cemmap
centre for microdata methods and practice

Sharp identification regions in models with convex moment predictions

Arie Beresteanu
Ilya Molchanov
Francesca Molinari

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP25/10

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Sharp Identification Regions in Models with Convex Moment Predictions*

Arie Beresteanu[†] Ilya Molchanov[‡] Francesca Molinari[§]

First draft: February 15, 2007. Current draft: August 30, 2010

Abstract

We provide a tractable characterization of the sharp identification region of the parameters θ in a broad class of incomplete econometric models. Models in this class have set valued predictions that yield a convex set of conditional or unconditional moments for the observable model variables. In short, we call these *models with convex moment predictions.* Examples include static, simultaneous move finite games of complete and incomplete information in the presence of multiple equilibria; best linear predictors with interval outcome and covariate data; and random utility models of multinomial choice in the presence of interval regressors data. Given a candidate value for θ, we establish that the convex set of moments yielded by the model predictions can be represented as the Aumann expectation of a properly defined random set. The sharp identification region of θ, denoted Θ_I, can then be obtained as the set of minimizers of the distance from a properly specified vector of moments of random variables to this Aumann expectation. Algorithms in convex programming can be exploited to efficiently verify whether a candidate θ is in Θ_I. We use examples analyzed in the literature to illustrate the gains in identification and computational tractability afforded by our method.

Keywords: Partial Identification, Random Sets, Aumann Expectation, Support Function, Finite Static Games, Multiple Equilibria, Random Utility Models, Interval Data, Best Linear Prediction.

---

[*] This paper supersedes two earlier working papers by the same authors, Beresteanu, Molchanov and Molinari (2008, 2009). We thank the co-Editor (Steven Berry) and three anonymous reviewers for very useful suggestions that substantially improved the paper. We are grateful to H. Bar, T. Bar, L. Barseghyan, I. Canay, G. Chamberlain, P. Ellickson, Y. Feinberg, K. Hirano, T. Magnac, T. O'Donoghue, M. Peski, A. Rosen, X. Shi, J. Stoye, A. Sweeting, and especially L. Blume, D. Easley and C. Manski for comments. We also thank seminar participants at many institutions and conferences where this paper was presented. We are grateful to M. Grant for advise on the use of the CVX software, which we used to run simulations in MatLab. Financial support from the USA NSF Grants: SES-0617559 and SES-0922373 (Beresteanu); SES-0617482 and SES-0922330 (Molinari); from the Swiss National Science Foundation Grant Nr. 200021-117606 and Grant Nr. 200021-126503 (Molchanov); and from the Center for Analytic Economics at Cornell University is gratefully acknowledged. Ilya Molchanov gratefully acknowledges the hospitality of the Department of Economics at Cornell University in October 2006.

[†] Department of Economics, Duke University, arie@econ.duke.edu.

[‡] Department of Mathematical Statistics and Actuarial Science, University of Bern, ilya@stat.unibe.ch.

[§] Department of Economics, Cornell University, fm72@cornell.edu.

1 Introduction

Overview. This paper provides a simple, novel, and computationally feasible procedure to determine the *sharp* identification region of the parameters θ characterizing a broad class of incomplete econometric models. Models in this class have set valued predictions which yield a convex set of conditional or unconditional moments for the model observable variables. In short, throughout the paper, we call these *models with convex moment predictions.* Our use of the term "model" encompasses econometric frameworks ranging from structural semi-parametric models, to non-parametric best predictors under square loss. In the interest of clarity of exposition, in this paper we focus on the semi-parametric case. We exemplify our methodology applying it to static, simultaneous move finite games of complete and incomplete information in the presence of multiple equilibria; and best linear predictors with interval outcome and covariate data.[1]

Models with convex moment predictions can be described as follows. For a given value of the parameter vector θ and realization of (a subset of) model variables, the economic model predicts a set of values for a vector of variables of interest. These are the model set valued predictions, which are *not* necessarily convex. No restriction is placed on the manner in which, in the data generating process, a specific model prediction is selected from this set. When the researcher takes conditional expectations of the resulting elements of this set, the unrestricted process of selection yields a convex set of moments for the model variables– this is the model's *convex set of moment predictions.* If this set were almost surely single valued, the researcher would be able to identify θ by matching the model-implied vector of moments to the one observed in the data. When the model's moment predictions are set valued, one may find many values for the parameter vector θ which, when coupled with specific selection mechanisms picking one of the model set valued predictions, generate the same conditional expectation as the one observed in the data. Each of these values of θ is observationally equivalent, and the question becomes how to characterize the collection of observationally equivalent θ's in a tractable manner.

