Quantitative Economics 6 (2015), 499–529 1759-7331/20150499

Combinatorial approach to inference in partially identified incomplete structural models

Marc Henry
Penn State

Romuald Méango
IFO Institute

Maurice Queyranne
Sauder School, UBC and CORE

We propose a computationally feasible inference method in finite games of complete information. Galichon and Henry (2011) and Beresteanu, Molchanov, and Molinari (2011) show that the empirical content in such models is characterized by a collection of moment inequalities whose number increases exponentially with the number of discrete outcomes. We propose an equivalent characterization based on classical combinatorial optimization methods that allows the construction of confidence regions with an efficient bootstrap procedure that runs in linear computing time. The method can be applied to the empirical analysis of cooperative and noncooperative games, instrumental variable models of discrete choice, and revealed preference analysis. We propose an application to the determinants of long term elderly care choices.

Keywords. Incomplete structural models, multiple equilibria, partial identification, sharp bounds, confidence regions, max-flow–min-cut, functional quantile, bootstrap, elderly care.

JEL classification. C13, C72.

Marc Henry: marc.henry@psu.edu
Romuald Méango: meango@ifo.de
Maurice Queyranne: maurice.queyranne@sauder.ubc.ca
Parts of this paper were written while Henry was visiting the Graduate School of Economics at the University of Tokyo; he gratefully acknowledges the CIRJE for its hospitality and support. Financial support from SSHRC Grants 410-2010-242 and 435-2013-0292 and NSERC Grant 356491-2013 and from the Leibniz Association (SAW-2012-ifo-3) is gratefully acknowledged. The authors thank Victor Chernozhukov, Russell Davidson, Alfred Galichon, Silvia Gonçalves, Hidehiko Ichimura, and Ismael Mourifié for helpful discussions, Rosa Matzkin, Jean-Marc Robin, and six anonymous referees for helpful suggestions, David Straley for providing the data, and Daniel Stubbs for computing assistance. Comments from seminar audiences in Cambridge, the Bank of Japan, Harvard–MIT, Hitotsubashi, UBC Sauder School, UCL, U-Kyoto, U-Tokyo, Rochester, Vanderbilt, and participants at the Inference in Incomplete Models Conference in Montreal, the First Workshop on New Challenges in Distributed Systems in Valparaiso, and the Second Workshop on Optimal Transportation and Applications to Economics in Vancouver are also gratefully acknowledged.

DOI: 10.3982/QE377

Introduction

With the conjoined advent of powerful computing capabilities and rich data sets, the empirical evaluation of complex structural models with equilibrium data is becoming prevalent, particularly in the analysis of social networks and industrial organization. However, in such models, multiple equilibria are the norm rather than the exception. Though multiplicity of equilibria and identifiability of the model's structural parameters are conceptually distinct, the former often leads to a failure of the latter, thereby invalidating traditional inference methods. This is generally remedied by imposing additional assumptions to achieve identification, such as imposing an equilibrium selection mechanism or a refinement of the equilibrium concept. Manski (1989) and Jovanovic (1989) were among the first to advocate a new inference approach that dispenses with identification assumptions and delivers confidence regions for partially identified structural parameters. A large literature has developed on the general problem of inference on partially identified parameters defined as minimizers of objective functions or more specifically as solutions to moment inequality restrictions, following the seminal work of Chernozhukov, Hong, and Tamer (2007).

In structural estimation using equilibrium conditions, the partial identification approach was initially applied, as in Haile and Tamer (2003), to achieve simple and robust inference from implications of the model in the form of a small number of moment inequalities. This partial identification approach was applied to inference in games by Andrews, Berry, and Jia (2003), Pakes, Porter, Ho, and Ishii (2015), and Ciliberto and Tamer (2009), among others. However, the approach of the latter papers to inference in games brings only part of the empirical content of the model to bear on the estimation, resulting in unnecessary loss of informativeness. In models with multiple equilibria and no additional prior information, nothing is known of the equilibrium selection mechanism. If a particular equilibrium selection mechanism is posited, the model likelihood can be derived and inference can be based on it. Jovanovic (1989) characterizes compatibility of an economic structure with the true data generating process as the existence of some (unknown) equilibrium selection mechanism, for which the likelihood is equal to the true data generating mechanism. Berry and Tamer (2006) define the identified set as the collection of structural parameter values for which the structure is compatible with the data generating mechanism in the sense of Jovanovic (1989). This definition of the identified set is not directly conducive to inference, as it involves an infinite dimensional (nuisance) parameter (the equilibrium selection mechanism). However, in the case of finite noncooperative games of complete information, Galichon and Henry (2011) and Beresteanu, Molchanov, and Molinari (2011) show equivalence of the Jovanovic (1989) definition with a system of inequalities. Hence, they show that the empirical content of such models is characterized by a finite collection of moment inequalities.

A large literature has developed on inference in moment inequality models since the seminal contribution of Chernozhukov, Hong, and Tamer (2007). We discuss and review it in Section 3. However, a major challenge in the framework of this paper is that the number of inequalities characterizing the empirical content of the model grows exponentially with the number of equilibrium strategy profiles. Hence the combinatorial optimization approach that we propose in this paper is, to the best of our knowledge, the only computationally feasible inference procedure for empirically relevant incomplete economic structures. The growing literature on "inference with many moment inequalities" addresses theoretical issues relating to the case where the number of inequalities grows with sample size and does not alleviate the computational burden mentioned here. This problem of exponential complexity goes a long way toward explaining the dearth of empirical studies using partial identification in such models. However, abandoning this partial identification approach would mean abandoning robust inference not only in noncooperative games of perfect information, but also in large classes of models that share exactly the same feature and fall into the framework of this paper. They include cooperative games, such as matching games and network formation games, revealed preference analysis of spacial preferences, and matching markets and instrumental variable models of discrete choice.

The objective of this paper is to propose a combinatorial solution to this problem, where the number of inequality restrictions grows exponentially with the number of strategy profiles or discrete outcomes. Ekeland, Galichon, and Henry (2010) have shown that generic partial identification problems can be formulated as optimal transportation problems. Developing ideas in Galichon and Henry (2011), we exploit the special structure of discrete choice problems and show that correct specification can be formulated as a problem of maximizing flow through a network, and that the identified set can be obtained from the max-flow–min-cut theorem. The dual problems of maximizing flow through a network and finding a minimum capacity cut are classics in combinatorial optimization and operations research, with applications in many areas such as traffic, communications, routing, and scheduling; see, for example, Schrijver (2004) for the theory and history, and Ahuja, Magnanti, and Orlin (1993) for numerous applications. To our knowledge this is the first application of the max-flow–min-cut theorem to statistical inference for equilibrium models. We apply this powerful combinatorial method to the problem of constructing confidence regions for structural parameters. We construct a functional quantile for the bootstrap process using a linear computing time algorithm and replace the unknown empirical process by this quantile in the system of moment inequalities to obtain the least relaxation of the moment inequalities, hence maximum informativeness, while controlling the confidence level of the covering region. Since the procedure involves bootstrapping the empirical process only, it does not suffer from the problems of bootstrap validity in partially identified models described in Chernozhukov, Hong, and Tamer (2007) and Bugni (2010). We illustrate and assess our procedure on a very simple full information game with two players and three strategies, easily derived equilibria, and yet a large number of inequalities to characterize its empirical content (namely 127). We simulate the game under a variety of parameter values and assumptions on the data generating process and with explanatory variables. Finally, we illustrate the approach, the procedure, and the interpretation of results on an application to the determinants of long term elderly care choices of American families.

In summary, the main contributions of this paper are as follows:

1. We present a new approach to inference in incomplete structural models.

2. We provide a simplified and insightful new proof for a characterization of the identified set.

3. We present a computationally efficient, combinatorial procedure that allows feasible inference in empirically relevant incomplete structural models. We demonstrate its practical efficiency in extensive simulations of a simple game.

4. We apply this methodology to an empirical example and demonstrate the type of econometric analysis and insights that it allows.

The paper is organized as follows. The next section introduces the general framework and the object of study. Section 2 derives the characterization of the identified set with the min-cut–max-flow theorem. Section 3 describes the combinatorial procedure to construct the confidence region efficiently. Section 4 contains the simulation evidence and Section 5 contains the empirical application. The last section concludes. Proofs are collected in the Appendix. Relevant definitions and theorems pertaining to the combinatorial optimization notions, and extended empirical results and discussions are collected in supplementary files on the journal website, http://qeconomics.org/supp/377/supplement.pdf and http://qeconomics.org/supp/377/code_and_data.zip.

1. Analytical framework

1.1 *Model specification*

We consider the model specification

$$Y \in G(X, \varepsilon; \theta), \tag{1.1}$$

where Y is an observable outcome variable, which takes values in a finite set $\mathcal{Y} = \{y_1, \dots, y_K\}$, X is a vector of exogenous explanatory variables with domain $\mathcal{X}$, ε is a vector of unobservable heterogeneity variables with domain $\Xi \subset \mathbb{R}^l$, and $\theta \in \Theta \subset \mathbb{R}^d$ is a vector of unknown parameters. Finally, $G : \mathcal{X} \times \Xi \rightrightarrows \mathcal{Y}$ is a set-valued mapping parameterized by θ. The random elements X, Y, and ε are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The sample consists of n observational units $i = 1, \dots, n$, which are independent and identical in distribution. To each unit i is attached a vector $(Y_i, X_i, \varepsilon_i)$, only the first two elements of which shall be observed. For each potential outcome $y \in \mathcal{Y}$, we denote by $P(y|X)$ the conditional probability $\mathbb{P}(Y = y|X)$. If Z is a subset of $\mathcal{Y}$, $P(Z|X)$ will denote $\sum_{y\in Z} P(y|X)$. It is important to emphasize here the fact that $P(\cdot|X)$ denotes the true outcome data generating process, which is unknown, but can be estimated from the data. It is not a function of the structural parameter vector and cannot be construed as the likelihood from the model. The vector of unobservable variables ε in the economic structure has conditional cumulative distribution function $F(\varepsilon|X; \theta)$ for some known function F parameterized by θ (the same notation is used for the parameters of the model correspondence and for the parameters of the error distribution to indicate that they may have common components). The economic structure is summarized by the multivalued mapping G. A special case of specification (1.1) arises when G is a function, in which case model (1.1) is a nonlinear nonseparable single equation discrete choice model as in Chesher (2010). Here, however, we entertain the possibility of G

having multiple values arising from multiple equilibria, data censoring, or endogeneity. The mapping G is entirely given by the economic structural model, up to an unknown parameter vector θ.

The analytical framework, concepts, and procedures proposed throughout the paper will be illustrated and discussed with the following simple example.

Example 1 (Partnership game). Our example is a simple noncooperative full information game of complementarities.

- *Strategies*. There are two players, who simultaneously decide, whether to invest strongly (strategy H), weakly (strategy L), or not at all (strategy O) in a partnership.
- *Payoffs*. Players pay a cost $c \geq 0$ (respectively $2c$) for a weak (respectively strong) investment. Benefits that accrue to players depend on the overall level of investment in the partnership and explanatory variables J_i, $i = 1, 2$, where $J_i = 1$ if player i is female and is zero otherwise. The benefits for player i are $3c(1 + \beta J_i)$ in case both players invest strongly, $2c(1 + \beta J_i)$ in case one player invests weakly and the other invests strongly, and $c(1 + \beta J_i)$ in case both players invest weakly. Finally player i also experiences an idiosyncratic random participation payoff ε_i, $i = 1, 2$, with a density with respect to Lebesgue measure. The payoff matrix for the game is given in Table 1.
- *Equilibrium concept*. We assume that outcomes are Nash equilibria in pure strategies. Other equilibrium concepts could be entertained, in particular with mixed strategies, as will be discussed in Section 3.1 and illustrated in the empirical application.

