcal{A} = (\{y_1, \ldots, y_i\}, \{y_i, \ldots, y_I\}: i = 1 \ldots, I)$ is core determining.*

Theorem 4 allows to reduce the cardinality of the power set $2^{\mathcal{Y}}$ to twice the cardinality of $\mathcal{Y}$ minus 2 (since the inequality need not be checked on the whole set $\mathcal{Y}$), as we illustrate in our Example 1.2.2.

(Example 1.2.2 continued). In the family bargaining example, Assumption 2 is satisfied, as seen on Figure 3, 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. 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 Theorem 4, and second, to illustrate the power of the combinatorial 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
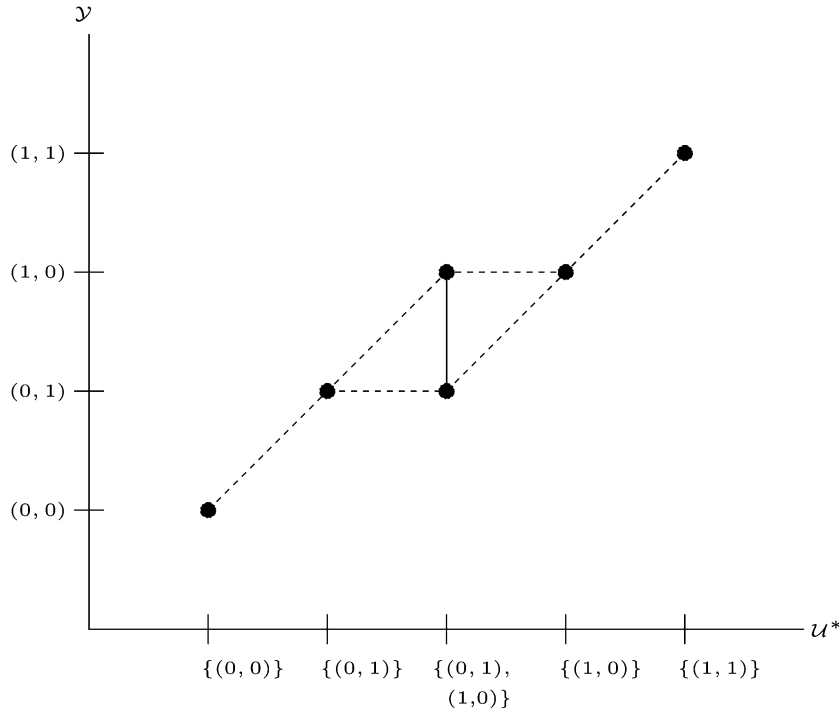
FIGURE 3

The monotonicity requirement in Assumption 2 is satisfied for this choice of orderings in the family bargaining example. (The thick dots represent the correspondence $G(\cdot|\theta)$). $\mathcal{U}^*$ denotes the ordered set of combinations of equilibria

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. (f_1, f_2) is uniformly distributed over $[0, 1]^2$. 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). 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. 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 4, which is taken from Berry and Tamer (2006).
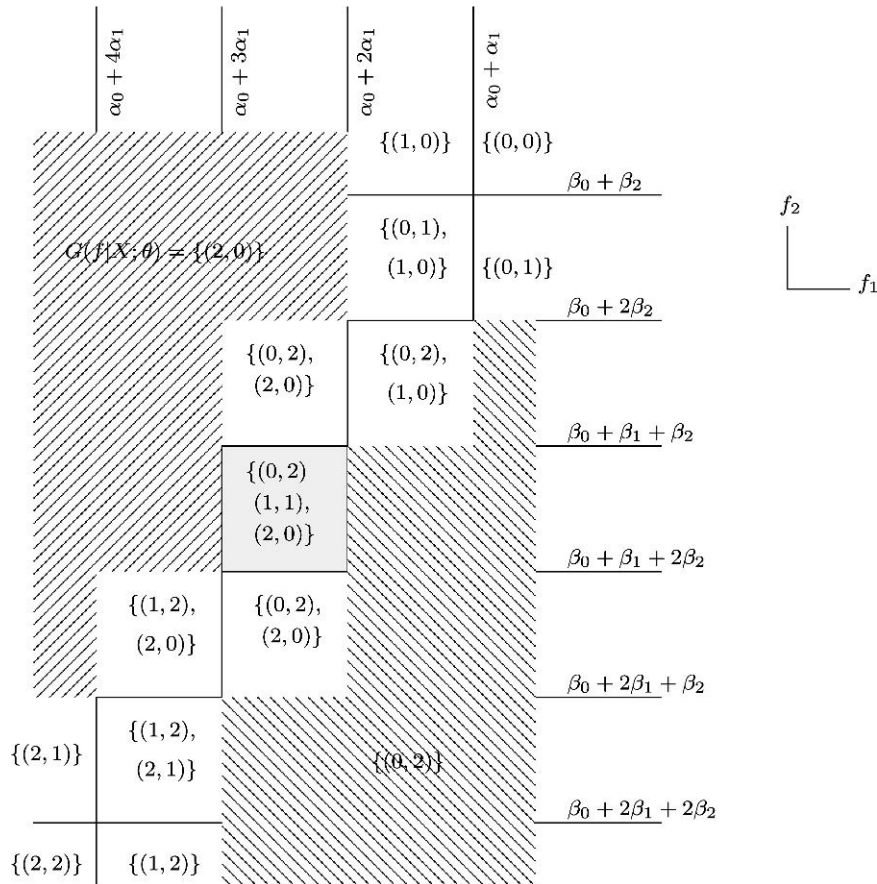
FIGURE 4

Model correspondence in the oligopoly entry game with two types of firms, two of each type. For instance, the top right cell indicates that when the productivity shock is lower than $\alpha_0 + \alpha_1$ for Firm 1 and smaller than $\beta_0 + \beta_1$ for Firm 2, the set of equilibria is the singleton $\{(0,0)\}$

3.1. Submodular optimization approach

The implementation of the submodular optimization method is extremely simple. Once the values of the generalized likelihood have been derived (analytically in the present case, but more often by simulation following the method proposed in Bajari, Hong and Ryan, 2010), all that remains to do to check whether a value θ belongs to the identified set, is to minimize the function defined for all $A \subseteq \mathcal{Y}$ by $A \mapsto \mathcal{L}(A|X;\theta) - P(A|X)$ using, for instance, the Matlab SFO toolbox routine $SFO-min-norm-point.m$ implementing Fujishige's algorithm (see p. 293 of Fujishige, 2005).

3.2. Optimal transportation approach

Consider now the optimal transportation strategy for computing the identified set. The bipartite graph corresponding to this example is represented in Figure 8. As shown in Theorem 3, 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 non-negative 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 8 and one 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 supported on the graph of Figure 8 (in the appendix) and has the correct marginal distributions, hence, it is equivalent to the fact that θ is in the identified set, as we showed in Theorem 3. 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 2. 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 $\sum_{y \in \mathcal{Y}} p_y = \sum_{u \in \mathcal{U}^*} q_u = 1$, and it is equal to 1 if and only if θ is in the identified set.

3.3. *Core-determining class approach*

We now illustrate the usefulness of Theorem 4 for the determination of a core-determining class in this example. Figure 5 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, $(1, 1)$ 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 $u = \{(0, 2), (1, 1), (2, 0)\}$, the mass p_{11} can be removed from $q_{02,11,20}$. Indeed, $(1, 1)$ is

TABLE 2
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}	
$\{(0,0)\}$	∞									q_{00}
$\{(0,1)\}$		∞								q_{01}
$\{(0,1),(1,0)\}$		∞	∞							$q_{01,10}$
$\{(1,0)\}$			∞							q_{10}
$\{(1,0),(0,2)\}$			∞	∞						$q_{02,10}$
$\{(0,2)\}$				∞						q_{02}
$\{(0,2),(2,0)\}$				∞		∞				$q_{02,20}$
$\{(0,2),(1,1),(2,0)\}$				∞	∞	∞				$q_{02,11,20}$
$\{(2,0)\}$						∞				q_{20}
$\{(2,0),(1,2)\}$						∞	∞			$q_{20,12}$
$\{1,2\}$							∞			q_{12}
$\{(1,2),(2,1)\}$							∞	∞		$q_{12,21}$
$\{(2,1)\}$								∞		q_{21}
$\{(2,2)\}$									∞	q_{22}

$\mathcal{Y}$

$(2,2)$

$(2,1)$

$(1,2)$

$(2,0)$

$(1,1)$

$(0,2)$

$(1,0)$

$(0,1)$

$(0,0)$

$\{(0,1)\}$ $\quad$ $\{(1,0)\}$ $\quad$ $\{(0,2)\}$ $\quad\quad$ $\{(2,0)\}$ $\quad$ $\{(1,2)\}$ $\quad$ $\{(2,1)\}$ $\quad$ $\mathcal{U}^*$