Although previous literature has provided tractable characterizations of the sharp identification region for certain models with convex moment predictions (see, e.g. Manski (2003) for the analysis of nonparametric best predictors under square loss with interval outcome data), there exist many important problems, including the examples analyzed in this paper, in which such a characterization

---

[1] When thinking about best linear prediction (BLP), no "model" is assumed in any substantive sense. However, with some abuse of terminology, for a given value of the BLP parameter vector θ, we refer to the set of prediction errors associated with each logically possible outcome and covariate variables in the observable random intervals, as the "model set valued predictions." In the Supplement to this paper, Beresteanu, Molchanov, and Molinari (2010b), we also analyze random utility models of multinomial choice in the presence of interval regressors data.

is difficult to obtain. The analyses of Horowitz, Manski, Ponomareva, and Stoye (2003, HMPS henceforth), and Andrews, Berry, and Jia (2004, ABJ henceforth) and Ciliberto and Tamer (2009, CT henceforth), are examples of research studying, respectively, the identified features of best linear predictors with missing outcome and covariate data, and finite games with multiple pure strategy Nash equilibria. HMPS provide sharp identification regions, but these may have prohibitive computational cost. In order to make progress not only on identification analysis but also on finite sample inference, ABJ and CT propose regions of parameter values which are not sharp.

Establishing whether a conjectured region for the identified features of an incomplete econometric model is sharp is a key question in identification analysis. Given the joint distribution of the observed variables, a researcher asks herself what parameters θ are consistent with this distribution. The sharp identification region is the collection of parameter values that could generate the same distribution of observables as the one in the data, for some data generating process consistent with the maintained assumptions. Examples of sharp identification regions for parameters of incomplete models are given in Manski (1989, 2003), Manski and Tamer (2002), and Molinari (2008), among others. In some cases, researchers are only able to characterize a region in the parameter space that includes all the parameter values that may have generated the observables, but may include other (infeasible) parameter values as well. These larger regions are called *outer regions*. The inclusion in the outer regions of parameter values which are infeasible may weaken the researcher's ability to make useful predictions, and to test for model misspecification.

Using the theory of random sets (Molchanov (2005)), we provide a general methodology that allows us to characterize the sharp identification region for the parameters of models with convex moment predictions in a computationally tractable manner. Our main insight is that for a given candidate value of θ, the (conditional or unconditional) *Aumann expectation* of a properly defined θ-dependent *random closed set* coincides with the convex set of model moment predictions.[2] That is, this Aumann expectation gives the convex set, implied by the candidate θ, of moments for the relevant variables which are consistent with *all* the model's implications. This is a crucial advancement compared to the related literature, where researchers are often unable to fully exploit the information provided by the model that they are studying, and work with just a subset of model's implications. In turn, this advancement allows us to characterize the sharp identification region of θ, denoted Θ_I, through a simple necessary and sufficient condition. Assume that the model is correctly specified. Then θ is in Θ_I if and only if the conditional Aumann expectation (a convex

---

[2] We formally define the notion of random closed set in Appendix A and the notion of conditional Aumann expectation in Section 2.

set) of the properly defined random set associated with θ, contains the conditional expectation of a properly defined vector of random variables observed in the data (a point).[3] The methodology that we propose allows us to verify this condition by checking whether the *support function* of such point is dominated by the support function of the θ-dependent convex set.[4] The latter can be evaluated exactly or approximated by simulation, depending on the complexity of the model. Showing that this dominance holds amounts to checking whether the difference between the support function of a point (a linear function) and the support function of a convex set (a sublinear function) in a direction given by a vector u attains a maximum of zero as u ranges in the unit ball of appropriate dimension. This amounts to maximizing a superlinear function over a convex set, a task which can be carried out efficiently using algorithms in convex programming (e.g., Boyd and Vandenberghe (2004), Grant and Boyd (2008)).