The strategies, payoffs, and equilibrium concept together define the economic structure. Element Y is an observed equilibrium strategy profile; $J = (J_1, J_2)$ is also observed by the analyst. The idiosyncratic participation benefit $\varepsilon = (\varepsilon_1, \varepsilon_2)$ is not observed, but it is common knowledge to the players. The structural parameter vector is $\theta = (c, \beta)$. The equilibrium correspondence, i.e., the set of equilibria for each value of ε, J and θ, can be easily derived and defines the multivalued mapping G in model specification (1.1), which is represented in the $(\varepsilon_1, \varepsilon_2)$ space in Figure 1 for the case $\beta = 0$. Since we assume that ε has absolutely continuous distribution with respect to Lebesgue measure, we do

Table 1. Payoff matrix for the partnership game.

	Player 2		
Player 1	H	L	O
H	$3c(1+\beta J_i) - 2c + \varepsilon_1$ $3c(1+\beta J_i) - 2c + \varepsilon_2$	$2c(1+\beta J_i) - 2c + \varepsilon_1$ $2c(1+\beta J_i) - c + \varepsilon_2$	$-2c + \varepsilon_1$ 0
L	$2c(1+\beta J_i) - c + \varepsilon_1$ $2c(1+\beta J_i) - 2c + \varepsilon_2$	$c(1+\beta J_i) - c + \varepsilon_1$ $c(1+\beta J_i) - c + \varepsilon_2$	$-c + \varepsilon_1$ 0
O	0 $-2c + \varepsilon_2$	0 $-c + \varepsilon_2$	0 0

Note: In each cell, the top expression is player 1's payoff and the bottom term is player 2's payoff.

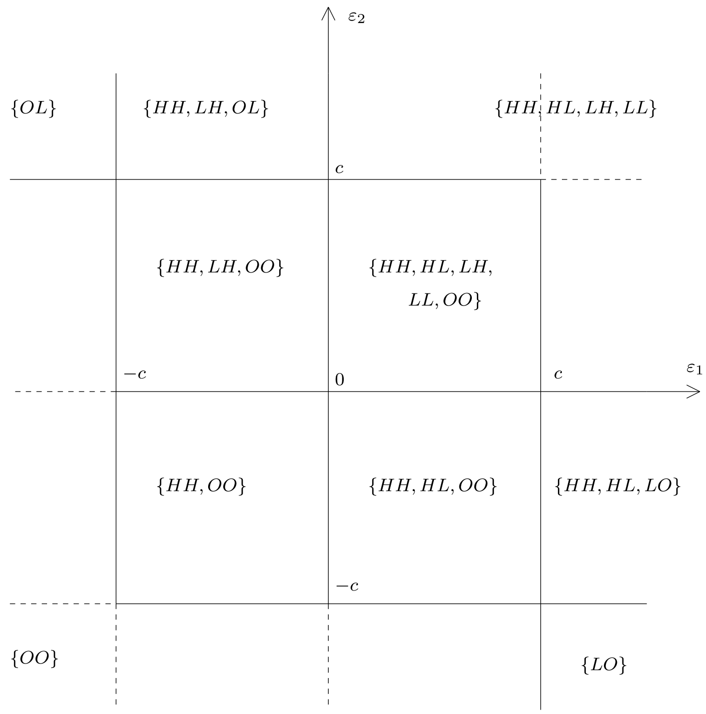

FIGURE 1. Representation of the equilibrium correspondence $G(J, \varepsilon; \theta)$ in the $(\varepsilon_1, \varepsilon_2)$ space, when $\beta = 0$.

not include zero probability predictions, such as $\{OO, OL\}$ when $\varepsilon_2 = c$ and $\varepsilon_1 < -c$, for instance.

1.2 *Object of inference*

Model (1.1) has the fundamental feature that G is multivalued (because of multiple equilibria in the example above, for instance). For a given value of (X, ε, θ), the model predicts a set of possible outcomes $G(X, \varepsilon; \theta)$. Only one of them, namely Y, is actually realized, but the economic structure is silent about how that particular Y was selected among $G(X, \varepsilon; \theta)$. In other words, the economic structure holds no information about the equilibrium selection mechanism. If the true (unknown) equilibrium selection mechanism is denoted $\pi_0(y|\varepsilon, X)$, which is a probability on $G(X, \varepsilon; \theta)$, then the likelihood of observation y can be written

$$L(\theta|y, X) = \int_{\Xi} \pi_0(y|\varepsilon, X)\, dF(\varepsilon|X; \theta)$$

and the true parameter θ_0 satisfies

$$P(y|X) = \int_{\Xi} \pi_0(y|\varepsilon, X)\, dF(\varepsilon|X; \theta_0), \quad X\text{-a.s. for all } y. \tag{1.2}$$

Jovanovic (1989) points out that the incomplete model (incomplete because the equilibrium selection is not modeled) is compatible with the true data generating process $P(\cdot|X)$ if and only if there exists a (generally nonunique) equilibrium selection mechanism π_0 such that (1.2) holds. The identified set is then defined as the set Θ_I of parameter values θ such that model (1.1) is compatible in the sense of Jovanovic (1989).

Definition 1 (Identified set). The identified set Θ_I is the set of parameter values $\theta \in \Theta$ such that there exists a probability kernel $\pi(\cdot|\varepsilon, X)$ with support $G(X, \varepsilon; \theta)$ for which (1.2) holds.

The identified set is empty if no value of the parameter can rationalize the data generating process, in which case the structural model is misspecified. The identified set is a singleton in case of point identification, which occurs if G happens to be single valued under the true parameter values (in case $c = \beta = 0$ in Example 1) or in very special cases under large support assumptions on X, as in Tamer (2003). The identified set is totally uninformative, i.e., $\Theta_I = \Theta$, in case the model has no empirical content (if, for instance, $G(X, \varepsilon; \theta_0)$ contains all selected outcome values for almost all ε at the true value θ_0).

1.3 *Applications of the framework*

Specification (1.1), hence the inference procedure presented in this paper, has a wide range of applications. Some of the most compelling applications are the empirical analysis of games, instrumental variable models of discrete choice with endogeneity, and revealed preference analysis.

- *Empirical analysis of games*. As illustrated in Example 1, model (1.1) applies to the empirical analysis of noncooperative games of perfect information (normal form games). They include the classic entry game of Bresnahan and Reiss (1990) and Berry (1992) as well as the social interaction game of Soetevent and Kooreman (2007). Noncooperative games of private information make for a less compelling application of this framework as point identification conditions are more easily derived and justified than in their perfect information counterparts (see, for instance, Aradillas-Lopez (2010) and Bajari, Hahn, Hong, and Ridder (2011) for a discussion). Finally, some cooperative games can be analyzed and estimated within the present framework, in particular matching and social network formation games, where the equilibrium correspondence is characterized by pairwise stability. Uetake and Watanabe (2013) present an empirical analysis of entry by merger, where the present inference procedure can be applied.

- *Discrete choice models with endogeneity*. Chesher, Rosen, and Smolinski (2013) study instrumental variable (IV) models of discrete choice, such as the IV extension of McFadden's conditional logit model with endogeneity, considered in their Example 1. When the endogenous and exogenous regressors have discrete support, that model falls within the framework of (1.1). Our paper provides a computationally attractive inference procedure in such settings.

- *Revealed preference analysis*. Henry and Mourifié (2013) apply the inference procedure proposed here to analyze voting behavior from a revealed preference standpoint. The same approach can be applied to revealed preference testing in matching markets as in Echenique, Lee, Shum, and Yenmez (2013) or the revealed preference approach to games taken in Pakes et al. (2015).

2. Operational characterization of the identified set

As noted in Berry and Tamer (2006), Definition 1 is not an operational definition of the identified set, as it includes the equilibrium selection mechanism as an infinite dimensional parameter. Galichon and Henry (2006, 2011) and Beresteanu, Molchanov, and Molinari (2011) show a characterization of the identified set with a finite collection of moment inequalities. In this section, we give an equivalent characterization of the identified set, whose proof is much simpler and relies on the min-cut–max-flow theorem, which brings classical efficient combinatorial optimization methods to bear on the problem. This will prove crucial for the feasibility of the inference procedure in realistic and relevant empirical examples.

First, we set out the main heuristic for the operational characterization of the identified set. Model specification (1.1) is a discrete choice model, hence the set $\mathcal{Y}$ of outcomes is finite and the correspondence G takes only a finite number of values, which we label $\mathcal{U} = \{u_1, \dots, u_J\}$. Each u is a set (possibly singleton) of outcomes in $\mathcal{Y}$. Because the model is incomplete, it does not predict the probabilities of individual outcomes in $\mathcal{Y}$, but it predicts the probability of each combination of equilibria listed in $\mathcal{U}$. We denote these probabilities $Q(u|X;\theta)$ as they depend on the structural parameter value.

Definition 2 (Predicted probabilities). For each $u \in \mathcal{U}$, we define $Q(u|X;\theta) := \mathbb{P}(G(X,\varepsilon;\theta) = u|X,\theta)$. If V is a subset of $\mathcal{U}$, we write $Q(V|X;\theta) = \sum_{u\in V} Q(u|X;\theta)$.

In most applications, it will be difficult to obtain closed forms for $Q(u|X;\theta)$. However, ε can be randomly generated. Given a sample $(\varepsilon^r)_{r=1,\dots,R}$ of simulated values, $Q(u|X;\theta)$ can be approximated by $\sum_{r=1}^{R} 1\{u = G(X,\varepsilon^r;\theta)\}/R$. Bajari, Hong, and Ryan (2010) propose an importance sampling procedure that greatly reduces the computational burden of this stage of the inference. The simulation procedure is now standard and cannot be avoided if one wishes, as we do here, to exhaust the empirical content of the structural model.

Example 1 (Continued). In the partnership example with $\beta = 0$, the model predicts the following values for the equilibrium correspondence: $\mathcal{U} = \{\{OL\}, \{LH, OL, HH\}, \{HH, LH, OO\}, \{OO\}, \{HH, OO\}, \{HH, LL, HL, LH\}, \{HH, LL, OO, HL, LH\}, \{HH, OO, HL\}, \{HH, HL, LO\}, \{LO\}\}$. The set $\mathcal{Y}$ of equilibrium strategy profiles (that may be observed) is $\{HH, HL, LH, LL, LO, OL, OO\}$ with 7 elements, while the set of predicted collections of equilibria (possible values of the equilibrium correspondence) $\mathcal{U}$ has 10 elements. The predicted probabilities can be computed in the following way. For instance, $Q(\{OL\}|c) = \mathbb{P}(\varepsilon_1 \le -c \text{ and } \varepsilon_2 \ge c)$ and $Q(\{HH, LH, OL\}|c) = \mathbb{P}(-c \le \varepsilon_1 \le 0 \text{ and } \varepsilon_2 \ge c)$, and the remaining 8 probabilities are determined similarly from Figure 1.

The model structure imposes a set of restrictions on the relation between the predicted probabilities of equilibrium combinations and the true probabilities of outcomes. For instance, the predicted probability $Q(\{HH, LH, OL\}|X; \theta)$ in the above example cannot be larger than the sum $P(HH) + P(LH) + P(OL)$ of probabilities of occurrence of each individual equilibrium in u, since Y is either HH, LH, or OL, when $u = \{HH, LH, OL\}$ is predicted. More generally, since P and Q are the marginals of the joint distribution of (Y, U) given X, we must have for all $u \in \mathcal{U}$,

$$\begin{aligned} Q(u|X;\theta) &= \sum_{y\in u} \mathbb{P}(Y = y \text{ and } U = u|X;\theta) \\ &\leq \sum_{y\in u} \mathbb{P}(Y = y|X;\theta) = \sum_{y\in u} P(y|X). \end{aligned} \tag{2.1}$$

Note that $Q(u|X; \theta)$ may be strictly smaller than $\sum_{y\in u} P(y|X)$ when some outcome $y \in u$ also belongs to other combinations u' that may arise under different values of ε, as its (marginal) probability $P(y|X)$ must then be split between $Q(u|X; \theta)$ and the probabilities $Q(u'|X; \theta)$ of such other combinations $u' \in \mathcal{U}$ containing y. However, inequalities (2.1) do not exhaust the information in the structure. They may all be satisfied and yet the structure may be incompatible with the data generating process as the following example shows. Hence more inequalities will be needed as derived below.