$\{(0,0)\}$ $\quad$ $\{(0,1),\ (1,0)\}$ $\quad$ $\{(0,2),\ (1,0)\}$ $\quad$ $\{(0,2),\ (2,0)\}$ $\quad$ $\{(2,0),\ (1,2)\}$ $\quad$ $\{(1,2),\ (2,1)\}$ $\quad$ $\{(2,2)\}$
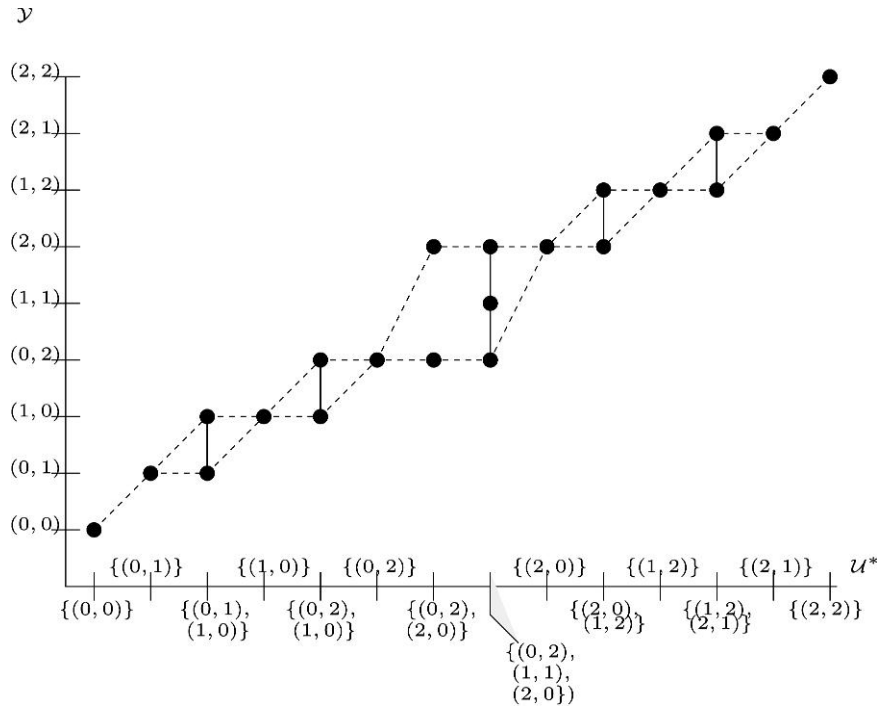
$\{(0,2),\ (1,1),\ (2,0)\})$

FIGURE 5

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

isolated in the sense that all its mass is necessarily transferred to $\{(0,2),(1,1),(2,0)\}$, which is the only predicted combination of equilibria in which it appears. Hence, after checking that $p_{11} \leq q_{02,11,20}$, we can remove p_{11} and $q_{02,11,20}$ and replace $q_{02,20}$ by $q_{02,20} + q_{02,11,20} - p_{11}$. Then the monotonicity and connectedness conditions hold, and Theorem 4 can be applied directly to $\mathcal{Y}\backslash\{(1,1)\}$ and $\mathcal{U}^*$, yielding the class $\mathcal{A} = (\{(0,0)\}, \{(0,0),\ (0,1)\}, \{(0,0),\ (0,1),\ (1,0)\}, \{(0,0),\ (0,1),\ (1,0),\ (0,2)\}, \{(0,0),\ (0,1),\ (1,0),\ (0,2),\ (2,0)\}, \{(0,0),\ (0,1),\ (1,0),\ (0,2),\ (2,0),\ (1,2)\}, \{(0,0),\ (0,1),\ (1,0),\ (0,2),\ (2,0),\ (1,2),\ (2,1)\}, \{(0,1),\ (1,0),\ (0,2),\ (2,0),\ (1,2),\ (2,1),\ (2,2)\}, \{(1,0),\ (0,2),\ (2,0),\ (1,2),\ (2,1),\ (2,2)\}, \{(0,2),\ (2,0),\ (1,2),\ (2,1),\ (2,2)\}, \{(2,0),\ (1,2),\ (2,1),\ (2,2)\}, \{(1,2),\ (2,1),\ (2,2)\}, \{(2,1),\ (2,2)\}, \{(2,2)\})$. Note that its cardinality is $2 \times 7 = 14$, as opposed to the cardinality of the power set of $\mathcal{Y}$ which is $2^9 = 512$.

3.4. *Efficiency comparison of the combinatorial methods*

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 $[0,1]^2$. α_0 and β_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 1 firm enters than a Type 2 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. The values of the model generalized likelihood are derived from these probabilities, and the function $\mathcal{L}(\cdot|X;\theta) - P(\cdot|X)$ is then minimized over all subsets of $\mathcal{Y}$ using the Matlab routine

SFO-min-norm-point.m from the SFO toolbox. An idea of the computational efficiency of this procedure is given by the fact that an order of 10^3 values of the parameter can be tested in 1 s on a standard portable computer. To implement the minimum cost of transportation/maximum flow method, the probabilities of each value of the equilibrium correspondence are entered together with the true frequencies of observable outcomes into the adjacency matrix of Table 2 and the Matlab routine maxflow.m of the BGL library 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 does not. 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. Hence, this method is faster than the general submodular minimization method but only applies to pure strategy equilibria.

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 *Singleton 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 α_1 and β_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 *Singleton class* restrictions, however, is much larger, as shown by its projection on the (β_2, α_1) space in Figure 6. There are also many values of the observed
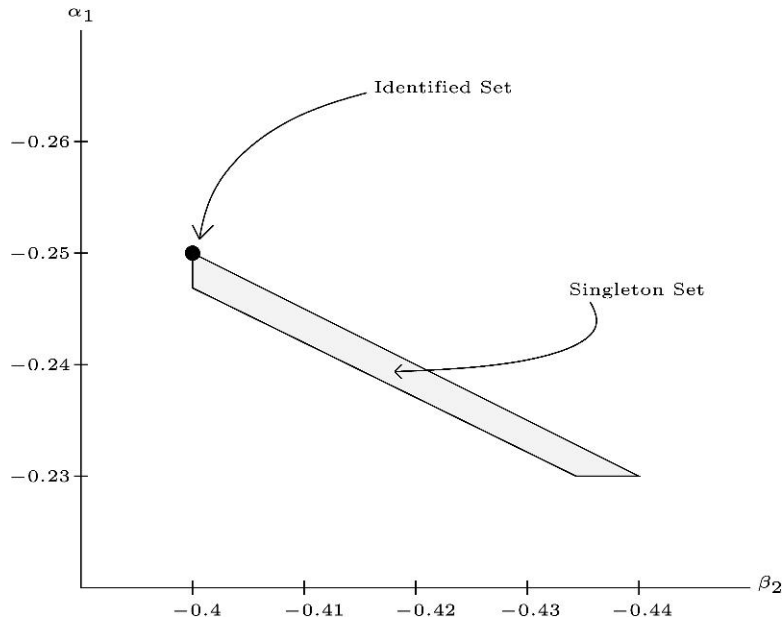

FIGURE 6

Projection of the identified set, and of the set characterized by the Singleton class on the (β_2, α_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 $(p_{00}, p_{01}, p_{10}, p_{02}, p_{11}, p_{20}, p_{12}, p_{21}, p_{22}) = (0.1, 0.15, 0.15, 0.1, 0, 0.5, 0, 0, 0)$

frequencies, for which the identified set is empty, so that the model is rejected, but the set defined by the *Singleton class* restrictions is non-empty, so that it fails to reject the model.

4. EXTENSION TO EQUILIBRIA IN MIXED STRATEGIES

We now show how our results extend to the case where mixed strategy and correlated equilibria are allowed[1]. As before, we consider a game parameterized by the deterministic parameter vector θ, a vector of covariates X and unobserved heterogeneity parameter ε with distribution $\nu(\cdot|X;\theta)$. The observable outcomes of the game are equilibrium actions profiles Y whose realizations belong to the finite-set $\mathcal{Y}$. Call $P(\cdot|X)$ the true distribution of observable outcomes Y conditionally on X. The equilibrium correspondence $G(\varepsilon|X;\theta)$ is now the set of Nash equilibria of the game in mixed strategies (including pure strategies as degenerate mixed strategies) or the set of correlated equilibria, depending on the modelling choice. The model stipulates that there exists an equilibrium selection mechanism $\pi(\cdot|\varepsilon,X;\theta)$ which selected an equilibrium strategy profile σ in $G(\varepsilon|X;\theta)$. The selection mechanism $\pi(\cdot|\varepsilon,X;\theta)$ is a probability distribution on the simplex with support contained in $G(\varepsilon|X;\theta)$. The outcome Y is therefore a realization from the distribution σ and the identified set can be defined as follows.

Definition 8 (Identified set). *The identified set Θ_I is equal to the set of parameter values θ such that X-almost surely there exists a probability $\pi(\cdot|\varepsilon,X;\theta)$ on the $\mathcal{Y}$ simplex with support contained in $G(\varepsilon|X;\theta)$ and for all $y \in \mathcal{Y}$,*

$$P(Y=y|X) = \int \left\{ \int P(Y=y|\varepsilon,\sigma,X;\theta)d\pi(\sigma|\varepsilon,X;\theta) \right\} d\nu(\varepsilon|X;\theta)$$

(Example 1.2.2 continued). We now return to the family bargaining game and derive the equilibrium correspondence $G(\varepsilon|X;\theta)$, which is now a set of probabilities on $\mathcal{Y}$. We note that in addition to the equilibria appearing in Figure 1(a), $\left\{\left(\frac{2\theta+\varepsilon_2}{3\theta},\frac{2\theta+\varepsilon_1}{3\theta}\right)\right\}$ is a Nash equilibrium in mixed strategies if and only if $\theta > \varepsilon_1, \varepsilon_2 > -2\theta$. With the notational convention that (α,β) corresponds to the equilibrium, where players independently participate with respective probabilities α and β, we can represent the equilibrium correspondence as a function of ε in Figure 7.

Remark 3. As noted above, the case with equilibria in correlated strategies can be handled within this framework. In case of correlated strategies, the definition is unchanged, with $G(\varepsilon|X;\theta)$ being the set of probabilities over $\mathcal{Y}$ that define a correlated equilibrium. We believe this to be critical in many settings, both for modelling and computational reasons. Indeed, this allows practitioners to entertain the possibility that players correlate their strategies, with obvious implications for competition policy. As for computation, correlated equilibria have the advantage of being characterized by linear constraints, hence much easier to enumerate than pure strategy Nash equilibria or independent mixed strategy Nash equilibria.