It is natural to wonder which model with set valued predictions may *not* belong to the class of models to which our methodology applies. Our approach is specifically tailored towards frameworks where Θ_I can be characterized via conditional or unconditional expectations of observable random vectors and model predictions.[5] Within these models, if restrictions are imposed on the selection process, non-convex sets of moments may result. We are chiefly interested in the case that no untestable assumptions are imposed on the selection process, and therefore exploring identification in models with non-convex moment predictions is beyond the scope of this paper.

There are no precedents to our general characterization of the sharp identification region of models with convex moment predictions. However, there is one precedent to the use of the Aumann expectation as a key tool to describe fundamental features of partially identified models. This is the work of Beresteanu and Molinari (2006, 2008), who were the first to illustrate the benefits of using elements of random sets theory to conduct identification analysis and statistical inference for incomplete econometric models in the space of sets, in a manner which is the exact analog of how these tasks are commonly performed for point identified models in the space of vectors.

In related work, Galichon and Henry (2009a) study finite games of complete information with multiple *pure* strategy Nash equilibria. They characterize the sharp identification region of θ

---

[3] This is because when such condition is satisfied, there exists a vector of conditional expectations associated with θ that is consistent with all the implications of the model, and coincides with the vector of conditional expectations observed in the data.

[4] "The support function [of a nonempty closed convex set B in direction u] $h(B,u)$ is the signed distance of the support plane to B with exterior normal vector u from the origin; the distance is negative if and only if u points into the open half space containing the origin," Schneider (1993, page 37). See Rockafellar (1970, Chapter 13) or Schneider (1993, Section 1.7) for a thorough discussion of the support function of a closed convex set, and its properties.

[5] In Section 2.2 we explain what mathematical features of conditional Aumann expectations yield the computational tractability that is novel to our approach.

through the *capacity functional* (i.e., the "probability distribution") of the random set of pure strategy equilibrium outcomes, by exploiting a result due to Artstein (1983).[6] They also show that under certain monotonicity restrictions, further computational simplifications may be obtained by using tools of optimal transportation theory. With pure strategies only, the characterization based on the capacity functional is "dual" to ours, as we formally establish in Beresteanu, Molchanov, and Molinari (2010b, Appendix B.2). It cannot, however, be extended to the general case where mixed strategies are allowed for, as discussed also in Galichon and Henry (2009a, Section 4), nor to other solution concepts such as, for example, correlated equilibrium. Hence, in order to deal with more general game theoretic models, Galichon and Henry (2009a) apply our methodology.

While our main contribution lies in the identification analysis that we carry out, our characterization leads to an obvious sample analog counterpart which can be used when the researcher is confronted with a finite sample of observations. This sample analog is given by the set of minimizers of a sample criterion function. We establish that the methodology of Andrews and Shi (2009) can be applied in our context, to obtain confidence sets that uniformly cover each element of the sharp identification region with a prespecified asymptotic probability. Related methods for statistical inference in partially identified models include, among others, Chernozhukov, Hong, and Tamer (2004, 2007), Pakes, Porter, Ho, and Ishii (2006), Beresteanu and Molinari (2008), Rosen (2008), Chernozhukov, Lee, and Rosen (2009), Galichon and Henry (2009b), Kim (2009), Andrews and Soares (2010), Bugni (2010), Canay (2010), and Romano and Shaikh (2010).