Example 1 (Continued). In the partnership example with $\beta = 0$, suppose that the true equilibrium selection mechanism is such that $Q(\{OL\}|\theta) = P(OL) > 0$ and $Q(\{HH, LH, OL\}|\theta) = P(HH) + P(LH) + P(OL)$. Then $Q(\{OL\} \cup \{HH, LH, OL\}|\theta) = Q(\{OL\}|\theta) + Q(\{HH, LH, OL\}|\theta) > P(HH) + P(LH) + P(OL)$ so that $\theta \notin \Theta_I$.

Extending this observation, consider a subset $V \subseteq \mathcal{U}$ and define

$$V^{\cup} := \{y \in Y : y \in u \text{ for some } u \in V\} = \bigcup_{u\in V} u.$$

Then we must have

$$\begin{aligned} Q(V|X;\theta) &= \sum_{u\in V}\sum_{y\in u} \mathbb{P}(Y = y \text{ and } U = u|X;\theta) \\ &= \sum_{y\in V^{\cup}}\sum_{u\in V:\, y\in u} \mathbb{P}(Y = y \text{ and } U = u|X;\theta) \\ &\leq \sum_{y\in V^{\cup}}\sum_{u\in \mathcal{U}} \mathbb{P}(Y = y \text{ and } U = u|X;\theta) \\ &= \sum_{y\in V^{\cup}} P(y|X), \end{aligned}$$

where the inequality is again due to the fact that some $y \in V^{\cup}$ may also belong to some $u' \notin V$. Since this inequality holds for every $V \subseteq \mathcal{U}$, we must have

$$\max_{V\subseteq\mathcal{U}}\left(\sum_{u\in V} Q(u|X;\theta) - \sum_{y\in V^{\cup}} P(y|X)\right) \leq 0.$$

This inequality must also hold for every realization x of X in the domain $\mathcal{X}$ of the explanatory variables, implying that every θ in the identified set Θ_I must satisfy

$$\sup_{x\subseteq\mathcal{X}}\max_{V\subseteq\mathcal{U}}\left(\sum_{u\in V} Q(u|x;\theta) - \sum_{y\in V^{\cup}} P(y|x)\right) \leq 0.$$

So far, we have shown implications of the model. It is far more difficult to show that these implication actually exhaust all the empirical content of the model, i.e., that they involve no loss of information and constitute sharp bounds. In Theorem 1 below, we will show this with an appeal to the classical max-flow–min-cut theorem of combinatorial optimization, providing our characterization (2.2) of the identified set. We thereby provide, for the case of a finite set of possible outcomes, a new and simpler proof of the characterization of the identified set with a finite collection of inequalities, without the complicated apparatus of the theory of random sets. This allows us to emphasize the combinatorial optimization formulation of our inference problem, which is key to its tractable solution in empirically relevant instances. Theorem 1 below also provides an alternative characterization (2.3) of the identified set from the "dual" perspective of outcome subsets $Z \subseteq \mathcal{Y}$, in addition to the preceding characterization (2.2) based on combination subsets $V \subseteq \mathcal{U}$, with the notation

$$Z^{\cap} := \{u \in \mathcal{U} : u \subseteq Z\} \quad \text{and} \quad Z^{-1} := \{u \in \mathcal{U} : u \cap Z \neq \emptyset\}.$$

This alternative characterization may be useful in situations where the number of possible outcomes is much smaller than the number of possible combinations (as is the case in Example 1, where the number of equilibrium outcomes (cardinality of $\mathcal{Y}$) is 7, so the corresponding number of inequalities to be checked is $2^7 - 1 = 127$, whereas the number of predicted equilibrium combinations (cardinality of $\mathcal{U}$) is 10, so the corresponding number of inequalities to check would be $2^{10} - 1 = 1023$).

Theorem 1. *The identified set is*

$$\Theta_I = \left\{\theta \in \Theta : \sup_{x\in\mathcal{X}}\max_{V\subseteq\mathcal{U}}\left(Q(V|x;\theta) - P\left(V^{\cup}|x\right)\right) \leq 0\right\} \tag{2.2}$$

$$= \left\{\theta \in \Theta : \sup_{x\in\mathcal{X}}\max_{Z\subseteq\mathcal{Y}}\left(Q\left(Z^{\cap}|x;\theta\right) - P(Z|x;\theta)\right) \leq 0\right\}. \tag{2.3}$$

Theorem 1 gives two characterizations of the identified set Θ_I, sometimes called *sharp identified region* in the literature. The set Θ_I contains all the values of the parameter such that (1.1) holds and only such values. Moreover, all elements of Θ_I are observationally equivalent. Hence no value of the parameter vector θ contained in Θ_I

can be rejected on the basis of the information available to the analyst. Thus, Θ_I completely characterizes the empirical content of the model. A by-product of the proof is the equivalence with the characterization of the identified set derived in Galichon and Henry (2006), which we give in (2.4) in our notation.

Corollary 1. *The identified set is*

$$\Theta_I = \left\{\theta \in \Theta : \sup_{x \in \mathcal{X}} \max_{Z \subseteq \mathcal{Y}} \big(P(Z|x;\theta) - Q\big(Z^{-1}|x;\theta\big)\big) \leq 0\right\}. \tag{2.4}$$

Example 1 (Continued). To illustrate the computation of the identified set, consider the case where it is known that $\beta = 0$. Assume that the true parameter value is $c_0 = 1/4$ and the idiosyncratic shocks are independent and uniformly distributed over $[-1/2, 1/2]$. Suppose further that the true data generating process is equal to the distribution implied by a uniform equilibrium selection rule, whereby all equilibrium strategy profiles within the equilibrium correspondence are selected with equal probability. For example, when $\varepsilon_1 \geq c_0 = 1/4$ and $-1/4 = -c_0 \leq \varepsilon_2 \leq 0$, each strategy profile within the equilibrium correspondence $\{HH, HL, LO\}$ is equally likely. The probability distribution of the true data generating process in this case is defined by $P(HH) = 167/960$, $P(OO) = 191/480$, $P(OL) = P(LO) = 1/12$, $P(LL) = 19/320$, and $P(HL) = P(LH) = 97/960$. The identified set is derived as the set of values of c such that the $2^7 - 1 = 127$ inequalities of the form $P(Z) \geq Q(Z^{\cap}|c)$, all $Z \subseteq \{HH, HL, LH, LL, LO, OL, OO\}$, are satisfied. For instance, one of those inequalities is $59/320 = P(LO \text{ or } HL) \geq Q(\{LO\}|c) = (1/2 - c)^2$ if $c \leq 1/2$ and is zero otherwise. The identified set can be computed using a min-cut–max-flow algorithm, which yields $[1/2 - 1/\sqrt{12}, 1/3] \simeq [0.2113, 0.3333]$, where the lower bound of the interval happens to be the smallest value of $c > 0$ for which the inequality in (2.3) with $Z = \{LO, OL\}$ is satisfied, and the upper bound happens to be the largest value for which that with $Z = \{HH, HL, LH, LL, OO\}$ is satisfied.

As illustrated in Example 1, even in simple examples where the equilibria are very easy to compute, the exponential size of the characterization of the identified set is a severe computational burden that is best approached with combinatorial optimization techniques, as developed in the next section.

3. Confidence region

3.1 *Objective*

We now turn to the problem of inference on Θ_I based on a sample of observations $((Y_1, X_1), \ldots, (Y_n, X_n))$. We seek coverage of the identified set with prescribed probability $1-\alpha$ for some $\alpha \in (0, 1)$. It would be tempting to appeal to the large literature on inference in moment inequality models. This includes several proposals for the construction of confidence regions covering each point in the identified set, which are generally preferred due to the fact that they may be more informative (although this may sometimes be misleading as pointed out in Henry and Onatski (2012)). Such proposals include Section 5 of Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Rosen (2008), Galichon and Henry (2009), and Andrews and Soares (2010), among others. All of the papers above propose to construct confidence regions by inverting specification tests. Hence, the confidence region is constructed through a search in the parameter space, with a computationally demanding testing procedure at each parameter value visited in the search. This becomes computationally infeasible for realistic parameter vector dimensions. With a reasonably precise grid search and five parameters (for example), the number of points to be visited is in the tens of billions. If the identified set is known to be convex, the search can be conducted from a central point with a dichotomy in polar coordinates, yet it remains computationally impractical to conduct a statistical procedure for each point in the search.

Hence, each parameter value in the search must be accepted or rejected based on a deterministic criterion. This means the significance of the confidence region must be controlled independently of the parameter value. This will automatically produce a confidence region that covers the identified set. Proposals for the construction of confidence regions covering the identified set include Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2010), Galichon and Henry (2013), and Bugni (2010), among others. These can be applied to realistic models defined by a small number of moment inequality restrictions. However, a major challenge in the framework of this paper is that the number of inequalities characterizing the empirical content of the model in Theorem 1 grows exponentially with the cardinality of $\mathcal{Y}$, which in the case of games is the number of equilibrium strategy profiles (in the very simple partnership game of Example 1, the number of inequalities is 127). Hence the combinatorial optimization approach that we propose in this paper is, to the best of our knowledge, the only computationally feasible inference procedure for empirically relevant economic structures defined by finite games and other models of discrete choice with endogeneity.

Definition 3 (Confidence region). A confidence region of asymptotic level $1-\alpha$ for the identified set Θ_I is defined as a sequence of regions Θ_n, $n \in \mathbb{N}$, satisfying $\liminf_n \mathbb{P}(\Theta_I \subseteq \Theta_n) \geq 1-\alpha$.

We seek coverage of the set of values of the parameter θ such that $Q(V|x,\theta) \leq P(V^{\cup}|x)$ for all values of x and all subsets V of $\mathcal{U}$; Q is determined from the model, but P is unknown. However, if we can construct random functions $\overline{P}_n(A|x)$ that dominate the probabilities $P(A|x)$ for all values of x and all subsets A of $\mathcal{Y}$ with high probability, then, in particular, $\overline{P}_n(V^{\cup}|x) \geq P(V^{\cup}|x)$ for each x and each subset V of $\mathcal{U}$. Hence any θ satisfying $Q(V|x,\theta) \leq P(V^{\cup}|x)$ for all values of x and all subsets V of $\mathcal{U}$ also satisfies $Q(V|x,\theta) \leq \overline{P}_n(V^{\cup}|x)$ for all values of x and all subsets V of $\mathcal{U}$. There remains to control the level of confidence of the covering region, which is achieved by requiring that $\overline{P}_n$ dominate P with probability asymptotically no less than the desired confidence level. Equivalently, when working from characterization (2.4), we impose the same requirement for dominated functions $\underline{P}_n$. Hence the following assumption.

Assumption 1. *Let the random functions $A \mapsto \overline{P}_n(A|x)$, $A \subseteq \mathcal{Y}$, satisfy*

$$\liminf_n \mathbb{P}\Big(\sup_{x\in\mathcal{X}} \max_{A\subseteq\mathcal{Y}} \big[P(A|x) - \overline{P}_n(A|x)\big] \leq 0\Big) \geq 1-\alpha. \tag{3.1}$$

Suppose now a value θ_0 of the parameter vector belongs to the identified set Θ_I. Then, by Theorem 1, for all x and $V \subseteq \mathcal{U}$, $Q(V|x;\theta_0) \le P(V^{\cup}|x)$, so that with probability tending to no less than $1-\alpha$, $Q(V|x;\theta_0) \le \overline{P}_n(V^{\cup}|x)$, hence Theorem 2.