Beresteanu, Molchanov and Molinari (2009) were the first to extend the characterization of the identified set to the case of equilibria in mixed strategies, which is both a very important and non-trivial problem. It is particularly useful, as it is well known that in normal form games, existence of equilibria is guaranteed in mixed strategies, but not necessarily in pure strategies,

---

1. We are grateful to a referee for pointing out the applicability of the definition and results to the case of correlated equilibria.
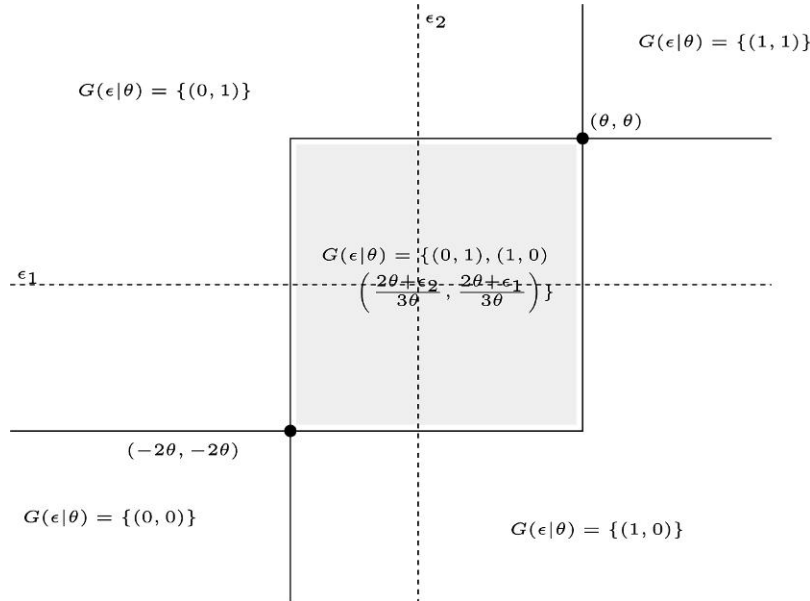
FIGURE 7

Model correspondence for the family bargaining case with mixed strategy equilibria

except in special classes of games such as supermodular games. It is a non-trivial problem since the generalized likelihood of the model is no longer the distribution function of a random set. Nonetheless, we go beyond the results of BMM and show that in certain classes of games, the identified set can be characterized by a finite set of inequalities. We also show that our efficient submodular optimization method is still applicable to this case, as the identified set continues to be characterized by the fact that true data distribution is contained in the core of the model generalized likelihood as in Theorem 1. In the general case, where the generalized likelihood is not necessarily submodular, we show that the identified set can be computed efficiently through a convex optimization problem, which we describe.

Proposition 2 (BMM)**.**

$$\Theta_I = \{\theta \in \Theta \colon \forall f \text{ function on } \mathcal{Y} \colon E_P(f(Y)|X) \leq \tilde{\mathcal{L}}(f|X;\theta)\},$$

where

$$\tilde{\mathcal{L}}(f|X;\theta) = \int \left(\max_{\sigma \in G(\varepsilon|X;\theta)} E_\sigma(f(Y)) \right) d\nu(\varepsilon|X;\theta), \ f \text{ function on } \mathcal{Y}.$$

For completeness, in the appendix, we give a different proof of this result, as it fits more naturally with our presentation than the original random set theoretic proof given in Beresteanu, Molchanov and Molinari (2009). We now show that this characterization allows efficient computation of the identified set as the solution of a convex optimization problem. In addition, under regularity conditions on the equilibrium correspondence of the game, we show that we can derive an exact characterization of the identified set with a finite collection of inequalities, as was the case when considering equilibria in pure strategies only.

4.1. *Efficient computation*

In this section, we give a simple procedure, which is valid for any normal form game, and is recommended to practitioners who do not wish to restrict attention to pure strategy Nash equilibria. In Proposition 2, the functional $\tilde{\mathcal{L}}$ is convex, as a maximum of linear functionals. As a result, we show that the identified set can be computed with a standard convex optimization program. Indeed, Θ_I is the set of θ's such that the convex functional $\phi(f) = \tilde{\mathcal{L}}(f|X; \theta) - E_P(f(Y)|X)$ remains non-negative for all function f. The only difficulty here is that the functional is positively homogeneous, so standard descent algorithms may be stuck at zero, so we propose the following procedure: suppose there is some v (a function on $\mathcal{Y}$, hence a vector in $\mathbb{R}^{d_y}$) such that $\phi(v) < 0$. Since ϕ is positively homogeneous (*i.e.* $\phi(\lambda v) = \lambda \phi(v)$ for all $\lambda \geq 0$), then for each $z \in \mathbb{R}^{d_y}$, such that $z \neq 0$, one of the following three cases applies:

- The inner product of v with z is negative, $\langle z, v \rangle < 0$. Then setting $\lambda = -\langle z, v \rangle^{-1}$, we have $\langle z, \lambda v \rangle = -1$ and $\phi(\lambda v) = \lambda \phi(v) < 0$, so that $\min_{\{w: \langle z, w \rangle = -1\}} \phi(w) < 0$.
- The inner product of v with z is positive, $\langle z, v \rangle > 0$. Then setting $\lambda = \langle z, v \rangle^{-1}$, we have $\langle z, \lambda v \rangle = 1$ and $\phi(\lambda v) = \lambda \phi(v) < 0$, so that $\min_{\{w: \langle z, w \rangle = 1\}} \phi(w) < 0$.
- The inner product is zero, $\langle z, v \rangle = 0$. Then by convexity of ϕ, there is a $\tilde{v}$ in the neighbourhood of v such that $\phi(\tilde{v}) < 0$ and $\langle z, \tilde{v} \rangle > 0$.

Hence, if we take any z, for instance $z = (1, 0, \ldots, 0)$, and define $H = \{w: \langle z, w \rangle = 0\}$, ϕ takes negative values if and only if the following statement is true

$$\min(\min_{w \in H} \phi(w + z), \quad \min_{w \in H} \phi(w - z)) < 0.$$

Since H is a convex set, the program above is a completely standard convex optimization program (we implement the procedure with the matlab optimization toolbox).

4.2. *Cases with a submodular generalized likelihood*

We now turn to the efficient computation of the identified set in the case where the generalized likelihood is submodular. This occurs when the game satisfies the following assumption, taken from Shapley (1971). Recall that the upper envelope ϕ of a family of probability functions $A \mapsto \mu_i(A)$, $i = 1, \ldots, I$ is defined as $A \mapsto \phi(A) = \sup_{i \in I} \mu_i(A)$.

Definition 9 (Regular core (Shapley)). *A normal form game is said to have a* regular core *if the upper envelope $A \mapsto \sup_{\sigma \in G} \sigma(A)$ of the equilibrium correspondence G is submodular.*

We now show a very simple and easy to check sufficient condition for a regular core.

Lemma 2 (Sufficient condition for a regular core). *Consider a normal form game with equilibrium correspondence G. Let S be the support of pure strategies in G. Let S^c be the complement of S in $\mathcal{Y}$. If there is some $\sigma \in G$ such that $\sigma(y) = \max_{\gamma \in G} \gamma(y)$ for all $y \in S^c$, then the game has a regular core.*

Remark 4. Note that a sufficient condition for Lemma 2 is that only one of the equilibria is not in pure strategies, and this condition is very easy to check, and it is satisfied for large classes of games (see Echenique and Edlin, 2004 for games of strategic complementarities and all 2×2 games, when unstable equilibria are removed). For other classes of games, the regularity of the core needs to be checked on a case by case basis.
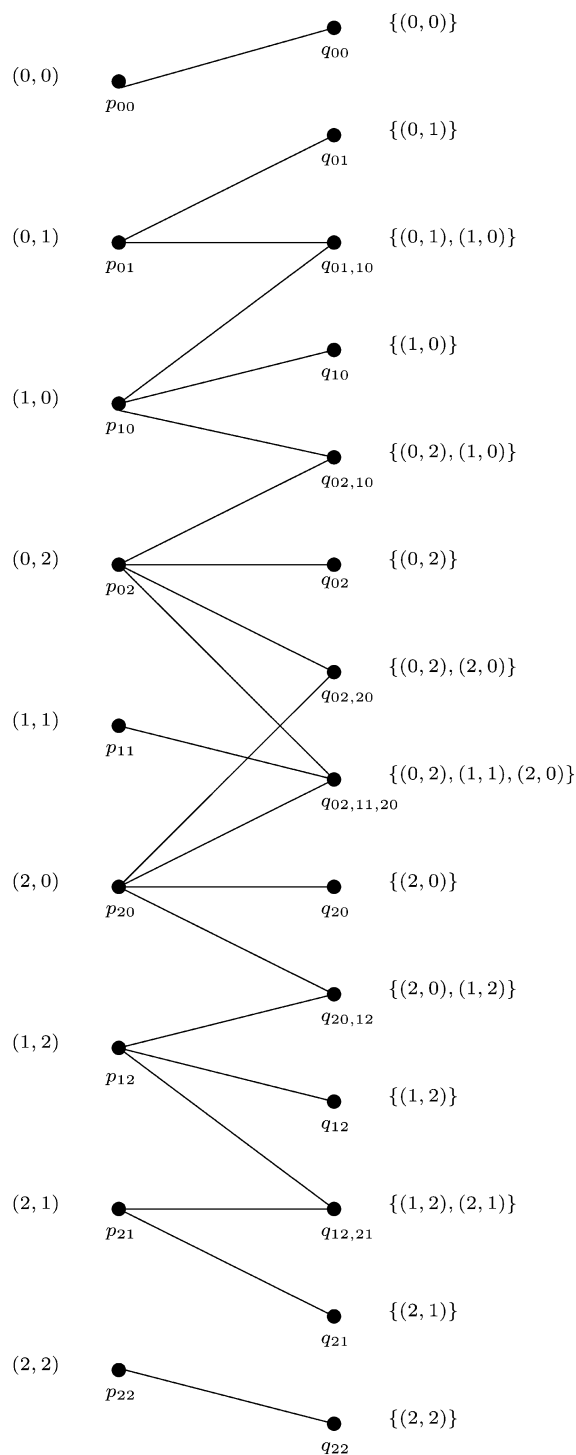
FIGURE 8