Structure of the Paper. In Section 2 we describe formally the class of econometric models to which our methodology applies, and we provide our characterization of the sharp identification region. In Section 3 we analyze in detail the identification problem in static, simultaneous move finite games of complete information in the presence of multiple mixed strategy Nash equilibria (MSNE), and show how the results of Section 2 can be applied. In Section 4, we show that our approach easily applies to finite games of incomplete information, and characterizes Θ_I through a finite number of moment inequalities. In Section 5 we show how the results of Section 2 can be applied to best linear prediction with interval outcome and covariate data. Section 6 concludes. Appendix A contains definitions taken from random sets theory, proofs of the results appearing in

---

[6] Galichon and Henry (2006) use the notion of capacity functional of a properly defined random set and the results of Artstein (1983), to provide a specification test for partially identified structural models, thereby extending the Kolmogorov-Smirnov test of correct model specification to partially identified models. They then define the notion of "core determining" classes of sets, to find a manageable class of sets for which to check that the dominance condition is satisfied. Beresteanu and Molinari (2006, 2008) use Artstein's (1983) result to establish sharpness of the identification region of the parameters of a best linear predictor with interval outcome data.

the main text,[7] and details concerning the computational issues associated with our methodology (for concreteness, we focus on the case of finite games of complete information).

Appendices B-D are given in the Supplement to this paper, Beresteanu, Molchanov, and Molinari (2010b). Appendix B specializes our results for the case that players are restricted to use pure strategies only and Nash equilibrium is the solution concept. In this case, Θ_I is characterized through a finite number of moment inequalities, and further insights are provided on how to reduce the number of inequalities to be checked in order to compute it. Appendix C shows that our methodology is applicable to static simultaneous move finite games regardless of the solution concept used.[8] Appendix D applies the results of Section 2 to the analysis of individual decision making in random utility models of multinomial choice in the presence of interval regressors data.

2 Semi-parametric Models with Convex Moment Predictions

Notation. Throughout the paper, we use capital Latin letters to denote sets and random sets. We use lower case Latin letters for random vectors. We denote parameter vectors and sets of parameter vectors, respectively by θ and Θ. For a given finite set W, we denote by κ_W its cardinality. We denote by Δ^{d-1} the unit simplex in $\Re^d$. Given two non-empty sets $A, B \subset \Re^d$, we denote the directed Hausdorff distance from A to B, the Hausdorff distance between A and B, and the Hausdorff norm of B, respectively, by

$$d_H(A,B) = \sup_{a\in A}\inf_{b\in B}\|a-b\|, \quad \rho_H(A,B) = \max\{d_H(A,B), d_H(B,A)\}, \quad \|B\|_H = \sup_{b\in B}\|b\|.$$

Outline. In this Section we describe formally the class of econometric models to which our methodology applies, and we provide our characterization of the sharp identification region. In Sections 3, 4 and 5 we illustrate how empirically relevant models fit into this general framework. In particular, we show how to verify, for these models, the assumptions listed below.

2.1 Framework

Consider an econometric model which specifies a vector z of random variables observable by the researcher, a vector ξ of random variables unobservable by the researcher, and an unknown parameter vector $\theta \in \Theta \subset \Re^p$, with Θ the parameter space. Maintain the following assumptions:

---

[7] The only proof appearing in the main text is that of our fundamental result, Theorem 2.1.

[8] Specifically, we illustrate this by looking at games where rationality of level-1 is the solution concept (a problem first studied by Aradillas-Lopez and Tamer (2008)), and by looking at games where correlated equilibrium is the solution concept.

Assumption 2.1 (Probability Space) *The random vectors* (z, ξ) *are defined on a probability space* $(\Omega, \mathfrak{F}, \mathbf{P})$. *The* σ*-algebra* $\mathfrak{F}$ *is generated by* (z, ξ). *The researcher conditions her analysis upon a sub-*σ*-algebra of* $\mathfrak{F}$, *denoted* $\mathfrak{G}$, *which is generated by a subvector of* z. *The probability space contains no* $\mathfrak{G}-$*atoms. Specifically,* $\forall\ A \in \mathfrak{F}$ *having positive measure, there is a* $B \subseteq A$ *such that* $0 < \mathbf{P}(B|\mathfrak{G}) < \mathbf{P}(A|\mathfrak{G})$ *with positive probability.*