THEOREM 2 (Confidence region). *Under Assumption 1, the sets*

$$\Theta_I(\overline{P}_n) = \Big\{\theta \in \Theta : \sup_{x\in\mathcal{X}} \max_{V\subseteq\mathcal{U}} \big(Q(V|x;\theta) - \overline{P}_n\big(V^{\cup}|x\big)\big) \le 0\Big\} \tag{3.2}$$

define a confidence region of asymptotic level $1-\alpha$ *for* Θ_I (*according to Definition* 3).

Theorem 2 has the fundamental feature that it dissociates search in the parameter space (or even possibly search over a class of models) from the statistical procedure necessary to control the confidence level. The upper probabilities $\overline{P}_n$ can be determined independently of θ in a procedure that is performed once and for all using only sample information, i.e., fully nonparametrically. Once the upper probabilities are determined, probabilities Q over predicted sets of outcomes are computed for particular chosen specifications of the structure and values of the parameter, and such specifications and values are tested with inequalities defining $\Theta_n(\overline{P}_n)$. This dissociation of the statistical procedure to control confidence level from the search in the parameter space is crucial to the computational feasibility of the proposed inference procedure in realistic examples (i.e., sample sizes in the thousands, two-digit dimension of the parameter space, and two-digit cardinality of the set of observed outcomes, as in the application to teen behavior in Soetevent and Kooreman (2007), or to entry into the airline market in Ciliberto and Tamer (2009)). The latter consider only equilibria in pure strategies, as we have until now. If equilibria in mixed strategies are also considered, as in Bajari, Hong, and Ryan (2010) and in the family bargaining application below, we can appeal to results in Beresteanu, Molchanov, and Molinari (2011) and Galichon and Henry (2011). In particular, Galichon and Henry (2011) show that if the game has a *Shapley regular core* (which is the case in the family bargaining application, by Lemma 2 of Galichon and Henry (2011)), then the identified set is characterized by (2.3) of Theorem 1 with the caveat that the set function $Z \mapsto Q(Z^{\cap}|x;\theta)$ is replaced by

$$\mathcal{L}(Z|x;\theta) = \int \min_{\sigma\in G(\varepsilon|X;\theta)} \sigma(Z)\, d\nu(\varepsilon), \tag{3.3}$$

where $G(\varepsilon|X;\theta)$ is now a set of mixed strategies, i.e., a set of probabilities on the set of outcomes, as opposed to a subset of the set of outcomes. Hence the methodology can be easily adapted, as in the application of Section 5.

3.2 *Control of confidence level*

We now turn to the determination of random functions satisfying Assumption 1. First, for each $y \in \mathcal{Y}$, let $\hat{P}_n(y|x)$ be the empirical analog (or, more generally, a nonparametric estimator) of $P(y|x)$ and $\hat{P}_n(A|x) = \sum_{y\in A} \hat{P}_n(y|x)$ for each $A \subseteq \mathcal{Y}$. A simple way to achieve (3.1) is by considering the random variable

$$M_n := \sup_{x\in\mathcal{X}} \max_{A\subseteq\mathcal{Y}} \big[P(A|x) - \hat{P}_n(A|x)\big].$$

Denoting by c_n^α the $(1-\alpha)$ quantile of the distribution of M_n, we have $\mathbb{P}(M_n \leq c_n^\alpha) = 1-\alpha$ by construction, hence

$$\mathbb{P}\Big(\sup_{x\in\mathcal{X}} \max_{A\subseteq\mathcal{Y}}\big[P(A|x) - \hat{P}_n(A|x) - c_n^\alpha\big] \leq 0\Big) \geq 1-\alpha, \tag{3.4}$$

and the desired result with $\overline{P}(A|x) = \hat{P}_n(A|x) + c_n^\alpha$. However, by construction, c_n^α is independent of A and x, so that the region obtained by plugging $\overline{P}(A|x) = \hat{P}_n(A|x) + c_n^\alpha$ into (3.2) of Theorem 2 will be unnecessarily conservative. We propose, instead, to replace c_n^α by a function $\beta_n(A|x)$ of A and x, which we interpret as a *functional quantile* of the distribution of the random function $P(A|x) - \hat{P}_n(A|x)$. Analogously to (3.4), we require it to satisfy

$$\mathbb{P}\Big(\sup_{x\in\mathcal{X}} \max_{A\subseteq\mathcal{Y}}\big[P(A|x) - \hat{P}_n(A|x) - \beta_n(A|x)\big] \leq 0\Big) \geq 1-\alpha. \tag{3.5}$$

We first give a heuristic description of our proposed functional quantile before precisely spelling out the bootstrap procedure involved in approximating it. If $\mathcal{X}$ is finite, the random matrix $P(A|x) - \hat{P}_n(A|x)$ with $A \subseteq \mathcal{Y}$ and $x \in \mathcal{X}$ has a finite population of possible realizations, at most one for each possible sample draw. These realizations can be ordered according to the maximum entry in the matrix $\max_{x\in\mathcal{X}} \max_{A\subseteq\mathcal{Y}}[P(A|x) - \hat{P}_n(A|x)]$. Now take all realizations that never exceed the $(1-\alpha)$ quantile c_n^α of $\max_{x\in\mathcal{X}} \max_{A\subseteq\mathcal{Y}}[P(A|x) - \hat{P}_n(A|x)]$ and define $\overline{P}_n(A|x) = \hat{P}_n(A|x) + \beta_n(A|x)$, where $\beta_n(A|x)$ is the pointwise maximum over all realizations that never exceed c_n^α. This guarantees that the resulting confidence region obtained in (3.2) of Theorem 2 with $\overline{P}_n(A|x) = \hat{P}_n(A|x) + \beta_n(A|x)$ will be valid and will be contained in the region obtained with $\overline{P}_n(A|x) = \hat{P}_n(A|x) + c_n^\alpha$ (hence more informative than the latter). In case the conditioning variables are finitely supported, it is well known (see Singh (1981) and Bickel and Freedman (1981)) that the nonparametric bootstrap version of c_n^α is a valid approximation, which, in turn, guarantees the validity of the bootstrap procedure described below.[1] In case X has continuous components, Chernozhukov, Lee, and Rosen (2013) derive the asymptotic distribution of the supremum (over $\mathcal{X}$) of the conditional empirical process, but nothing is known of its nonparametric bootstrap approximation.

Definition 4 (Nonparametric bootstrap). Let $\mathbb{P}_n^*$ denote probability statements relative to the bootstrap distribution and conditional on the original sample $((Y_1, X_1), \dots, (Y_n, X_n))$. A bootstrap sample takes the form $((Y_1^*, X_1), \dots, (Y_n^*, X_n))$, where the explanatory variable is not resampled and for each i, Y_i^* is drawn from distribution $\hat{P}_n(\cdot|X_i)$. Let $((Y_1^b, X_1), \dots, (Y_n^b, X_n))$, $b = 1, \dots, B$, be a sequence of B bootstrapped samples. Denote by $\hat{P}_n^*(\cdot|\cdot)$ the bootstrap version (i.e., constructed identically from a bootstrap sample) of $\hat{P}_n(\cdot|\cdot)$ and let $\hat{P}_n^b$, $b = 1, \dots, B$, be its values taken on the B realized bootstrap samples. Finally, for each $A \subseteq \mathcal{Y}$ and $1 \leq j \leq n$, denote $\zeta_n^*(A|X_j) = \sum_{y\in A}[\hat{P}_n(y|X_j) - \hat{P}_n^*(y|X_j)]$ and define $\zeta_n^b(A|X_j)$ analogously.

[1] To dispell the common misconception that the use of the bootstrap is always suspect in partially identified environments, we provide a short heuristic discussion in Section S2 of the Supplement.

In the bootstrap version of the problem, we are seeking functions β_n satisfying

$$\mathbb{P}_n^*\Big(\max_{1\leq j\leq n}\max_{A\subseteq\mathcal{Y}}\big[\hat{P}_n(A|X_j)-\hat{P}_n^*(A|X_j)-\beta_n(A|X_j)\big]\leq 0\Big)\geq 1-\alpha \quad *\text{-a.s.} \tag{3.6}$$

If there was a total order on the space of realizations of ζ_n^*, we could choose β_n as the quantile of level $1-\alpha$ of the distribution of ζ_n^*. However, the $\zeta_n^*(\cdot, X_j)$'s are random functions defined on $2^{\mathcal{Y}}\times\{X_1,\dots,X_n\}$; hence there is no such total order. We propose to determine β_n from a subset of $\lceil B(1-\alpha)\rceil$ bootstrap realizations determined as follows (where $\lfloor x\rfloor$ is the largest integer below x).

Bootstrap functional quantile algorithm.

Step 1. Draw bootstrap samples $((Y_1^b, X_1),\dots,(Y_n^b, X_n))$ for $b=1,\dots,B$.

Step 2. For each $b\leq B$, $j\leq n$, and $A\subseteq\mathcal{Y}$, compute $\zeta_n^b(A|X_j)=\hat{P}_n(A|X_j)-\hat{P}_n^b(A|X_j)$.

Step 3. Discard at most a proportion α of the bootstrap indices, and compute $\beta_n(A|X_j)$ as the maximum over the remaining bootstrap realizations $\zeta_n^b(A|X_j)$.

Discarding at most $B\alpha$ among the bootstrap realizations guarantees the control of the level of confidence, and we wish to choose the set $D\subseteq\{1,\dots,B\}$ of discarded indices so as to make β_n *as small as possible, to maximize informativeness* of the resulting confidence region. Again, if there was a total order, we would be similarly discarding the $B\alpha$ largest realizations of ζ_n^b, effectively choosing β_n as the quantile of the distribution of ζ_n^b, $b=1,\dots,B$. Instead, we discard all realizations of the matrix $\zeta_n^b(A|X_j)$ that have at least one entry that strictly exceeds the $(1-\alpha)$ quantile of $w_b=\max_{1\leq j\leq n}\max_{A\subseteq\mathcal{Y}}\zeta_n^b(A|X_j)$. Hence, we choose D solving the optimization problem

$$\min\Big\{\max_{b\notin D} w_b : D\subseteq\{1,\dots,B\}, |D|\leq B\alpha\Big\}. \tag{3.7}$$

The procedure is explained graphically in Figure 2.

Problem (3.7) can be solved as follows.

Bootstrap realization selection (BRS) algorithm.

BRS Step 1. For each $b\leq B$, set $w_b'=\max_{1\leq j\leq n}\sum_{y\in\mathcal{Y}}\max\{0,\hat{P}_n(y|X_j)-\hat{P}_n^b(y|X_j)\}$.

BRS Step 2. Let D be the set of indices b of the $\lfloor B\alpha\rfloor$ largest w_b'.