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

in the two-type oligopoly entry model

Theorem 5 (Characterization of the identified set). *If the game has a regular core for all θ, (X, ε)-almost surely, then*

$$\Theta_I = \{\theta \in \Theta \colon (\forall B \subseteq \mathcal{Y} \colon P(B|X) \leq \mathcal{L}(B|X; \theta)); X - \text{a.s.}\},$$

where

$$\mathcal{L}(B|X; \theta) = \int \left(\max_{\sigma \in G(\varepsilon|X; \theta)} \sigma(B) \right) d\nu(\varepsilon|X; \theta).$$

In this case, we give an exact characterization of the identified set with only a finite number of inequalities, unlike BMM's characterization. One implication is trivial: it is obvious to see that $E_P(f(Y)|X) \leq \tilde{\mathcal{L}}(f|X; \theta)$ implies $P(B|X) \leq \mathcal{L}(B|X; \theta)$ (it suffices to take $f(y) = 1_B(y)$). The converse is far from trivial, and follows from the characterization of Choquet integrals (see Schmeidler, 1986). This implies that the problem of checking whether a value of the parameter is in the identified set is equivalent to the problem of checking whether the distribution of Y is in the core of a Choquet capacity. This was already the case in Galichon and Henry (2006) when we restricted ourselves to pure strategy equilibria. However, in the latter case, the Choquet capacity was infinitely alternating, as it was defined as the distribution of a random set. With mixed strategy equilibria, it no longer has this property, hence, it is no longer the distribution of a random set.

(Example 1.2.2 continued).   In the family bargaining example, the model generalized likelihood can be derived in the following way: if the support of ε belongs to $[-2\theta, \theta]^2$, then

- $\mathcal{L}(\{(0,1)\}|\theta) = \mathcal{L}(\{(1,0)\}|\theta) = 1$.
- $\mathcal{L}(\{(0,0)\}|\theta) = \mathbb{E}_\nu[(\theta - \varepsilon_2)(\theta - \varepsilon_1)/(9\theta^2)]$.
- $\mathcal{L}(\{(1,1)\}|\theta) = \mathbb{E}_\nu[(2\theta + \varepsilon_2)(2\theta + \varepsilon_1)/(9\theta^2)]$.
- $\mathcal{L}(\{(1,1),(0,0)\}|\theta) = \mathcal{L}(\{(1,1)\}|\theta) + \mathcal{L}(\{(0,0)\}|\theta)$.

The general case, where the domain of ε is unrestricted, can be deduced very simply from the above (see details in Appendix B.1).

As explained in Section 2.1, the strategy to compute the identified set efficiently in the case, where the generalized likelihood is submodular is based on the following observation.

Corollary 1 (Computation of the identified set). *The identified set is obtained by minimization of a submodular set function:* $\Theta_I = \{\theta \in \Theta \colon \min_{B \subseteq \mathcal{Y}}(\mathcal{L}(B|X; \theta) - P(B|X)) = 0, X - \text{a.s.}\}$.

As stated in Definition 6, a set function φ is called submodular if $\varphi(A \cup B) + \varphi(A \cap B) \leq \varphi(A) + \varphi(B)$, and $\varphi(B) = \mathcal{L}(B|X; \theta) - P(B|X)$ is indeed submodular, as shown in the proof of Theorem 5. The notion of submodularity is the analog of convexity for functions defined on lattices, such as set functions. Hence, minimizing a submodular set function is akin to minimizing a convex function, but unlike convex optimization, it is an exact method, and it is generically much more computationally efficient. As noted in Section 2.1, it is a classical problem in combinatorial optimization (see, for instance, Topkis, 1998, Chapter 2), and several polynomial-time algorithms exist (see, for instance, Fujishige, 2005). We implement this using Andreas Krause's Matlab SFO toolbox. Note that, as explained in Section 2.2, the more efficient optimal transportation method does not apply when mixed strategy equilibria are considered.

4.3. *Monte Carlo simulations*

In order to illustrate our fundamental characterization of the identified set, we propose a series of Monte Carlo simulations based on our Example 1.2.2. The objective of the simulations is to illustrate how the core of the generalized likelihood characterizes the identifying power of the model. To this end, we represent the simplex, which is the set of all probability measures for the observable outcomes. Inside the simplex, for a given value of the parameter vector, we represent the core of the generalized likelihood, which is the set of potential data generating processes predicted by the model. We plot the true data-generating process. Recalling that the chosen value of the parameter is in the identified set if and only if the true data-generating process is within the core, the ability to detect a violation of the model will depend on the position of the true data-generating process relative to the boundary of the core of the generalized likelihood. We illustrate this by simulating samples drawn from the true data-generating process and plotting the cloud of resulting empirical distributions. We trim this cloud by removing from the picture a few of the most distant realizations ("most distant" is understood in Euclidian distance in the simplex plane). For three different values of the interaction parameter $\theta = 0 \cdot 25, 0 \cdot 5, 0 \cdot 75$, we represent the core of the generalized likelihood in the four-dimensional simplex. We simulate observed strategy profiles according to probability distributions P in the core of the model. We consider three cases. First, the case where the probability distribution P of the simulated data is the barycenter of the core (this case is called "data generating process [hereafter DGP]"). Second, a case where P is an extreme point of the core (this case is called "extreme DGP"), and finally, an intermediate case (called "intermediate DGP"). In each case, and for each value of θ, we simulated $10,000$ samples of size $n = 100, 1000$, excluded the 5% most distant (case with "confidence = $0 \cdot 95$") or the 10% most distant (case with "confidence = $0 \cdot 9$") and showed graphically how the set of remaining empirical distributions P_n intersects with the core. The 12 graphics pertaining to all cases with $\theta = 0 \cdot 25$ are given in the appendix. In all graphs on Figure 9(a)–(h), the red tetrahedron is the four-dimensional simplex, whose extreme points correspond to the dirac masses on each of the equilibrium profiles $00, 01, 10, 11$. All points inside the tetrahedron have barycentric coordinates corresponding to the vector of probabilities attached to each profile. The blue diamond-shaped polyhedron is the core computed for the value of the parameter $\theta = 0 \cdot 25$, *i.e.* the set of distributions of equilibrium profiles that are compatible with the model for that particular value of θ. The true distribution is a point in the simplex. If the true distribution is a point in the core, then the value of θ is in the identified set.

In Figure 9(a), the core is plotted for the case $\theta = 0 \cdot 25$ and $10,000$ data series of size $n = 100$ are drawn from the analytical center of the core. For each simulated sample, the empirical distribution is computed, and the green ball-shaped polyhedron is the set of 9000 out of the $10,000$ computed empirical distributions. The 1000 that are excluded are the most distant (in Euclidian distance) from the data-generating process and they are excluded for visual convenience. The sensitivity to the proportion of plotted empirical distributions (called "confidence" in the captions) can be seen by comparing Figure 9(a) and (b) in which only 500 outliers are excluded. We see that in both Figure 9(a) and (b), a sizeable proportion of draws falls outside the core, which does not mean, however, that a confidence ball around such empirical distributions would not intersect with the core, hence, it does not necessarily lead to the exclusion of θ from a confidence region for the identified set. In the following graphics, Figure 9(c) and (d), the set-up is identical, except for the fact that the samples drawn have size 1000. In that case, it is clear that all drawn empirical distributions are well within the core, which would lead to include θ in any reasonably constructed confidence region for the identified set. In Figure 9(e)–(h), the data are generated from an extreme point of the core. In such a case, a point in the neighbourhood of the data-generating process is more likely to fall outside the core than inside it. Hence, for all sample
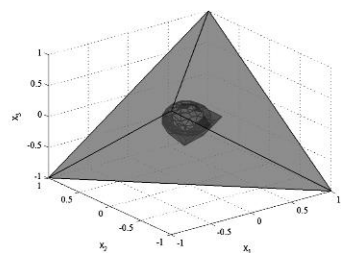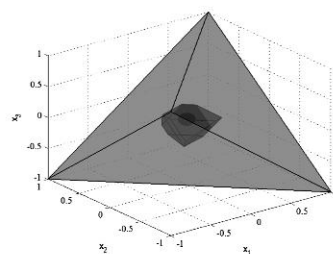
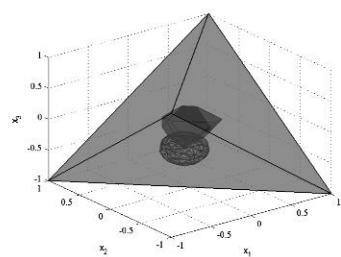(a) $\theta = 0.25$, Central DGP, $n = 100$, 90% confidence
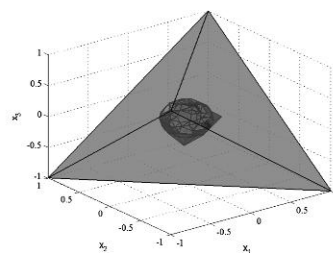
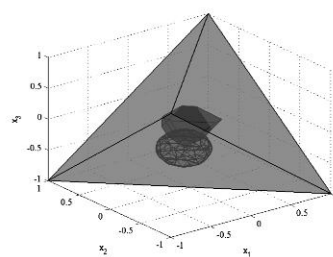(b) $\theta = 0.25$, Central DGP, $n = 100$, 95% confidence
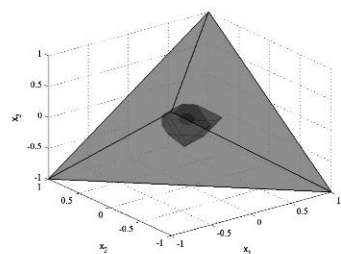
(c) $\theta = 0.25$, Central DGP, $n = 1000$, 90% confidence
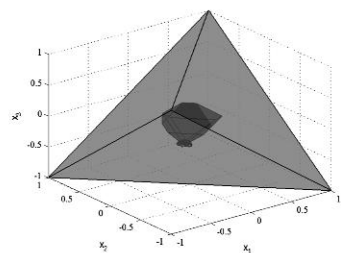
(d) $\theta = 0.25$, Central DGP, $n = 1000$, 95% confidence
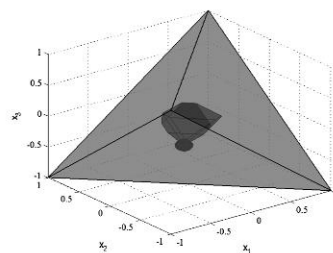
(e) $\theta = 0.25$, Extreme DGP, $n = 100$, 90% confidence