Assumption 2.2 (Set Valued Predictions) *For a given value of* θ, *the model maps each realization of* (z, ξ) *to a non-empty closed set* $Q_\theta(z, \xi) \subset \Re^d$. *The functional form of this map is known to the researcher.*

Assumption 2.3 (Absolutely Integrable Random Closed Set) *For every compact set* C *in* $\Re^d$ *and all* $\theta \in \Theta$,

$$\{\omega \in \Omega : Q_\theta(z(\omega), \xi(\omega)) \cap C \neq \emptyset\} \in \mathfrak{F}.$$

Moreover, $\mathbf{E}(\|Q_\theta(z, \xi)\|_H) < \infty$.

Assumption 2.1 requires the probability space to be non-atomic with respect to the σ-algebra $\mathfrak{G}$ upon which the researcher conditions her analysis. This technical assumption is not restrictive for most economic applications, as we show in Sections 3, 4 and 5. For example, it is satisfied whenever the distribution of ξ conditional on $\mathfrak{G}$ is continuous.

Assumption 2.2 requires the model to have set valued predictions.[9] As we further explain below, the set $Q_\theta(z, \xi)$ is the fundamental object that we use to relate the convex set of model moment predictions, to the observable moments of random vectors. In Sections 3, 4 and 5 we provide examples of how $Q_\theta(z, \xi)$ needs to be constructed in specific applications to exploit all the model information.

Assumption 2.3 is a measurability condition, requiring $Q_\theta(z, \xi)$ to be an *integrably bounded random closed set*, see Definitions A.1-A.2 in Appendix A. It guarantees that any ($\mathfrak{F}$-measurable) random vector q such that $q(\omega) \in Q_\theta(z(\omega), \xi(\omega))$ *a.s.* is absolutely integrable.

In what follows, for ease of notation, we write the set $Q_\theta(z, \xi)$ and its realizations, respectively, as Q_θ and $Q_\theta(\omega) \equiv Q_\theta(z(\omega), \xi(\omega))$, $\omega \in \Omega$, omitting the explicit reference to z and ξ. The researcher wishes to learn θ from the observed distribution of z. Because the model makes set valued predictions, we maintain the following assumption:

---

[9] A model which makes singleton predictions with probability one is a special case of the more general class of models analyzed here.

Assumption 2.4 (Selected Prediction) *The econometric model can be augmented with an unknown selection mechanism which selects one of the model predictions, yielding a map* ψ *which depends on* z *and* ξ*, and may depend on* θ*, and which satisfies the following conditions:*
(i) $\psi(z(\omega), \xi(\omega), \theta) \in Q_\theta(\omega)$ *for almost all* $\omega \in \Omega$*;*
(ii) $\psi(z(\omega), \xi(\omega), \theta)$ *is* $\mathfrak{F}$*-measurable for all* $\theta \in \Theta$*.*

Assumption 2.4 requires that the econometric model can be "completed" with an unknown selection mechanism. Economic theory often provides no guidance on the form of the selection mechanism, which therefore we leave completely unspecified. For each $\omega \in \Omega$, the process of selection results in a random element ψ which takes values in Q_θ, i.e., is a model's selected prediction.[10] The map ψ is unknown and constitutes a nonparametric component of the model; it may depend on unobservable variables even after conditioning on observable variables. We insert θ as an argument of ψ to reflect the fact that Assumption 2.4-(i) requires ψ to belong to the θ-dependent set Q_θ.