Proposition 1. *The BRS algorithm determines an optimal solution to problem* (3.7) *in* $O(nB|\mathcal{Y}|)$ *time.*

Remark 1. Problem (3.6) may have alternate optimum solutions. As observed by a referee, this may arise when the sample size n is small, since $\hat{P}_n(y|X_j)$ and $\hat{P}_n^b(y|X_j)$ are multiples of $1/n$ and thus distinct w_b's are more likely to have the same value when the sample size n is small. In case of ties, *any* optimum solution D to Problem (3.6) may be used to discard bootstrap realizations and determine functions β_n. If one desires a specific tie-breaking rule, e.g., for robustness or reproducibility, then we suggest

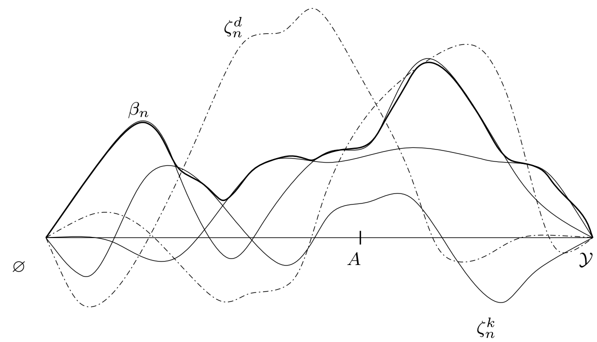

FIGURE 2. Stylized representation of the determination of the functional quantile β_n in a case without explanatory variables. The subsets A of $\mathcal{Y}$ are represented on the horizontal axis, ranging from $\emptyset$ to $\mathcal{Y}$. ζ_n^d is one of two discarded realizations of the empirical process (dotted lines), whereas ζ_n^k is one of three realizations that are not discarded (solid lines). β_n is the pointwise maximum over the realizations that were not discarded (thick line).

the following lexicographic selection rule as a refinement to BRS Step 2: Let w^b be the vector with components $w_j^b = \sum_{y\in\mathcal{Y}} \max\{0, \hat{P}_n(y|X_j) - \hat{P}_n^b(y|X_j)\}$ for $j = 1, \dots, n$, and let $[w]^b$ be the vector w^b with its components sorted in nonincreasing order, i.e., with $[w]_1^b = w_b \geq [w]_2^b \geq \cdots \geq [w]_n^b = \min_j w_j^b$; then discard the $\lfloor B\alpha \rfloor$ bootstrap realizations b with the lexicographically largest vector $[w]^b$. In other words, we refine problem (3.6) as $\text{lexmin}\{\text{lexmax}_{b\notin D}[w]^b : D \subseteq \{1, \dots, B\}, |D| \leq B\alpha\}$, where lexmin and lexmax denote the minimum and maximum relative to the lexicographic total order of vectors with n components. This rule aims at simultaneously minimizing all the values $\beta(A|X_j)$ without going through extensive additional computations.

In problem (3.7), we chose to minimize the maximum, over all $j \in \{1, \dots, n\}$ and $A \subseteq \mathcal{Y}$, of the nondiscarded bootstrap realizations $\zeta_n^b(A|X_j)$. Other objectives are possible, for example, the $\mathbb{L}^1$ objective $\sum_{b\notin D} w_b$. The main justification for the $\mathbb{L}^\infty$ norm objective $\max_{d\notin D} w_b$ in (3.7) is that it leads to a problem solvable in linear time. In contrast, the problem with an $\mathbb{L}^1$ objective is computationally difficult, namely NP-hard in the strong sense, as shown in the next result.

PROPOSITION 2. *Minimization of* $\{\sum_{b\notin D} w_b : |D| \leq \lfloor B\alpha \rfloor, D \subseteq \{1, \dots, B\}\}$ *is NP-hard in the strong sense.*

This result implies that unless $P = NP$, there exists no algorithm for this problem that runs in polynomial time. This is a severe computational drawback relative to the linear-time algorithm achieved with BRS.

3.3 *Search in the parameter space*

Once the functional quantile has been computed, there remains to search in the parameter space for the values of θ that satisfy (3.2). As shown in the Lemma 1, the function to be optimized in characterization (2.2) of the identified set is supermodular.

Definition 5 (Supermodular function). A set function $\rho : A \mapsto \rho(A) \in \mathbb{R}$ is called supermodular (resp. submodular) if for all pairs of sets (A, B), $\rho(A \cup B) + \rho(A \cap B) \geq$ (resp. $\leq$) $\rho(A) + \rho(B)$.

Lemma 1. *The function $V \mapsto P(V^{\cup}|x)$ is submodular for all $x \in \mathcal{X}$.*

In the computation of $\Theta_n(\overline{P}_n)$, it may be desirable to require $\overline{P}_n(V^{\cup}|x)$ to also be submodular as a function of $V \subseteq \mathcal{U}$, so that the function to be maximized in (3.2) can be maximized using submodular optimization techniques. This can be achieved by adding the following additional linear constraints (see Schrijver (2004)): $\forall u \neq v \in \mathcal{U}, \forall V \subseteq \mathcal{U} \setminus \{u, v\}, j = 1, \dots, n$,

$$\begin{aligned} &\overline{P}_n\big(\big[V \cup \{u\} \cup \{v\}\big]^{\cup}|X_j\big) - \overline{P}_n\big(\big[V \cup \{u\}\big]^{\cup}|X_j\big) \\ &\quad - \overline{P}_n\big(\big[V \cup \{v\}\big]^{\cup}|X_j\big) + \overline{P}_n\big([V]^{\cup}|X_j\big) \leq 0. \end{aligned} \tag{3.8}$$

The problem of checking whether θ is in the confidence regions can then be solved in polynomial time. Moreover, since submodular optimization has far ranging applications in all areas of operations research, many extremely efficient algorithms and implementations are readily available.

3.4 *Summary of results relevant to set inference*

Derive function $\beta_n : 2^{\mathcal{Y}} \times \mathcal{X} \to [0, 1]$ with the bootstrap functional quantile algorithm of Section 3.2. Define $\overline{P}_n$ for each $A \subseteq \mathcal{Y}$ and $x \in \mathcal{X}$ as $\overline{P}_n(A|x) = \hat{P}_n(A|x) + \beta_n(A|x)$ and construct confidence region $\Theta_n(\overline{P}_n)$ according to (3.2). The results of Section 3 can be summarized in the following result.

Theorem 3. *If $\mathcal{X}$ is finite, $\Theta_I(\overline{P}_n)$ is a valid $1 - \alpha$ confidence region for the identified set Θ_I.*

Theorem 3 readily follows from the combined facts that β_n satisfies (3.6) by construction, and the bootstrap empirical process $\sqrt{n}[\hat{P}_n^*(A|x) - \hat{P}_n(A|x)]$ finitely indexed by $A \subseteq \mathcal{Y}$ and $x \in \mathcal{X}$ has the same limiting distribution as the empirical process $\sqrt{n}[\hat{P}_n(A|x) - P(A|x)]$ by Theorem 4.1 of Bickel and Freedman (1981).

4. Simulation based on Example 1

We now illustrate and assess the performance of our procedure on the game described in Example 1. Throughout the experiment, we assume that $(\varepsilon_1, \varepsilon_2)$ is uniformly distributed on $[-1/2, 1/2]^2$ and $J = (J_1, J_2)$ is a vector of independent Bernoulli$(1/2)$ random variables. True values for the parameters are indicated with a 0 subscript. We consider the following true parameter specifications: $(\beta_0, c_0) = (0, 0)$ (point identified case) and $(\beta_0, c_0) = (0, 1/4)$ (which corresponds in some sense to the greatest possible indeterminacy). For the true data generating process, we consider four distinct equilibrium selection rules (which, like the true parameter values, are of course supposed unknown in the inference procedure). The first rule specifies that in case of multiplicity, all equilibrium strategy profiles in the equilibrium correspondence are selected with equal probability: we call this case *uniform selection*. The second selection rule specifies that in case of multiplicity, only symmetric equilibria will be selected, and the latter with equal probability if there are more than one. We call this case *symmetric selection*. The third selection rule specifies that in case of multiplicity, the equilibrium with largest aggregate investment is selected: Suppose, for instance, that the equilibrium correspondence takes the value $\{HH, HL, LO\}$; then equilibrium strategy profile HH is realized. We call this case *maximal selection*. Finally, the case, where the equilibrium with the lowest aggregate investment (the poverty trap) is selected is dubbed *minimal selection*. Maximal and minimal selections are identical in case $(\beta_0, c_0) = (0, 0)$, so we consider seven simulation experiments in total. In all cases with $(\beta_0, c_0) = (0, 0)$, $\beta_0 = 0$ is assumed known a priori by the analyst performing inference (to avoid an unbounded identified set in the simulations). In the remaining cases, β_0 is unknown a priori.

The experiment is run as follows. In each of the seven cases above, we calculate the distribution of the true data generating process. With the latter, we compute the identified set. In the point identified cases, namely $(\beta_0, c_0) = (0, 0)$ in all cases and $(\beta_0, c_0) = (0.25, 0)$ in the minimal selection case, the identified set is equal to the true value. In the case $c_0 = 0.25$, with $\beta = 0$ known a priori, the identified set is $[0.211, 0.333]$ for maximal selection as explained in the example at the end of Section 2; for uniform selection, it is $[0.250, 0.275]$. In case $(c_0 = 0.25, \beta_0 = 0)$ and β_0 a priori unknown, with uniform selection, the identified set projects to $[0, 0.375]$ on the c coordinate and to $[0, 0.320]$ on the β coordinate.

We then simulate 5000 samples of sizes $n = 100$, $n = 500$, and $n = 1000$ from these distributions. We use 999 bootstrap replications. We consider confidence levels 90%, 95%, and 99%. Coverage probabilities of the true value and of the identified set by the confidence region, as computed from the 5000 samples, are displayed in Tables 2, 3, 4, and 5 for the data generating process obtained with uniform, symmetric, maximal, and minimal selections, respectively.

We report Monte Carlo coverage of the true value (Point Coverage) and of the identified set (Set Coverage). The two are identical in case of point identification, hence the dashes in the tables under Set Coverage for such cases. We also report the effective level at which condition (3.5) is satisfied to directly assess the bootstrap functional quantile approximation.

We find only 12 cases of undercoverage of the identified set out of a possible 72. We also find 12 cases of undercoverage of the true value out of 72. All cases of undercoverage except 4 occur for sample size 100. We find very high levels of point coverage in all partially identified cases. Coverage of the identified set is higher in cases where the equilibrium selection involves randomization, i.e., when the true data generating process involves uniform and symmetric selection, as opposed to maximal and minimal selection.

Monte Carlo frequency of violation of condition (3.5) is lower than the theoretical level in 21 cases out of a possible 72. This overrejection occurs mostly for sample size 100.

TABLE 2. Coverage probabilities of (c_0, β_0) and of the identified set by the confidence region, as computed from 5000 samples.

(c_0, β_0)	n	Level	Point Coverage	Set Coverage	Condition (3.5)
$(0, 0)$	100	0.99	0.9888	–	0.9866
		0.95	0.9790	–	0.9750
		0.90	0.9674	–	0.9636
	500	0.99	0.9936	–	0.9930
		0.95	0.9874	–	0.9866
		0.90	0.9802	–	0.9784
	1000	0.99	0.9928	–	0.9916
		0.95	0.9882	–	0.9868
		0.90	0.9836	–	0.9812
$(0.25, 0)$	100	0.99	1	0.9986	0.9828
		0.95	0.9998	0.9972	0.9740
		0.90	0.9994	0.9968	0.9640
	500	0.99	1	1	0.9930
		0.95	1	0.9998	0.9884
		0.90	1	0.9996	0.9822
	1000	0.99	1	0.9988	0.9914
		0.95	1	0.9988	0.9854
		0.90	1	0.9988	0.9786

Note: The data generating process is obtained with uniform selection. The dashes indicate that the coverage probability is identical to that in the cell on the left, as a consequence of the point identification of (c_0, β_0) in the case $(0, 0)$.

TABLE 3. Coverage probabilities of (c_0, β_0) and of the identified set by the confidence region, as computed from 5000 samples.

(c_0, β_0)	n	Level	Point Coverage	Set Coverage	Condition (3.5)
$(0, 0)$	100	0.99	0.9896	–	0.9878
		0.95	0.9764	–	0.9720
		0.90	0.9608	–	0.9528
	500	0.99	0.9924	–	0.9920
		0.95	0.9864	–	0.9844
		0.90	0.9762	–	0.9724
	1000	0.99	0.9936	–	0.9908
		0.95	0.9876	–	0.9838
		0.90	0.9802	–	0.9752
$(0.25, 0)$	100	0.99	0.9978	0.9974	0.9676
		0.95	0.9976	0.9968	0.9490
		0.90	0.9974	0.9962	0.9182
	500	0.99	1	1	0.9896
		0.95	1	1	0.9780
		0.90	1	0.9994	0.9588
	1000	0.99	1	0.9998	0.9882
		0.95	1	0.9998	0.9758
		0.90	1	0.9996	0.9592

Note: The data generating process is obtained with symmetric selection. The dashes indicate that the coverage probability is identical to that in the cell on the left, as a consequence of the point identification of (c_0, β_0) in the case $(0, 0)$.