(f) $\theta = 0.25$, Extreme DGP, $n = 100$, 95% confidence

(g) $\theta = 0.25$, Extreme DGP, $n = 1000$, 90% confidence

(h) $\theta = 0.25$, Extreme DGP, $n = 1000$, 95% confidence

FIGURE 9

Core representations in the probability simplex

sizes, the proportion of empirical distributions that fall outside the core is relatively stable and larger than the proportion that falls inside the core. However, for larger sample sizes, we see that simulated empirical distributions are more likely to be within a predetermined neighbourhood of the core, so that the value θ is less likely to be excluded from a confidence region for the identified set. The cases where $\theta = 0.5$ and $\theta = 0.75$ are very similar and we do not report them here.

5. EMPIRICAL ILLUSTRATION: DETERMINANTS OF LONG-TERM CARE FOR ELDERLY PARENTS

The applicability of the methodology proposed is best illustrated on data pertaining to our Example 1.2.2. This is intended to serve as a template for the application of our methodology beyond empirical industrial organization to models of group decision making with free-riding incentives, such as investment in the procurement of a public good. We estimate the determinants of long-term care option choices for elderly parents in American families. The data consist of a sample of 948 families with two children drawn from the National Long Term Care Survey, sponsored by the National Institute of Aging and conducted by the Duke University Center for Demographic Studies under Grant number U01-AG007198, Duke (1999). Elderly people were interviewed in 1984 about their living and care arrangements. The survey questions include gender and age of the children, the distance between homes of parents and each of the children, the disability status of the elderly parent (where disability is referred to as problems with "Activities of Daily Living or Instrumental Activities of Daily Living (ADL)") and the number of hours per week each of the children devote to the care of the elderly parent. The dependent variable is the care provision for the parent. The parent is asked to list children (either at home or away from home) and how much each provides help. If only one child is listed as providing significant help, that child is designated the primary care giver. If more than one child is listed, the one providing the most hours is designated the primary care giver. If no child is listed or if the parent lives in a nursing home, then the parent is designated as "living alone."

The observable choice of care option is modelled as in Engers and Stern (2002) as the outcome of a family bargaining game. The pay-off to family member i, $i = 0, 1, 2$ is represented as the sum of three terms. The first term V_{ij} represents the value to parent $(i = 0)$ or child i, $i = 1, 2$, of care option j, where $j > 0$ means child j becomes the primary care giver and $j = 0$ means the parent remains alone or is moved to a nursing home. The matrix $V = (V_{ij})_{ij}$ is known to both children. We suppose it takes the form

$$
V = \begin{pmatrix}
0 & \beta_0 + \text{ADL}\beta & \beta_0 + \text{ADL}\beta \\
0 & \beta_0 + \text{ADL}\beta + D_1\psi + F\alpha & \beta_0 + \text{ADL}\beta \\
0 & \beta_0 + \text{ADL}\beta & \beta_0 + \text{ADL}\beta + D_2\psi
\end{pmatrix},
$$

where ADL is a dummy variable that takes the value 1 if the parent has problems with activities of daily living or instrumental activities of daily living, D_i is the distance between parent and child i, $F = 1$ if child 1 (who is first born) is female, and zero otherwise and $\theta = (\beta_0, \beta, \psi, \alpha)'$ is a four-dimensional parameter vector, unknown to the analyst. The value a family attaches to the fact that a non-disabled parent is taken care of by one of the children is measured by β_0, whereas for a disabled parent, it is $\beta_0 + \beta$. ψ measures the disutility to the caregiver of living more than an hour away from the parent, and α measures the incremental utility of taking care of the parent for the firstborn daughter, as compared with a first-born son.

In the data, 85% of interviewed parents have problems with activities of daily living. Seventy percent of parents live alone. Forty-six percent of first-born children are female, whereas 56%

of first born who are primary care takers are female. The female first-born effect was identified in Engers and Stern (2002), and we wish to see whether it is robust to our analysis that takes multiple equilibria in the game seriously.

Both children simultaneously decide whether or not to take part in the long-term care decision. Suppose M is the set of children who participate. The option chosen is option j that maximizes the sum $\sum_{i \in M} V_{ij}$ among the available options (only participating children can become primary care givers). It is assumed that participants abide with the decision and that benefits are then shared equally among children participating in the decision through a monetary transfer s_i, which is the second term in the children's pay-off. The third-term ε_i in the pay-off is a random benefit from participation, which is 0 for children who decide not to participate and distributed according to $\nu(\cdot|\theta)$ for children who participate. All children observe the realizations of ε, whereas the analyst only knows its distribution (the negative shock on participation is calibrated with a normal distribution $N(-3, 1)$ for ε. An alternative would have been a negative exponential distribution for ε).

The pay-off matrix of the participation game can be determined in the following way:

- If both children decide to participate, child i's pay-off (for $i = 1, 2$) is

$$X_i = \max\left(0, \beta_0 + \mathrm{ADL}\beta + \frac{1}{2}(D_1\psi + F\alpha), \beta_0 + \mathrm{ADL}\beta + \frac{1}{2}D_2\psi\right) + \varepsilon_i.$$

- If neither child participates, their pay-offs are 0.
- If only Child 1 participates, Child 1's pay-off is

$$X_1 = \max(0, \beta_0 + \mathrm{ADL}\beta + D_1\psi + F\alpha) + \varepsilon_1$$

  and Child 2's pay-off is

$$X_2 = (\beta_0 + \mathrm{ADL}\beta)1_{\{\beta_0 + \mathrm{ADL}\beta + D_1\psi + F\alpha \geq 0\}}.$$

- If only Child 2 participates, Child 2's pay-off is

$$X_2 = \max(0, \beta_0 + \mathrm{ADL}\beta + D_2\psi) + \varepsilon_2$$

  and Child 1's pay-off is

$$X_1 = (\beta_0 + \mathrm{ADL}\beta)1_{\{\beta_0 + \mathrm{ADL}\beta + D_2\psi \geq 0\}}.$$

We do not observe participation, but only the chosen care option. To each equilibrium strategy profile corresponds a (almost surely) unique care option choice, hence for each participation shock ε, we can derive $G(\varepsilon|F, D, \mathrm{ADL}; \theta)$ as the set of probability measures on the set of care options $\{0, 1, 2\}$ induced by mixed strategy profiles, which are probabilities on the set of participation profiles $\{(N, N), (N, P), (P, N), (P, P)\}$. The Gambit software is a good option for the computation of $G(\varepsilon|F, D, \mathrm{ADL}; \theta)$.

The methodology proposed in the paper allows the construction of the identified set based on the hypothetical knowledge of the true distribution of the data. In order to account for sampling uncertainty, we appeal to the bootstrap procedure proposed in Henry, Méango and Queyranne (2010) to construct confidence regions for the identified set in such situations. The method relies on a bootstrap determination of a set function $A \mapsto \underline{P}(A|F, D, \mathrm{ADL})$ that is is dominated by $P(A|F, D, \mathrm{ADL})$ (uniformly over $A \subseteq \{0, 1, 2\}$, X, D, and ADL) with probability $1 - \alpha$ (the chosen level of significance, here 0·95). Once $\underline{P}$ is determined, one proceeds as recommended

in Section 4.2. That is to say, we keep in the identified set only values of θ such that for all observed values of the explanatory variables, the minimum over $A \subseteq \{0, 1, 2\}$ of the function $\mathcal{L}(A|F, D, \text{ADL}; \theta) - \underline{P}(A|F, D, \text{ADL})$ is non-negative. The search over the four-dimensional parameter space is conducted in the following way. First, one finds a value of the parameter that lies within the confidence region (for instance, the corresponding estimates in the analysis of Engers and Stern, 2002 which achieves identification by removing multiplicity of equilibria), then one chooses a coarse discretization of $[-\pi, \pi]^3$ to search in all directions for the frontier point of the confidence region, which is assumed to be star-shaped (on each arc, we use a dichotomic search). The resulting frontier points are the extreme points of the confidence polytope. Each value of the parameter can be tested in a fraction of a second on a standard laptop, and the region can be constructed in a few hours, again on a standard laptop without parallel processing. The confidence region is a four-dimensional polytope and three-, two-, or one-dimensional confidence regions for subsets or linear combinations of parameters can be easily visualized as cuts from the confidence region using the matlab multiparametric MPT toolbox.