In this paper we restrict attention to models where the set of observationally equivalent parameter vectors θ, denoted Θ_I, can be characterized via conditional expectations of observable random vectors and model predictions. One may find many values for the parameter vector θ which, when coupled with maps ψ satisfying Assumption 2.4, generate the same moments as the ones observed in the data. Hence, we assume that Θ_I can be characterized through selected predictions as follows:

Assumption 2.5 (Sharp Identification Region) *Given the available data and Assumptions 2.1-2.3, the sharp identification region of* θ *is*

$$\Theta_I = \left\{\theta \in \Theta : \exists\ \psi(z,\xi,\theta) \text{ satisfying Assumption 2.4, s.t. } \mathbf{E}\left(w(z)\middle|\,\mathfrak{G}\right) = \mathbf{E}\left(\psi(z,\xi,\theta)\middle|\,\mathfrak{G}\right) \quad a.s.\right\}, \tag{2.1}$$

where $w(\cdot)$ *is a known function mapping* z *into vectors in* $\Re^d$ *and* $\mathbf{E}(w(z)|\,\mathfrak{G})$ *is identified by the data.*

The process of "unrestricted selection" yielding ψ's satisfying Assumption 2.4 builds all possible mixtures of elements of Q_θ. When one takes expectations of these mixtures, the resulting set of expectations is the *convex set of moment predictions*:

$$\{\mathbf{E}\left(\psi(z,\xi,\theta)|\,\mathfrak{G}\right) : \psi(z,\xi,\theta) \text{ satisfies Assumption 2.4}\}.$$

Convexity of this set is formally established in the next Section.

---

[10] For expository clarity, we observe that even for $\omega_1 \neq \omega_2$ such that $z(\omega_1) = z(\omega_2)$ and $\xi(\omega_1) = \xi(\omega_2)$, $\psi(z(\omega_1), \xi(\omega_1), \theta)$ may differ from $\psi(z(\omega_2), \xi(\omega_2), \theta)$.

Using the notion of selected prediction, Assumption 2.5 characterizes abstractly the sharp identification region of a large class of incomplete econometric models, in a fairly intuitive manner. This characterization builds on previous ones given by Berry and Tamer (2007) and Tamer (2009, Section 3). However, because ψ is a rather general random function, it may constitute an infinite dimensional nuisance parameter, which creates great difficulties for the computation of Θ_I and for inference. In this paper, we provide a complementary approach based on tools of random sets theory. We characterize Θ_I avoiding altogether the need to deal with ψ, thereby contributing to a stream of previous literature which has provided tractable characterizations of sharp identification regions without making any reference to the selection mechanism or the selected prediction (see, e.g., Manski (2003) and Manski and Tamer (2002)).

2.2 Representation Through Random Sets Theory

As suggested by Aumann (1965), one can think of a random closed set (or random correspondence in Aumann's work) as a bundle of random variables – its *measurable selections*, see Definition A.3 in Appendix A. We follow this idea, and denote by $\mathrm{Sel}\,(Q_\theta)$ the collection of $\mathfrak{F}$-measurable random elements q with values in $\Re^d$ such that $q(\omega) \in Q_\theta\,(\omega)$ for almost all $\omega \in \Omega$. As it turns out, there is not just a simple assonance between "selected prediction" and "measurable selection." Our first result establishes a one-to-one correspondence between them.

Lemma 2.1 *Let Assumptions 2.1-2.3 hold. For any given $\theta \in \Theta$, $q \in \mathrm{Sel}\,(Q_\theta)$ if and only if there exists a selected prediction $\psi(z, \xi, \theta)$ satisfying Assumption 2.4, such that $q\,(\omega) = \psi(z\,(\omega)\,, \xi\,(\omega)\,, \theta)$ for almost all $\omega \in \Omega$.*

The definition of the sharp identification region in Assumption 2.5 indicates that one needs to take conditional expectations of the elements of $\mathrm{Sel}\,(Q_\theta)$. Observe that by Assumption 2.3, Q_θ is an integrably bounded random closed set, and therefore all its selections are integrable. Hence, we can define the conditional Aumann expectation (Aumann (1965)) of Q_θ as

$$\mathbb{E}\left(Q_\theta \middle| \mathfrak{G}\right) = \left\{\mathbf{E}\left(q \middle| \mathfrak{G}\right) : q \in \mathrm{Sel}\,(Q_\theta)\right\},$$

where the notation $\mathbb{E}\left(\cdot \middle| \mathfrak{G}\right)$ denotes the conditional Aumann expectation of the random set in parentheses, while we reserve the notation $\mathbf{E}\left(\cdot \middle| \mathfrak{G}\right)$ for the conditional expectation of a random vector. By Theorem 2.1.46 in Molchanov (2005) the conditional Aumann expectation exists and is unique. Because $\mathfrak{F}$ contains no $\mathfrak{G}$-atoms, and because the random set Q_θ takes its realizations in a subset of the finite dimensional space $\Re^d$, it follows from Theorem 1.2 of Dynkin and Evstigneev