Table 4. Coverage probabilities of (c_0, β_0) and of the identified set by the confidence region, as computed from 5000 samples.

(c_0, β_0)	n	Level	Point Coverage	Set Coverage	Condition (3.5)
$(0, 0)$	100	0.99	0.9824	–	0.9792
		0.95	0.9608	–	0.9592
		0.90	0.9320	–	0.9308
	500	0.99	0.9928	–	0.9926
		0.95	0.9794	–	0.9792
		0.90	0.9618	–	0.9606
	1000	0.99	0.9874	–	0.9870
		0.95	0.9782	–	0.9776
		0.90	0.9616	–	0.9610
$(0.25, 0)$	100	0.99	0.9368	0.9368	0.9278
		0.95	0.9364	0.9354	0.9068
		0.90	0.9250	0.9226	0.8770
	500	0.99	0.9884	0.9884	0.9834
		0.95	0.9770	0.9766	0.9648
		0.90	0.9590	0.9584	0.9366
	1000	0.99	0.9922	0.9918	0.9858
		0.95	0.9802	0.9798	0.9652
		0.90	0.9658	0.9648	0.9392

Note: The data generating process is obtained with maximal selection. The dashes indicate that the coverage probability is identical to that in the cell on the left, as a consequence of the point identification of (c_0, β_0) in the case $(0, 0)$.

Table 5. Coverage probabilities of (c_0, β_0) and of the identified set by the confidence region, as computed from 5000 samples.

(c_0, β_0)	n	Level	Point Coverage	Set Coverage	Condition (3.5)
$(0.25, 0)$	100	0.99	0.9394	–	0.9362
		0.95	0.9296	–	0.9218
		0.90	0.9126	–	0.8964
	500	0.99	0.9856	–	0.9846
		0.95	0.9722	–	0.9680
		0.90	0.9548	–	0.9464
	1000	0.99	0.9900	–	0.9878
		0.95	0.9778	–	0.9732
		0.90	0.9606	–	0.9530

Note: The data generating process is obtained with maximal selection. The identified set and its coverage in the case $(0, 0)$ are identical with maximal and minimal selections, and are omitted here. The dashes indicate that the coverage probability is identical to that in the cell on the left, as a consequence of the point identification of (c_0, β_0) in the case $(0.25, 0)$.

In almost all cases, the Monte Carlo set coverage is higher than the Monte Carlo frequency of condition (3.5) being satisfied.

Improvements with sample size (in the sense of reducing undercoverage or overrejection) occurs in all 12 cases of set undercoverage, in all 12 cases of true value undercoverage, and in all 21 cases of overrejection of condition (3.5). Overall, the procedure is conservative and this feature persists with increasing sample sizes.

5. Application to long term elderly care decisions

We estimate the determinants of long term care option choices for elderly parents in American families. The model we use closely follows that proposed by Engers and Stern (2002), who present these choices as the result of a nonfamily participation game. The family members decide simultaneously whether to participate in a family reunion where the care option maximizing the participants' utility is chosen. Profits are then split among these participants according to some benefit-sharing rule. The data consist of a sample of 1212 elderly Americans 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 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 the elderly parent 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 days per week each of the children devotes 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 who provides the most time is designated the primary care giver. If the elderly parent lives in a nursing home, then the nursing home is the primary care giver. If no child is listed and the parent does not live in a nursing home, then the parent is designated as "living alone." Table 6 presents the list of variables used in the analysis. They include parent characteristics, characteristics of the children, and the care option chosen. A more detailed discussion, summary statistics, and additional results can be found in the Supplement.

Table 6. List of variables.

Variables	Equal to 1 if	Percentage of Sample
Care option	Living with child 1	26.81
	Living with child 2	6.75
	Living in nursing home	19.92
	Living home alone	46.54
Parent variables		
DA	Highly disabled	33.81
DM	Living with the spouse	40.36
Children variables		
DD	Distance from parent: 31 min and more	49.45
DS	Female	49.26

5.1 *The game*

The observable choice of care option is modeled as in Engers and Stern (2002) as the outcome of a family bargaining game. We index family members as follows: parent, 0; firstborn child, 1; second born child, 2. The payoff to family member i, $i = 0, 1, 2$, is the sum of three terms. The first term V_{ij} is the value to parent 0 and to child i of care option j, where $j \in 1, 2$ means child j becomes the primary care giver, $j = 0$ means the parent remains self-reliant, and $j = 3$ means the parent is moved to a nursing home. The matrix $V = (V_{ij})_{ij}$ is known to both children and the parent. We suppose it takes the form

$$V(X;\theta) = \begin{bmatrix} \begin{pmatrix} \beta_{00} \\ +\beta_m DM \\ +\beta_{ah} DA \end{pmatrix} & (\psi_s DS_1) & 0 & 0 \\ \begin{pmatrix} \beta_m DM \\ +\beta_{ah} DA \end{pmatrix} & \begin{pmatrix} \beta_{11} \\ +\psi_1 DD_1 \\ +\beta_{ac} DA \end{pmatrix} & 0 & 0 \\ \begin{pmatrix} \beta_m DM \\ +\beta_{ah} DA \end{pmatrix} & 0 & \begin{pmatrix} \beta_{11} \\ +\psi_1 DD_2 \\ +\beta_{ac} DA \end{pmatrix} & 0 \end{bmatrix} + \begin{bmatrix} u_1 \; u_2 \; u_3 \; u_4 \\ u_1 \; u_2 \; u_3 \; u_4 \\ u_1 \; u_2 \; u_3 \; u_4 \end{bmatrix},$$

where all explanatory variables are defined in Table 6 and u is a matrix of $\text{nid}(0, \sigma_u^2)$ unobserved utility shocks that are common knowledge to the participants.

The second term in the payoff results from the family bargaining process as follows. We assume that it is always in the interest of the parent to attend the family reunion. However, child i $(i = 1, 2)$ can refrain from participating in the meeting. By choosing not to participate, a member of that family agrees on whatever is decided but can neither assume the role of primary care giver nor be involved in any side payment. Both children simultaneously decide whether or not to participate in the long term care decision. Suppose M is the set of children who participate. The option chosen is option $j \in M \cup \{0, 3\}$, which maximizes the participants's total utility $\sum_{i \in M} V_{ij}$. It is assumed that participants abide by the decision and that benefits are then shared equally among parent and children participating in the decision through a monetary transfer s_i, which is the second term in the children's payoff. The third term ϵ_i in the payoff is a random benefit from participation, which is 0 for children who decide not to participate and is distributed according to absolutely continuous distribution $\nu(\cdot|\theta)$ for each child who does participate. All children observe the realizations of ϵ, whereas the analyst only knows its distribution. The payoff matrix is given in Table 7, where overall benefit shares w_i^{IJ}, $i = 1, 2$, $I, J = N, P$, are defined and derived in the Supplement. Multiple Nash equilibria in pure and mixed strategies are also derived in the Appendix. Each equilibrium action profile results in a (almost surely) unique care option choice; hence for each participation shock ϵ, we can derive $G(\epsilon|X;\theta)$ as the set of probability measures on the set of care options $\{0, 1, 2, 3\}$ induced by mixed strategy profiles, which are probabilities on the set of participation profiles $\{NN, NP, PN, PP\}$.

Table 7. Payoffs for the family participation game.

	Child 2	
Child 1	N	P
N	w_1^{NN}, w_2^{NN}	$w_1^{NP}, \varepsilon_2 + w_2^{NP}$
P	$\varepsilon_1 + w_1^{PN}, w_2^{PN}$	$\varepsilon_1 + w^{PP}, \varepsilon_2 + w^{PP}$

5.2 *Estimation methodology*

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. As described in Section 3, we account for sampling uncertainty and control the level of confidence by constructing set functions $A \mapsto \overline{P}(A|X)$, which dominate $P(A|X)$ (uniformly over $A \subseteq \{0, 1, 2, 3\}$ and X) with probability $1-\alpha$ (the chosen level of confidence; here 0.95). We implement the method detailed in Section 3 with a number of bootstrap replications $B = 999$. Second, we obtain the model likelihood by simulating the valuation matrix and computing the equilibrium correspondence from the payoff matrix, for given values of X and θ. The procedure, for a given X and θ, is as follows.

- We generate and store R draws of the vector u and ε from the distribution ν_θ. Here $R = 1000$. The components of random vector u (resp. ε) are independent and normally distributed with mean 0 (resp. μ) and variance σ_1 (resp. σ_2). Note that μ, σ_1, σ_2 belong to the parameter θ.
- For each value u^r, we compute the valuation matrix $V(X, u^r, \theta)$ and the corresponding payoff matrix.
- Then we determine the equilibrium correspondence $G(X, u^r, \varepsilon^r; \theta)$ from the analytical results derived in the preceding section. The Gambit software provides an alternative for computing numerically the sets of Nash equilibria for more complex games.
- The last step of the simulation is to compute an estimator of the model likelihood $\mathcal{L}$ defined in (3.3) as $\hat{\mathcal{L}}(A|X; \theta) = \frac{1}{R}\sum_{r=1}^{R} \min\{\sigma(A) : \sigma \in G(X, u^r, \varepsilon^r; \theta)\}$.

Having constructed those two elements, the identified set comprises all values of θ such that for all observed values of the explanatory variables, the minimum over $A \subseteq \{0, 1, 2, 3\}$ of the function $\overline{P}(A|X) - \hat{\mathcal{L}}(A|X; \theta)$ is nonnegative, as explained in Section 3. We used a grid search over the parameter space.[2] In the following application, a standard laptop is used to compute the five-dimensional parameter space in a few hours. In the Supplement, we construct confidence regions for specifications involving 12-dimensional parameter space. In the latter case, since estimation time grows exponentially with the number of parameters induced by the model, parallel processing becomes necessary. We use an open-MP procedure for parallel processing, which is perfectly suited to the method we propose. The computation resources have been provided

[2]The grid was adaptive in the sense that more points were investigated in the interval between an accepted and a rejected point, so as to better approximate the frontier of the identified set.

by the Réseau Québécois de Calcul de Haute Performance (RQCHP). All computation where made under the system Cottos, which provides up to 128 computation nodes (1024 CPU cores) equipped with two Intel Xeon E5462 quad-core processors at 3 GHz. Under one node, approximately 5×10^7 parameter points can be tested in 24 hours.

5.3 *Results*

We perform the estimation under different values of the mean and variance of the error term. We first test the significance of some of the individual parameters by checking whether the hyperplanes defined by $\theta_i = 0$—where θ_i is a component of θ—intersect the 95% confidence region. We fail to reject the null hypothesis if the estimation procedure returns a nonempty set. We then obtain a constrained confidence region for the remaining parameters. For each value of mean and variance of the error term, we find a nonempty intersection between the confidence region and the hyperplane defined by $\beta_{11} = 0$. This means we fail to reject (at the 5% level) the null hypothesis that there is no additional constant disutility for a child to take care of an elderly parent. Since, this hypothesis is not rejected, we obtain a constrained confidence region for the remaining parameters.

We construct the confidence region under the additional constraint $\beta_{ah} = -\beta_m$ (see the Supplement for a discussion of this hypothesis). We note that the null hypothesis $H_0 : \beta_{00} = 0$ is always rejected. Hence, when we control for all other effects, parents are not indifferent between the first two options. They show a clear preference in favor of living in their own home (option called *living alone*) instead of living in a nursing home (β_{00} is always positive). Note that the identified set is not a compact set. In particular, β_{ac}, β_{ah}, β_m, and ψ_d are allowed to diverge to $-\infty$.

The following results are generally consistent with expectations and previous results on the subject.