The variables chosen were those that were significant in Engers and Stern (2002). We test the significance of each of the individual parameters by checking whether the hyperplanes defined by $\beta_0 = 0$, $\beta = 0$, $\psi = 0$, and $\alpha = 0$ intersect with our 95% confidence region. The range of values of β_0 is $\beta_0 \in [-4 \cdot 297, 1 \cdot 426]$. Hence, the hyperplane $\beta_0 = 0$ has non-empty intersection with the confidence region, which means we fail to reject (at the 5% level) the null hypothesis that the family values identically the fact that a non-disabled parent lives alone or is taken care of by a family member. On the other hand, the range of values for β is $\beta \in [1 \cdot 976, 7 \cdot 297]$, so the half space $\beta \leq 0$ has empty intersection with the confidence region, and we reject the null hypothesis that a parent's disability does not increase the value of care provided by the family. We also reject the hypothesis that distance between parent and caretaker is insignificant, as the range for ψ is $\psi \in [-2 \cdot 981, -1 \cdot 306]$. Finally, we reject the hypothesis that gender of the first-born child does not affect the chosen care option as the range for α is $\alpha \in [0 \cdot 409, 2 \cdot 405]$. Indeed, our results are still consistent with Engers and Stern (2002) in the finding that families value care provided by a first-born daughter more than care provided by a first-born son. Since the hypothesis that $\beta_0 = 0$ is not rejected, we slice the confidence region at $\beta_0 = 0$ to obtain the constrained confidence region, which is three-dimensional and is represented (from four different angles) in Figure 10. In the constrained region, the ranges for the three remaining parameters are considerably tighter, as $\beta \in [2 \cdot 717, 4 \cdot 062]$, $\psi \in [-2 \cdot 867, -1 \cdot 365]$ and $\alpha \in [0 \cdot 532, 2 \cdot 290]$. Two-dimensional regions are also given for (α, ψ), when $\beta = 3$ and (β, α) when $\psi = -2$. These regions are plotted in Figure 11. Note the thin diagonal shapes of the regions, which makes them rather informative, as for instance simultaneous large values of ψ and α are rejected, which roughly means that first-born daughters living close by are not the only possible caregivers. Similarly, β and ψ are not simultaneously very large, so that either the disability effect is very large or the distance effect is very large but not both.

6. DISCUSSION AND EXTENSIONS

In the context of models with multiple equilibria in pure and mixed strategies, we have proposed an equivalence result between the existence of an equilibrium selection mechanism compatible with the data and a finite set of inequalities characterizing the core of the model generalized 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 easily implementable combinatorial methods. We identify several extensions of our work. First, our analysis of mixed strategies shows that the identified set is characterized by a finite collection of inequalities under a property we call *regular core*. We derive sufficient conditions to show that this property holds
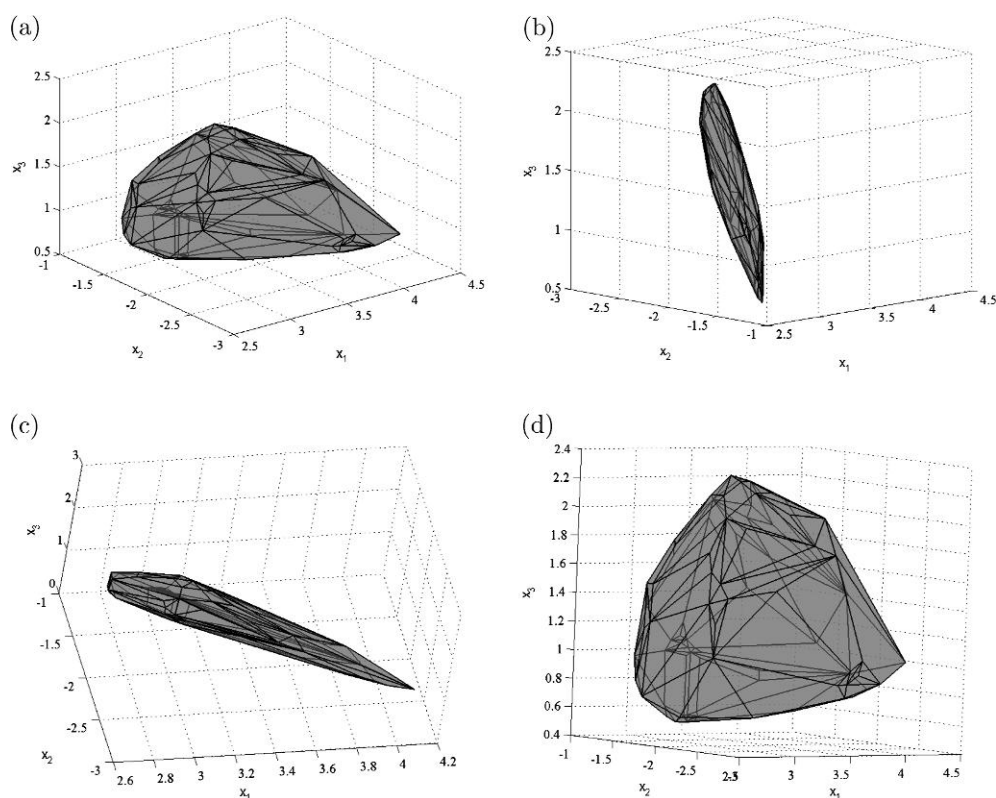
(a)

(b)

(c)

(d)

FIGURE 10

(β, ψ, α) region at $\beta_0 = 0$

for many games, including games with an equilibrium in mixed strategies and multiple equilibria in pure strategies, but more research is needed to characterize the scope and applicability of the regular core property of games. Second, we present a characterization of the identified set in models where the unobservable heterogeneity is parameterized, and the parametric assumption plays a role in the tightness of the identified region. This assumption is characteristic of the literature in empirical industrial organization. However, different assumptions regarding unobserved heterogeneity need to be entertained, including moment conditions, as is customary in econometrics. Some results pertaining to that case are presented in Ekeland, Galichon and Henry (2010). Once we depart from parametric assumptions for the unobserved heterogeneity, the identified region is likely to be much larger, and the effect of incorporating extra prior information, such as shape restrictions on the distribution of unobserved heterogeneity needs to be analysed. Third, the present work proposes an analysis of the identifying power of models of interaction. For practical implementation purposes, the effect of sampling uncertainty needs to be analysed. The issue of constructing confidence regions for the identified set characterized here is taken up in Galichon and Henry (2009), which complements the seminal work of Chernozhukov, Hong and Tamer (2007), but much remains to be done, both to improve efficiency of computation of confidence regions, and to sharpen their statistical properties. Finally, results are presented here in a very general framework, including any normal form game. We have shown that the tools provided here allow efficient inference in a large class of empirically relevant models, such

(a) (ψ, α) region at $(\beta_0, \beta) = (0, 3)$.   (b) (β, ψ) region at $(\beta_0, \alpha) = (0, 2)$.
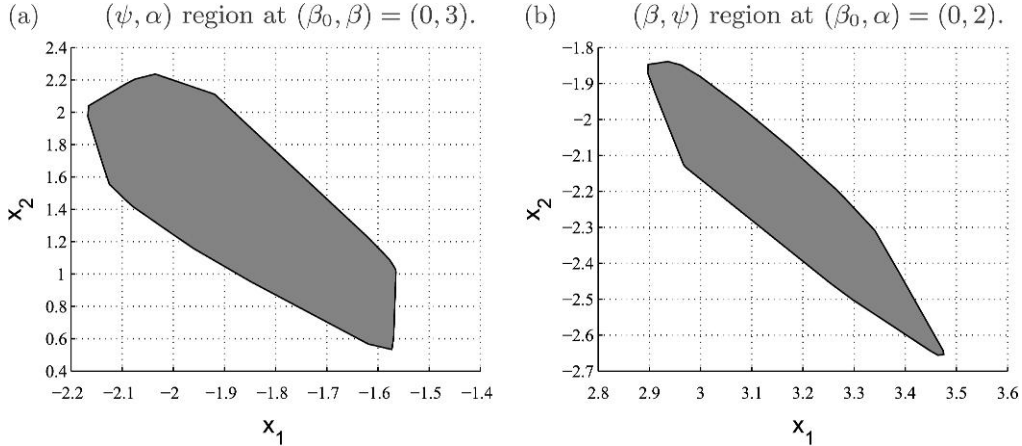
FIGURE 11
Bivariate confidence regions

as models of oligopolistic competition or group decision making. However, ideas developed in the paper can be fruitfully applied to specific classes of models, such as matching models with non-transferable utility as in Echenique, Lee and Shum (2010) and more generally models of social network formation such as in Christakis *et al.* (2010), where computational issues remain daunting.