(1976) and from Theorem 2.1.24 of Molchanov (2005), that $\mathbb{E}(Q_\theta|\mathfrak{G})$ is a closed convex set *a.s.*, such that $\mathbb{E}(Q_\theta|\mathfrak{G}) = \mathbb{E}(\mathrm{co}[Q_\theta]|\mathfrak{G})$, with $\mathrm{co}[\cdot]$ the convex hull of the set in square brackets.

Our second result establishes that $\mathbb{E}(Q_\theta|\mathfrak{G})$ coincides with the convex set of model's moment predictions:

Lemma 2.2 *Let Assumptions 2.1-2.3 hold. For any given* $\theta \in \Theta$, $\mathbb{E}(Q_\theta|\mathfrak{G}) = \{\mathbf{E}(\psi(z,\xi,\theta)|\mathfrak{G}) :$ $\psi(z,\xi,\theta)$ *satisfies Assumption 2.4*$\}$, *and therefore the latter set is convex.*

Hence, the set of observationally equivalent parameter values in Assumption 2.5 can be written as

$$\Theta_I = \{\theta \in \Theta : \mathbf{E}(w(z)|\mathfrak{G}) \in \mathbb{E}(Q_\theta|\mathfrak{G}) \quad a.s.\}$$

The fundamental characterization result of this paper is the following:

Theorem 2.1 *Let Assumptions 2.1-2.5 be satisfied, and no other information be available. Let* $h(Q_\theta,u) \equiv \sup_{q\in Q_\theta} u'q$ *denote the support function of* Q_θ *in direction* $u \in \Re^d$. *Then*

$$\Theta_I = \left\{\theta \in \Theta : \max_{u\in B}\left(u'\mathbf{E}(w(z)|\mathfrak{G}) - \mathbf{E}[h(Q_\theta,u)|\mathfrak{G}]\right) = 0 \quad a.s.\right\} \tag{2.2}$$

$$= \left\{\theta \in \Theta : \textstyle\int \left(u'\mathbf{E}(w(z)|\mathfrak{G}) - \mathbf{E}[h(Q_\theta,u)|\mathfrak{G}]\right)_+ d\mathcal{U} = 0 \quad a.s.\right\}, \tag{2.3}$$

where $B = \left\{u \in \Re^d : \|u\| \leq 1\right\}$, $\mathcal{U}$ *is a probability measure on* B *with support equal to* B, *and for any* $a \in \Re$, $(a)_+ = \max\{0,a\}$.

Proof. The equivalence between equations (2.2)-(2.3) follows immediately, observing that the integrand in equation (2.3) is continuous in u and both conditions inside the curly brackets are satisfied if and only if

$$u'\mathbf{E}(w(z)|\mathfrak{G}) - \mathbf{E}[h(Q_\theta,u)|\mathfrak{G}] \leq 0 \;\forall u \in B \quad a.s. \tag{2.4}$$

In order to establish sharpness, it suffices to show that for a given $\theta \in \Theta$ expression (2.4) holds if and only if $\theta \in \Theta_I$ as defined in equation (2.1). Take $\theta \in \Theta$ such that expression (2.4) holds. Theorem 2.1.47-(iv) in Molchanov (2005) assures that

$$\mathbf{E}[h(Q_\theta,u)|\mathfrak{G}] = h(\mathbb{E}(Q_\theta|\mathfrak{G}),u) \;\;\forall u \in \Re^d \quad a.s. \tag{2.5}$$