(1) The existence of several problems with the parent's functional ability is a key determinant of the decision to enter a nursing home. The parameters β_{ah} and β_{ac} are both negative and can both be (very) large. The negative sign of β_{ah} captures the fact that a parent's disability increases the value of care provided by the family or a specialized institution. In addition, $\beta_{ac} < 0$ means that the disability entails a utility cost for the child if he is chosen as the primary care giver.

(2) Parameter β_m associated with the parent living with a spouse is positive and large. This implies that married parents are more likely to remain self-reliant. In families where the parent is disabled, the effect of living with the spouse compensates the disutility of disability and preserves the incentive for parents to live at home.

(3) While we cannot rule out parents being indifferent to the gender and birth order of their primary care giver, estimation shows a tilt of the confidence interval toward positive values for both parameters, with a possible positive and large magnitude of the parameter α.

(4) Children living more than 30 minutes from the parents are less likely to provide care than those living closer to the parents. Distance has a (possibly strong) disutility effect on children's incentives to participate in the care decision.

Figure 3 shows two-dimensional projections and cuts of the confidence region for column 2 of Table 8, i.e., $\mu_\varepsilon = 0$, $\sigma_\varepsilon = 1$, and $\sigma_u = 1$. Notice the triangular shape of the region plotted in Figure 3(a), which entails that large values of β_{ah} are only permitted when β_{ac} is also large.

Conclusion

We have considered the problem of statistical inference in incomplete partially identified structural models, such as models of discrete choice with interactions and other forms of endogeneity. A characterization of the identified set for structural parameters was given with an appeal to a classical theorem in combinatorial optimization, the max-flow min-cut theorem, thereby emphasizing the optimization formulation of the problem of inference in such models. Finally, we have shown how to apply combinatorial optimization methods within a bootstrap procedure so as to compute informative confidence regions very efficiently, hence feasibly, in empirically relevant applications. An application of the methodology was carried out on a family bargaining example and it was shown that most findings in the literature on the determinants of long term elderly care by American families were supported in this more robust framework where the effects of interaction are accounted for. This procedure applies to very general classes of models, and its efficiency and coverage properties could no doubt be improved when tailored to more specific applications. In particular, the application to matching games and revealed preference testing of stability in matching still poses considerable challenges. Other perspectives for further work include the application of max-flow–min-cut algorithms to the detection of redundant inequalities at the identification stage so as to improve the performance at the inference stage, possibly by appealing to other existing procedures if the number of nonredundant inequalities is small enough.

Appendix: Proofs of results in the main text

Proof of Theorem 1. By Proposition 1 of Galichon and Henry (2011), a value θ of the parameter vector belongs to Θ_I if and only if $\mathbb{P}(Y \in G(X, \varepsilon; \theta)) = 1$, X-a.s. (which we drop from the notation from this point on). Hence, if there exists a pair (Y, U) of random vectors on $\mathcal{Y} \times \mathcal{U}$ such that Y has probability mass $P(y|X)$, $y \in \mathcal{Y}$, then U has probability mass $Q(u|X; \theta)$, $u \in \mathcal{U}$, and $\mathbb{P}(Y \in U|X) = 1$. This is equivalent to the existence of nonnegative weights π_y^u, $(y, u) \in \mathcal{Y} \times \mathcal{U}$, such that $\sum_{u\in\mathcal{U}} \pi_y^u = P(y|X)$, $\sum_{y\in\mathcal{Y}} \pi_y^u = Q(u|X)$; $\pi_y^u = 0$ when $y \notin u$. The latter is equivalent to the following programming problem with auxiliary variables a_y, $y \in \mathcal{Y}$, and a^u, $u \in \mathcal{U}$, having zero as a solution. The programming problem is the following: $\min(\sum_{y\in\mathcal{Y}} a_y + \sum_{u\in\mathcal{U}} a^u)$ subject to the constraints $\sum_{u\in\mathcal{U}} \pi_y^u + a_y \le P(y|X)$, $\sum_{y\in\mathcal{Y}} \pi_y^u + a^u \le Q(u|X; \theta)$, a_y, a^u, $\pi_y^u \ge 0$, and $\pi_y^u = 0$ when $y \notin u$. Since $\sum_{y\in\mathcal{Y}} a_y + \sum_{u\in\mathcal{U}} a^u \le \sum_{y\in\mathcal{Y}} P(y|X) + \sum_{u\in\mathcal{U}} Q(u|X; \theta) - 2\sum_{y\in\mathcal{Y}}\sum_{u\in\mathcal{U}} \pi_y^u = 2 - 2\sum_{y\in\mathcal{Y}}\sum_{u\in\mathcal{U}} \pi_y^u$, the latter is also equivalent to $\max \sum_{y\in\mathcal{Y}}\sum_{u\in\mathcal{U}} \pi_y^u \ge 1$ subject to the constraints $\sum_{u\in\mathcal{U}} \pi_y^u \le P(y|X)$, $\sum_{y\in\mathcal{Y}} \pi_y^u \le Q(u|X)$, $\pi_y^u \ge 0$ and $\pi_y^u = 0$ when $y \notin u$.

The latter is equivalent to a maximum flow $\max \sum_{y\in\mathcal{Y}}\sum_{u\in\mathcal{U}} \pi_y^u$ of at least 1 in the directed network (N, E) with nodes in $N = \{S\} \cup \mathcal{Y} \cup \mathcal{U} \cup \{T\}$, edges in $E = \{(S, y) : y \in \mathcal{Y}\} \cup$

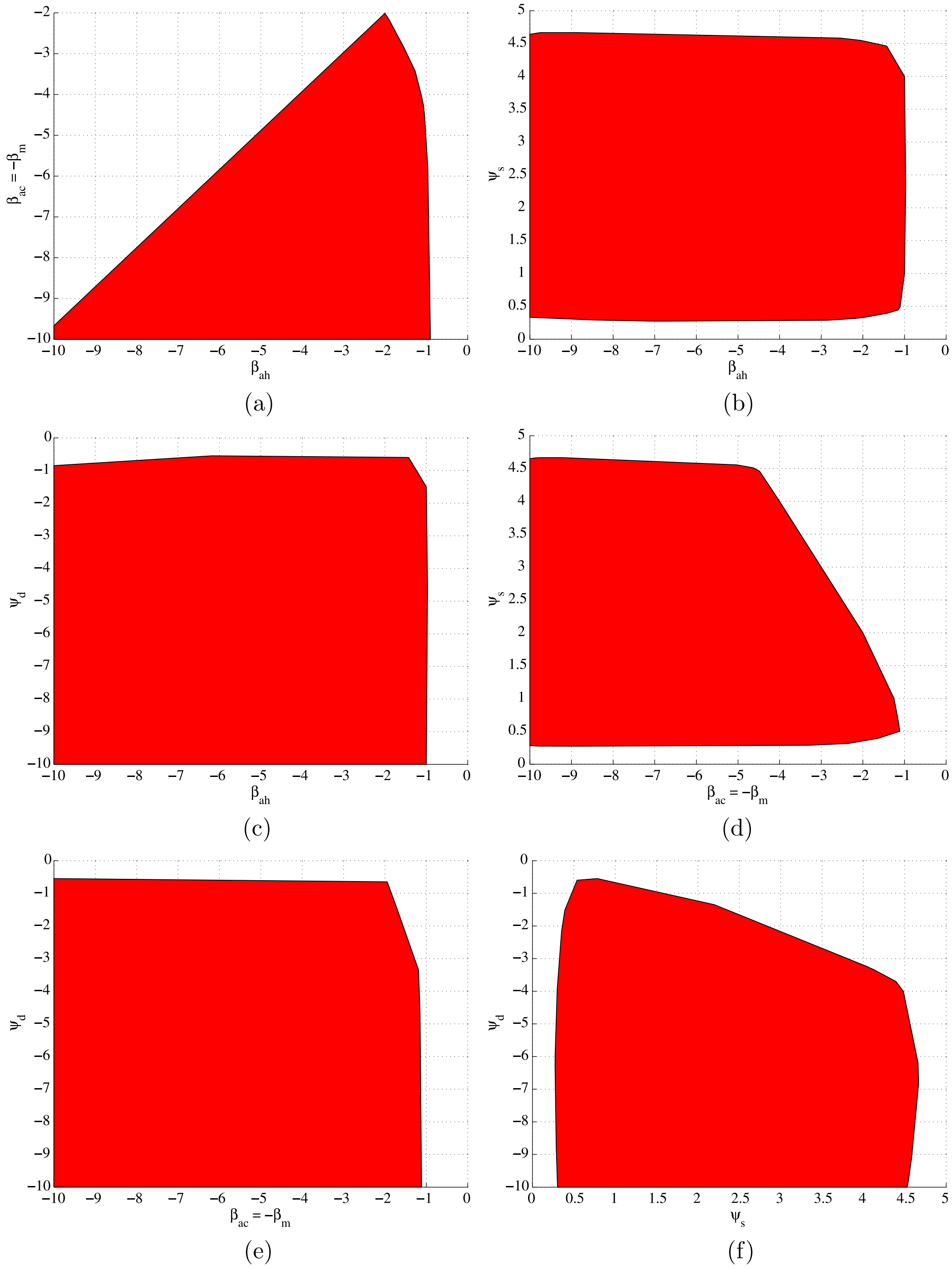

FIGURE 3. Two-dimensional representations of the confidence region at $\beta_{00} = 1$, $\mu = 0$, $\sigma_\varepsilon = 1$, $\sigma_u = 1$.

Table 8. Parameter range for estimation of the specified participation game at $\beta_{11}=0$, $\beta_{ah}=-\beta_m$, $\mu=0$, $\sigma_\varepsilon=1$ and for different values of the parameter $\sigma_u\in\{0.5,1,2\}$.

Parameters	Min	Max	Min	Max	Min	Max
β_{00}	0.35	1.50	0.70	3.85	2.00	8.85
β_{ah}	∞	-0.48	∞	-0.88	∞	-3.65
β_{ac}	∞	-1.05	∞	-1.80	∞	-4.35
ψ_s	0.05	4.70	0.00	8.80	0.00	17.70
ψ_d	∞	-0.55	∞	-0.45	∞	-0.85
σ_u	0.5	0.5	1	1	2	2

$\{(y,u):y\in\mathcal{Y},u\in\mathcal{U},y\in u\}\cup\{(u,T):u\in\mathcal{U}\}$, and capacity constraints $C(S,y)=P(y|X)$, each $y\in\mathcal{Y}$ and $C(u,T)=Q(u|X)$, each $u\in\mathcal{U}$, the remaining edges being unconstrained (infinite capacity). Consider cuts (Z,V) such that $S\in Z$ and $T\in V$. Since the capacity of an edge from y to u is such that $y\in u$ is infinite, cut (Z,V) has finite capacity if and only if $y\in u$ and $u\in V$ jointly imply $y\in Z$. Such a cut has capacity $C(Z,V)=\sum_{y\in Z}P(y|X)+\sum_{u\in\mathcal{U}\setminus V}Q(u|X;\theta)=\sum_{y\in Z}P(y|X)+1-\sum_{u\in V}Q(u|X;\theta)$. This capacity is minimal when $y\notin u$ and $u\in V$ jointly imply $y\notin Z$, hence if $Z=V^\cup=\bigcup\{u:u\in V\}$. Therefore, the capacity of a minimal cut is $C(V^\cup,V)=\sum_{y\in V^\cup}P(y|X)+1-\sum_{u\in V}Q(u|X;\theta)=P(V^\cup|X)+1-Q(V|X;\theta)$. By the max-flow–min-cut theorem, the capacity of any minimal cut is equal to the maximum flow through the network; hence $\theta\in\Theta_I$ if and only if for all subsets V of $\mathcal{U}$, $P(V^\cup|X)+1-Q(V|X;\theta)\geq 1$, i.e., $Q(V|X;\theta)\leq P(V^\cup|X)$, and (2.2) follows. (2.3) is obtained by inverting the network and repeating the reasoning. □