APPENDIX A PROOFS OF RESULTS IN THE MAIN TEXT

A.1. Proof of Theorem 1

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

A.2. Proof of proposition 3

We first 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 $A \mapsto \mathcal{L}(A|X;\theta) = \nu(\varepsilon\colon G(\varepsilon|X;\theta) \cap A \neq \varnothing|X;\theta)$ if for every measurable subset A of $\mathcal{Y}$, there exists non-negative integers K, L, N, α_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 $\varepsilon \in \mathcal{U}$, $1_A(y) \leq \frac{1}{N}\left(\sum_{k=1}^{K} \alpha_k 1_{A_k}(y) - L\right)$ and $1_{\{G(\varepsilon|X;\theta) \cap A \neq \varnothing\}}(\varepsilon) \geq \frac{1}{N}\left(\sum_{k=1}^{K} \alpha_k 1_{\{G(\varepsilon|X;\theta) \cap A_k \neq \varnothing\}}(\varepsilon) - L\right)$.*

Proof of proposition 3. Subtracting the second inequality in Proposition 3 yields $1_A - 1_{\{\varepsilon\colon G(\varepsilon|X;\theta) \cap A \neq \varnothing\}} \leq \frac{1}{N}\left(\sum_{k=1}^{K} \alpha_k(1_{A_k} - 1_{\{\varepsilon\colon G(\varepsilon|X;\theta) \cap A_k \neq \varnothing\}})\right)$. Taking expectations (conditionally on X) on both sides of the previous equation yields $P(A|X) - \nu(\varepsilon\colon G(\varepsilon|X;\theta) \cap A \neq \varnothing|X;\theta) \leq \frac{1}{N}\left(\sum_{k=1}^{K} \alpha_k(P(A_k|X) - \nu(\varepsilon\colon G(\varepsilon|X;\theta) \cap A_k \neq \varnothing|X;\theta))\right)$. This in turn implies that $P(A|X) \leq \nu(\varepsilon\colon G(\varepsilon|X;\theta) \cap A \neq \varnothing|X;\theta)$ if $P(A_k|X) \leq \nu(\varepsilon\colon G(\varepsilon|X;\theta) \cap A_k \neq \varnothing|X;\theta)$ for each k, which means that $\mathcal{A}(\theta)$ is indeed core determining, which completes the proof. ‖

We now turn to the proof of Theorem 4. 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 θ in the notation, so that $G\colon u \mapsto G(u)$ 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 $\{u \in \mathcal{U}^*\colon G(u) \cap A \neq \varnothing\}$. 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

$$
\begin{aligned}
&\sum_{k=1}^{K_u} 1_{\{\inf G(u_k), \ldots, y_{K_y}\}}(y) \Delta 1^{+}_{G^{-1}(A)}(u_k) + \sum_{k=1}^{K_u} 1_{\{y_1, \ldots, \sup G(u_{k-1})\}}(y) \Delta 1^{-}_{G^{-1}(A)}(u_k) - \sum_{k=1}^{K_u} \Delta 1^{-}_{G^{-1}(A)}(u_k) \\
&= \sum_{k=1}^{K_u} 1_{\{\inf G(u_k), \ldots, y_{K_y}\}}(y) \Delta 1^{+}_{G^{-1}(A)}(u_k) - \sum_{k=1}^{K_u} 1_{\{y_{k_{\sup G(u_{k-1})}+1}, \ldots, y_{K_u}\}}(y) \Delta 1^{-}_{G^{-1}(A)}(u_k). \quad (A1)
\end{aligned}
$$

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 (A1) is equal to

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

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

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

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

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

Which completes the proof.

A.3. Proof of Theorem 3

First note that any specification of the latent variable ε 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 $Q(u|X;\theta) = \nu(\varepsilon: G(\varepsilon|X;\theta) = u|X;\theta)$ and redefine G as the correspondence from $\mathcal{U}^*$ to $\mathcal{Y}$ defined by $G(u) = u$. By Theorem 1, θ belongs to the identified set if and only if for any subset A of $\mathcal{Y}$, $P(A|X) \leq \nu(\varepsilon: G(\varepsilon|X;\theta) \cap A \neq \varnothing|X;\theta)$ or equivalently $P(A|X) \leq Q(G^{-1}(A)|X;\theta)$. By Proposition 1 of Galichon and Henry (2009), this is equivalent to the existence of a probability Λ on $\mathcal{Y} \times \mathcal{U}^*$ with marginal distributions $P(\cdot|X)$ and $Q(\cdot|X;\theta)$ and such that $\Lambda\{(y,u) \in \mathcal{Y} \times \mathcal{U}^*: y \in u\} = 1$, in other words such that it is supported on the subset of pairs (y,u) such that $y \in u$. This completes the proof.

A.4. Proof of Proposition 2

The term in curly brackets in equation (1) is a convex combination of elements of $G(\varepsilon|X;\theta)$, hence the existence of an equilibrium mechanism with the required property is equivalent to the existence of a family $\pi(\cdot|\varepsilon, X;\theta)$ of probability measures on $\mathcal{Y}$ such that $\pi(\cdot|\varepsilon, X;\theta) \in \text{co}(G(\varepsilon|X;\theta))$, ε-almost surely. By Choquet's Theorem (see Choquet, 1954), this implies that for any function f on $\mathcal{Y}$, $\sum_{\mathcal{Y}} f(y)\pi(y|\varepsilon, X;\theta) \leq \max\{\sum_{\mathcal{Y}} f(y)s(y); s \in G(\varepsilon|X;\theta)\}$, which, by integration over ε, implies $\sum_{\mathcal{Y}} f(y)P(y|X) \leq \int\{\max_{s \in G(\varepsilon|X;\theta)} \sum_{\mathcal{Y}} f(y)s(y)\}d\nu(\varepsilon|X)$. The latter yields the desired representation after dividing both sides by the norm of f.

Conversely, suppose that for all functions f on $\mathcal{Y}$, we have

$$\sum_{\mathcal{Y}} f(y)P(y|X) \leq \int L_{\varepsilon|X;\theta}(f)d\nu(\varepsilon|X;\theta),$$

where $L_{\varepsilon|X;\theta}(f) = \max\{\sum_{\mathcal{Y}} f(y)s(y); s \in G(\varepsilon|X;\theta)\}$. By the Strassen Disintegration Theorem (see Strassen, 1965), there exists a map $\varepsilon \mapsto \pi(\cdot|\varepsilon, X;\theta)$ such that for all $y \in \mathcal{Y}$, $P(y|X) = \int \pi(y|\varepsilon, X;\theta)d\nu(\varepsilon|X)$ and $\pi(\cdot|\varepsilon, X;\theta) \leq L_{\varepsilon|X;\theta}(\cdot)$, ε-almost surely. Again by the Choquet Theorem, the latter condition is equivalent to $\pi(\cdot|\varepsilon, X;\theta) \in \text{co}(G(\varepsilon|X;\theta))$, ε-almost surely, and the result follows.

A.5. Proof of Lemma 2

Consider a game with equilibrium correspondence G and call μ the upper envelope of the equilibrium correspondence, i.e. for all $A \subset \mathcal{Y}$, $\mu(A) = \sup_{\sigma \in G} \sigma(A)$. Let S be the support of pure strategies and $S^c = \mathcal{Y} \backslash S$. Take any two subsets A and B of the set of strategy profiles $\mathcal{Y}$. We need to show that under the assumptions of lemma 2 we have

$$\mu(A \cup B) + \mu(A \cap B) \leq \mu(A) + \mu(B). \tag{A2}$$

If $A \cap S \neq \varnothing$, we have $\mu(A) = \mu(A \cup B) = 1$ and $\mu(A \cap B) \leq \mu(B)$, so that equation (A2) holds. If $B \cap S \neq \varnothing$, we have $\mu(B) = \mu(A \cup B) = 1$ and $\mu(A \cap B) \leq \mu(A)$, so that equation (A2) holds. Finally, if $A \subseteq S^c$ and $B \subseteq S^c$, we have

by assumption $\mu(A) = \sigma(A)$, $\mu(B) = \sigma(B)$, $\mu(A \cap B) = \sigma(A \cap B)$, and $\mu(A \cup B) = \sigma(A \cup B)$. Hence, equation (A2) holds by the inclusion–exclusion principle. The result follows.

A.6. Proof of Theorem 5

By Proposition 2 , the identified set is characterized as the set of parameter values θ such that for all functions f on $\mathcal{Y}$, $E_P(f(Y)|X) \leq \int \left(\max_{\sigma \in G(\varepsilon|X;\theta)} E_\sigma(f(Y)) \right) dv(\varepsilon|X;\theta)$. By Schmeidler (1986), the functional

$$\tilde{\mathcal{L}}: f \mapsto \int \left(\max_{\sigma \in G(\varepsilon|X;\theta)} E_\sigma(f(Y)) \right) dv(\varepsilon|X;\theta)$$

is the Choquet integral associated with the Choquet capacity $\mathcal{L}: B \mapsto \mathcal{L}(B|X;\theta) = \tilde{\mathcal{L}}(1_B|X;\theta)$ because the latter is submodular, as a mixture of submodular capacities, and (dropping the dependence on X and θ from notation) the Choquet integral satisfies the following two properties:

- Monotonicity: If $f(y) \leq t(y)$ for all $y \in \mathcal{Y}$, then $\tilde{\mathcal{L}}(f) \leq \tilde{\mathcal{L}}(t)$,
- Comonotonic additivity: If f and t are comonotonic, *i.e.* if $(f(y_1) - f(y_2))(t(y_1) - t(y_2)) \geq 0$ for all $y_1, y_2 \in \mathcal{Y}$, then $\tilde{\mathcal{L}}(f + t) = \tilde{\mathcal{L}}(f) + \tilde{\mathcal{L}}(t)$.

Note that submodularity of $\mathcal{L}: B \mapsto \mathcal{L}(B) = \tilde{\mathcal{L}}(1_B)$ (i.e. $\mathcal{L}(A \cap B) + \mathcal{L}(A \cup B) \leq \mathcal{L}(A) + \mathcal{L}(B)$ for all $A, B \subseteq \mathcal{Y}$) is equivalent to convexity of the Choquet integral $\tilde{\mathcal{L}}$ (see Proposition 3 of Schmeidler, 1986 or Theorem 6.13 p. 212 of Fujishige, 2005). By definition of the Choquet integral, $\tilde{\mathcal{L}}$ can be written

$$\tilde{\mathcal{L}}(f) = \int_0^\infty \mathcal{L}(f \geq x) dx + \int_{-\infty}^0 (\mathcal{L}(f \geq x) - 1) dx.$$

Whenever a probability P on $\mathcal{Y}$ satisfies $P(B) \leq \mathcal{L}(B)$ for all $B \in \subseteq \mathcal{Y}$, it follows that

$$E_P(f(Y)) = \sum_{y \in \mathcal{Y}} f(y) P(y)$$

$$= \int_0^\infty P(f \geq x) dx + \int_{-\infty}^0 (P(f \geq x) - 1) dx$$

$$\leq \int_0^\infty \mathcal{L}(f \geq x) dx + \int_{-\infty}^0 (\mathcal{L}(f \geq x) - 1) dx$$

$$= \tilde{\mathcal{L}}(f).$$

The converse is immediately seen by taking the indicator functions for f, and the result follows.

APPENDIX B COMPLEMENTS

B.1. Generalized likelihood in the example 1.2.2

- $\mathcal{L}(\{(0,1)\}|\theta) = v(\varepsilon_1 < \theta, \varepsilon_2 > -2\theta)$.
- $\mathcal{L}(\{(1,0)\}|\theta) = v(\varepsilon_2 < \theta, \varepsilon_1 > -2\theta)$.
- $\mathcal{L}(\{(0,1),(1,0)\}|\theta) = 1 - v(\varepsilon_1 < -2\theta, \varepsilon_2 < -2\theta) - v(\varepsilon_1 > \theta, \varepsilon_2 > \theta)$.
- $\mathcal{L}(\{(0,1),(1,0),(1,1)\}|\theta) = 1 - v(\varepsilon_1 < -2\theta, \varepsilon_2 < -2\theta)$.
- $\mathcal{L}(\{(0,1),(1,0),(0,0)\}|\theta) = 1 - v(\varepsilon_1 > \theta, \varepsilon_2 > \theta)$.
- $\mathcal{L}(\{(0,0)\}|\theta) = \mathbb{E}_v[(\theta - \varepsilon_2)(\theta - \varepsilon_1)/(9\theta^2)1\{\varepsilon \in [-2\theta, \theta]^2\}] + v(\varepsilon_1 < -2\theta, \varepsilon_2 < -2\theta)$.
- $\mathcal{L}(\{(1,1)\}|\theta) = \mathbb{E}_v[(2\theta + \varepsilon_2)(2\theta + \varepsilon_1)/(9\theta^2)1\{\varepsilon \in [-2\theta, \theta]^2\}] + v(\varepsilon_1 > \theta, \varepsilon_2 > \theta)$.
- $\mathcal{L}(\{(1,1),(0,0)\}|\theta) = \mathcal{L}(\{(1,1)\}|\theta) + \mathcal{L}(\{(0,0)\}|\theta)$.
- $\mathcal{L}(\{(0,0),(0,1)\}|\theta) = \mathcal{L}(\{(0,1)\}|\theta) + v(\varepsilon_1 < -2\theta, \varepsilon_2 < -2\theta)$.
- $\mathcal{L}(\{(0,0),(1,0)\}|\theta) = \mathcal{L}(\{(1,0)\}|\theta) + v(\varepsilon_1 < -2\theta, \varepsilon_2 < -2\theta)$.
- $\mathcal{L}(\{(1,1),(0,1)\}|\theta) = \mathcal{L}(\{(0,1)\}|\theta) + v(\varepsilon_1 > \theta, \varepsilon_2 > \theta)$.
- $\mathcal{L}(\{(1,1),(1,0)\}|\theta) = \mathcal{L}(\{(1,0)\}|\theta) + v(\varepsilon_1 > \theta, \varepsilon_2 > \theta)$.
- $\mathcal{L}(\{(1,1),(0,0),(0,1)\}|\theta) = \mathcal{L}(\{(0,1)\}|\theta) + v(\varepsilon_1 > \theta, \varepsilon_2 > \theta) + v(\varepsilon_1 < -2\theta, \varepsilon_2 < -2\theta)$.
- $\mathcal{L}(\{(1,1),(0,0),(1,0)\}|\theta) = \mathcal{L}(\{(1,0)\}|\theta) + v(\varepsilon_1 > \theta, \varepsilon_2 > \theta) + v(\varepsilon_1 < -2\theta, \varepsilon_2 < -2\theta)$.

REFERENCES

ACKERBERG, D., BENKARD, L., BERRY, S. and PAKES, A. (2007), "Econometric Tools for Analyzing Market Outcomes", *Handbook of Econometrics*, **6A**, 4171–4276.

ANDREWS, D., BERRY, S. and JIA, P. (2003), "Placing Bounds on Parameters of Entry Games in the Presence of Multiple Equilibria" (unpublished manuscript).

ARTSTEIN, Z. (1983), "Distributions of Random Sets and Random Selections", *Israel Journal of Mathematics*, **46**, 313–324.

BAJARI, P., HONG, H. and RYAN, S. (2010), "Identification and Estimation of Discrete Games of Complete Information", *Econometrica*, **78**, 1529–1568.

BERESTEANU, A., MOLCHANOV, I. and MOLINARI, F. (2009), "Sharp Identification Regions in Models with Convex Predictions: Games, Individual Choice, and Incomplete Data" (Working Paper No. CWP27/09, Cemmap).

BERRY, S. (1992), "Estimation of a Model of Entry in the Airline Industry", *Econometrica*, **60**, 889–917.

BERRY, S. and TAMER, E. (2006), Identification in Models of Oligopoly Entry", in Blundell, R., Newey, W. and Persson, T. (eds) *Advances in Economics and Econometrics* (Cambridge: Cambridge University Press) 46–85.

BISIN, A., MORO, A. and TOPA, G. (2002), "Empirical Content of Models with Multiple Equilibria" (unpublished manuscript).

BJORN, P. and VUONG, Q. (1985), "Simultaneous Equations Models for Dummy Endogenous Variables" (Working Paper No. 537, Caltech).

BRESNAHAN, T. and REISS, P. (1990), "Entry in Monopoly Markets", *Review of Economic Studies*, **57**, 531–553.

BROCK, B. and DURLAUF, S. (2001), "Discrete Choice with Social Interactions", *Review of Economic Studies*, **68**, 235–265.

CHERNOZHUKOV, V., HONG, H. and TAMER, E. (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.

CHRISTAKIS, N. A., FOWLER, J. H., IMBENS, G. W. and KALYANARAMAN, K. (2010), "An Empirical Model for Strategic Network Formation" (Working Paper No. w16039, NBER).

CILIBERTO, F. and TAMER, E. (2009), "Market Structure and Multiple Equilibria in Airline Markets", *Econometrica*, **70**, 1791–1828.

DAGSVIK, J. and JOVANOVIC, B. (1994), "Was the Great Depression a Low-Level Equilibrium", *European Economic Review*, **38**, 1711–1729.

DUKE (1999), "National Long Term Care Survey", Public use data set produced and distributed by the Duke University Center for Demographic Studies with funding from the National Institute on Aging under Grant No. U01-AG007198.

ECHENIQUE, F. and EDLIN, A. (2004), "Mixed Equilibria in Games of Strategic Complementarities", *Journal of Economic Theory*, **22**, 33–44.

ECHENIQUE, F., LEE, S. and SHUM, M. (2010), "Aggregate Matchings" (Working Paper No. 1316, Caltech SS).

EKELAND, I., GALICHON, A. and HENRY, M. (2010), "Optimal Transportation and the Falsifiability of Incompletely Specified Economic Models", *Economic Theory*, **42**, 355–374.

ENGERS, M. and STERN, S. (2002), "Family Bargaining and Long Term Care", *International Economic Review*, **43**, 73–114.

FUJISHIGE, S. (2005), *Submodular Functions and Optimization* (Amsterdam, The Netherlands: Elsevier).

GALICHON, A. and HENRY, M. (2006), "Inference in Incomplete Models" (unpublished, available from SSRN at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=886907).

GALICHON, A. and HENRY, M. (2009), "A Test of Non-Identifying Restrictions and Confidence Regions for Partially Identified Parameters", *Journal of Econometrics*, **152**, 186–196.

GOURIÉROUX, C., LAFFONT, J.-J. and MONFORT, A. (1980), "Coherency Conditions in Simultaneous Linear Equations Models with Endogenous Switching Regimes", *Econometrica*, **48**, 675–695.

HECKMAN, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equations System", *Econometrica*, **46**, 931–960.

HECKMAN, J., SMITH, J. and CLEMENTS, N. (1997), "Making the Most Out of Programme Evaluation and Social Experiments: Accounting for Heterogeneity in Programme Impacts", *Review of Economic Studies*, **64**, 487–535.

HENRY, M., MÉANGO, R. and QUEYRANNE, M. (2010), "Bootstrap Inference in Normal Form Games" (unpublished, available at SSRN: http://ssrn.com/abstract=1758743).

JOVANOVIC, B. (1989), "Observable Implications of Models with Multiple Equilibria", *Econometrica*, **57**, 1431–1437.

MANSKI, C. (1990), "Nonparametric Bounds on Treatment Effects", *American Economic Review*, **80**, 319–323.

MILGROM, P. and ROBERTS, J. (1990), "Rationalizability, Learning and Equilibrium in Games with Strategic Complementarities", *Econometrica*, **58**, 1255–1277.

MOULIN, H. (1995), *Cooperative Microeconomics* (Princeton: Princeton University Press).

SCHMEIDLER, D. (1986), "Integral Representation Without Additivity", *Proceedings of the American Mathematical Society*, **97**, 255–261.

SHAPLEY, L. (1971), "Cores of Convex Games", *International Journal of Game Theory*, **1**, 11–26.

STRASSEN, V. (1965), "The Existence of Probability Measures with Given Marginals", *Annals of Mathematical Statistics*, **36**, 423–439.

SWEETING, A. (2009), "Coordination Games, Multiple Equilibria and the Timing of Radio Commercials", *RAND Journal of Economics*, **40**, 710–742.

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: Princeton University Press).

VIVES, X. (1990), "Nash Equilibrium with Strategic Complementarities", *Journal of Mathematical Economics*, **19**, 305–321.