Proof of Corollary 1. Since $(Z^\cap)^c=\{u\in\mathcal{U}:u\subseteq Z\}^c=\{u\in\mathcal{U}:u\cap Z^c\neq\emptyset\}=Z^{-1}$, the result follows from (2.3) by taking complements. □

Proof of Lemma 1. Take an $x\in\mathcal{X}$. Take any $u\in\mathcal{U}$ and $V\subseteq\mathcal{U}\setminus\{u\}$. We have $P([V\cup\{u\}]^\cup|x)-P(V^\cup|x)=\sum_{y\in\bigcup_{v\in V\cup\{u\}}v}P(y|x)-\sum_{y\in\bigcup_{v\in V}v}P(y|x)=\sum_{y\in u\setminus V^\cup}P(y|x)=P(u\setminus V^\cup|x)$, which is nonincreasing in V, hence the result. □

Proof of Theorem 2. Given a value $\theta\in\Theta_I$, by Theorem 1, we have

$$\sup_{x\in\mathcal{X}}\max_{V\subseteq\mathcal{U}}\big(Q(V|x;\theta)-P(V^\cup|x)\big)\leq 0.$$

Under Assumption 1,

$$\sup_{x\in\mathcal{X}}\max_{V\subseteq\mathcal{U}}\big(P(V^\cup|x)-\overline{P}_n(V^\cup|x)\big)\leq 0,$$

with limiting probability larger than $1-\alpha$. Hence, with probability at least $1-\alpha$,

$$\sup_{x\in\mathcal{X}}\max_{V\subseteq\mathcal{U}}\big(Q(V|x;\theta)-\overline{P}_n(V^\cup|x)\big)\leq 0$$

and, thus, $\theta\in\Theta_I(\overline{P}_n)$. □

Proof of Proposition 1. We first justify BRS Step 1 by showing that $w_b = w_b'$ for all b. Indeed observe that for any $j \in \{1, \dots, n\}$ and $A \subseteq \mathcal{Y}$, we have

$$\zeta_n^b(A|X_j) = \sum_{y \in A} \hat{P}_n(y|X_j) - \sum_{y \in A} \hat{P}_n^b(y|X_j) = \sum_{y \in A} \big[\hat{P}_n(y|X_j) - \hat{P}_n^b(y|X_j)\big]$$

and, thus, $\max_{A \subseteq \mathcal{Y}} \zeta_n^b(A|X_j)$ is attained by selecting all the elements $y \in \mathcal{Y}$ with $\hat{P}_n(y|X_j) - \hat{P}_n^b(y|X_j) > 0$. It follows that

$$w_b' = \max_{1 \le j \le n} \max_{A \subseteq \mathcal{Y}} \left[\sum_{y \in \mathcal{Y}} \hat{P}_n(y|X_j) - \sum_{y \in A} \hat{P}_n^b(y|X_j)\right]$$

and, therefore, $w_b = w_b'$. To justify BRS Step 2, let w^{opt} denote the optimum objective value of problem (3.7). If D fails to include any b such that $w_b > w^{\text{opt}}$, then $\max_{b \notin D} w_b > w^{\text{opt}}$; therefore, an optimal D must include all b such that $w_b > w^{\text{opt}}$. Alternatively, if D is any optimal subset and some $b' \in D$ satisfies $w_{b'} \le w^{\text{opt}}$, then discarding b' from D yields a feasible subset $D \setminus \{b'\}$ (since $|D \setminus \{b'\}| < |D| \le \overline{d}$) such that $\max_{b \in D \setminus \{b'\}} w_b \le \max_{b \in D} w_b$; hence $D \setminus \{b'\}$ is an alternate optimal solution. Therefore, an optimal D consists of all indices b such that $w_b > w^{\text{opt}}$. Concerning the running time, BRS Step 1 requires $O(nB|\mathcal{Y}|)$ time and BRS Step 2 requires $O(B)$ time using a linear-time selection (or median-finding) algorithm (see Blum, Floyd, Pratt, Rivest, and Tarjan (1973)). □

Proof of Proposition 2. The problem corresponds to the following decision problem: Given an $m \times n$ matrix H, an integer k, and a target value t, can one find a subset $S \subseteq \{1, \dots, n\}$ such that $|S| \ge k$ and $\sum_{i=1}^m \max_{j \in S} H_{ij} \le t$? Denote (H, k, t) as an instance of the latter problem. Consider the well known NP-hard decision problem CLIQUE (see, for instance, Section 4.8, p. 43, of Schrijver (2004)): Given a graph $G = (V, E)$ and an integer q satisfying $2 \le q \le |V|$, does there exist a subset $Q \subseteq V$ such that $|Q| \ge q$ and for all $i, j \in V$, $ij \in E$ (i.e., Q is a *clique*). To any instance (G, q) of the problem CLIQUE, we associate an instance (H, k, t) of our decision problem, where lines of H corresponds to vertices of G (elements of V), columns of H corresponds to edges in G (elements of E), and $H_{ij} = 1$ if vertex i belongs to edge j and is 0 otherwise. For any subset $S \subseteq E$ of edges in G, we have for all $i \in E$, $\max_{j \in S} H_{ij} = 1$ if i belongs to at least one element of S and is 0 otherwise. Hence, $\sum_{i \in E} \max_{j \in S} H_{ij}$ is the number of vertices that belong to at least one edge in S. Define $k = q(q-1)/2$ and $t = q$. Then a set S of k edges involves at least (hence exactly) q vertices if and only if S is the set of edges of a CLIQUE. Hence the answer to the decision problem (H, k, t) thus defined is YES if and only if G contains a CLIQUE with q vertices. Since CLIQUE is NP-complete, it follows that our decision problem is NP-hard. Since $k = O(|V|^2)$ and $t = O(|V|)$, the input size (in unitary notation) of such instances of our problem is polynomially bounded by the input size (in unitary or binary notation) $\Omega(|V|)$ of the corresponding instance of CLIQUE. Hence our decision problem is NP-hard in the strong sense. □

References

Ahuja, R. K., T. L. Magnanti, and J. B. Orlin (1993), *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, NJ. [501]

Andrews, D., S. Berry, and P. Jia (2003), "Placing bounds on parameters of entry games in the presence of multiple equilibria." Unpublished manuscript. [500]

Andrews, D. and G. Soares (2010), "Inference for parameters defined by moment inequalities using generalized moment selection." *Econometrica*, 78, 119–157. [509]

Aradillas-Lopez, A. (2010), "Semiparametric estimation of a simultaneous game with incomplete information." *Journal of Econometrics*, 157, 409–431. [505]

Bajari, P., J. Hahn, H. Hong, and G. Ridder (2011), "A note on semiparametric estimation of finite mixtures of discrete choice models with applications to game theoretic models." *International Economic Review*, 52, 807–824. [505]

Bajari, P., H. Hong, and S. Ryan (2010), "Identification and estimation of a discrete game of complete information." *Econometrica*, 78, 1529–1568. [506, 511]

Beresteanu, A., I. Molchanov, and F. Molinari (2011), "Sharp identification regions in models with convex predictions." *Econometrica*, 79, 1785–1821. [499, 500, 506, 511]

Berry, S. (1992), "Estimation of a model of entry in the airline industry." *Econometrica*, 60, 889–917. [505]

Berry, S. and E. Tamer (2006), "Identification in models of oligopoly entry." In *Advances in Economics and Econometrics*, 46–85, Cambridge University Press, Cambridge. [500, 506]

Bickel, P. and D. Freedman (1981), "Some asymptotic theory for the bootstrap." *The Annals of Statistics*, 9, 1196–1217. [512, 515]

Blum, M., R. Floyd, V. Pratt, R. Rivest, and R. Tarjan (1973), "Time bounds for selection." *Journal of Computational Systems Science*, 7, 448–461. [526]

Bresnahan, T. and P. Reiss (1990), "Entry in monopoly markets." *Review of Economic Studies*, 57, 531–553. [505]

Bugni, F. (2010), "Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set." *Econometrica*, 78, 735–753. [501, 510]

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models." *Econometrica*, 75, 1243–1284. [500, 501, 509, 510]

Chernozhukov, V., S. Lee, and A. Rosen (2013), "Inference on intersection bounds." *Econometrica*, 81 (2), 667–737. [512]

Chesher, A. (2010), "Instrumental variable models for discrete outcomes." *Econometrica*, 78, 575–601. [502]

Chesher, A., A. Rosen, and K. Smolinski (2013), "An instrumental variable model of multiple discrete choice." *Quantitative Economics*, 4 (2), 157–196. [505]

Ciliberto, F. and E. Tamer (2009), "Market structure and multiple equilibria in airline markets." *Econometrica*, 77, 1791–1828. [500, 511]

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 U01-AG007198. [519]

Echenique, F., S. Lee, M. Shum, and B. Yenmez (2013), "The revealed preference theory of stable and extremal stable matchings." *Econometrica*, 81 (1), 153–171. [506]

Ekeland, I., A. Galichon, and M. Henry (2010), "Optimal transportation and the falsifiability of incompletely specified economic models." *Economic Theory*, 42, 355–374. [501]

Engers, M. and S. Stern (2002), "Long-term care and family bargaining." *International Economic Review*, 43, 73–114. [519, 520]

Galichon, A. and M. Henry (2006), "Inference in incomplete models." Available from SSRN at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=886907. [506, 509]

Galichon, A. and M. Henry (2009), "A test of non-identifying restrictions and confidence regions for partially identified parameters." *Journal of Econometrics*, 152, 186–196. [509]

Galichon, A. and M. Henry (2011), "Set identification in models with multiple equilibria." *Review of Economic Studies*, 78, 1264–1298. [499, 500, 501, 506, 511, 523]

Galichon, A. and M. Henry (2013), "Dilation bootstrap." *Journal of Econometrics*, 177 (1), 109–115. [510]

Haile, P. and E. Tamer (2003), "Inference with an incomplete model of English auctions." *Journal of Political Economy*, 111, 1–51. [500]

Henry, M. and I. Mourifié (2013), "Euclidean revealed preferences: Testing the spatial voting model." *Journal of Applied Econometrics*, 28 (4), 650–666. [506]

Henry, M. and A. Onatski (2012), "Set coverage and robust policy." *Economics Letters*, 115 (2), 256–257. [509]

Jovanovic, B. (1989), "Observable implications of models with multiple equilibria." *Econometrica*, 57, 1431–1437. [500, 505]

Manski, C. (1989), "Anatomy of the selection problem." *Journal of Human Resources*, 24, 343–360. [500]

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015), "Moment inequalities and their application." *Econometrica*, 83, 315–334. [500, 506]

Romano, J. and A. Shaikh (2008), "Inference for identifiable parameters in partially identified econometric models." *Journal of Statistical Planning and Inference*, 138, 2786–2807. [509]

Romano, J. and A. Shaikh (2010), "Inference for the identified set in partially identified econometric models." *Econometrica*, 78, 169–211. [510]

Rosen, A. (2008), "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities." *Journal of Econometrics*, 146, 107–117. [509]

Schrijver, A. (2004), *Combinatorial Optimization: Polyhedra and Efficiency*. Springer, Berlin. [501, 515, 526]

Singh, K. (1981), "On the asymptotic accuracy of Efron's bootstrap." *The Annals of Statistics*, 9, 1187–1195. [512]

Soetevent, A. and Kooreman (2007), "A discrete choice model with social interactions: With an application to high school teen behaviour." *Journal of Applied Econometrics*, 22, 599–624. [505, 511]

Tamer, E. (2003), "Incomplete simultaneous discrete response model with multiple equilibria." *Review of Economic Studies*, 70, 147–165. [505]

Uetake, K. and Y. Watanabe (2013), "Entry by merger: Estimates from a two-sided matching model with externality." Unpublished manuscript. [505]

Submitted July, 2013. Final version accepted June, 2014.