Recalling that the support function is positive homogeneous, equation (2.4) holds if and only if

$$u'\mathbf{E}(w(z)|\mathfrak{G}) \leq h(\mathbb{E}(Q_\theta|\mathfrak{G}),u) \;\;\forall u \in \Re^d \quad a.s. \tag{2.6}$$

Standard arguments in convex analysis (see, e.g. Rockafellar (1970, Theorem 13.1)) assure that equation (2.6) holds if and only if $\mathbf{E}\left(w\left(z\right)|\,\mathfrak{G}\right) \in \mathbb{E}\left(Q_\theta|\,\mathfrak{G}\right)$ *a.s.*, and therefore by Lemma 2.2 $\theta \in \Theta_I$. Conversely, take $\theta \in \Theta_I$ as defined in equation (2.1). Then there exists a selected prediction ψ satisfying Assumption 2.4, such that $\mathbf{E}\left(w\left(z\right)|\,\mathfrak{G}\right) = \mathbf{E}\left(\psi(z,\xi,\theta)|\,\mathfrak{G}\right)$. By Lemma 2.2 and the above argument, it follows that expression (2.4) holds. ■

It is well known (e.g., Rockafellar (1970, Chapter 13), Schneider (1993, Section 1.7)) that the support function of a non-empty closed convex set is a continuous convex sublinear function.[11] This holds also for the support function of the convex set of moment predictions. However, calculating this set is computationally prohibitive in many cases. The fundamental simplification comes from equation (2.5), which assures that one can work directly with the conditional expectation of $h\left(Q_\theta, u\right)$. This expectation is quite straightforward to compute. Hence, the characterization in equation (2.2) is computationally very attractive, because for each candidate $\theta \in \Theta$ it requires to maximize an easy-to-compute superlinear, hence concave, function over a convex set, and check if the resulting objective value is equal to zero. This problem is computationally tractable and several efficient algorithms in convex programming are available to solve it, see for example the book by Boyd and Vandenberghe (2004), and the MatLab software for disciplined convex programming CVX by Grant and Boyd (2010). Similarly, the characterization in equation (2.3) can be implemented by calculating integrals of concave functions over a convex set, a task which can be carried out in random polynomial time (see, e.g. Dyer, Frieze, and Kannan (1991) and Lovász and Vempala (2006)).

Remark 2.1 Using the method proposed by Andrews and Shi (2009), expression (2.4) can be transformed, using appropriate instruments, into a set of unconditional moment inequalities indexed by the instruments and by $u \in B$, even when the conditioning variables have a continuous distribution. Equations (2.2)-(2.3) can be modified accordingly, to yield straightforward criterion functions which are minimized by every parameter in the sharp identification region. When faced with a finite sample of data, one can obtain a sample analog of these criterion functions by replacing the unconditional counterpart of the moment $u'\mathbf{E}\left(w\left(z\right)|\,\mathfrak{G}\right) - \mathbf{E}\left[h\left(Q_\theta, u\right)|\,\mathfrak{G}\right]$ with its sample analog. The resulting statistics correspond,[12] respectively, to the Kolmogorov-Smirnov (KS) and

---

[11] In particular, for a given set $A \subset \Re^d$, $h\left(A, u+v\right) \leq h\left(A,u\right) + h\left(A,v\right)$ for all $u, v \in \Re^d$ and $h\left(A, cu\right) = ch\left(A,u\right)$ for all $c > 0$ and for all $u \in \Re^d$. Additionally, one can show that the support function of a bounded set $A \subset \Re^d$ is Lipschitz with Lipschitz constant $\|A\|_H$, see Molchanov (2005, Theorem F.1).

[12] Because $u = 0 \in B$, the function $\left(u'\mathbf{E}\left(g\left(z\right)|\,\mathfrak{G}\right) - \mathbf{E}\left[h\left(Q_\theta, u\right)|\,\mathfrak{G}\right]\right)$ and its positive part achieve the same maximum value for $u \in B$. Andrews and Shi's test statistics are obtained by replacing $\left(u'\mathbf{E}\left(g\left(z\right)|\,\mathfrak{G}\right) - \mathbf{E}\left[h\left(Q_\theta, u\right)|\,\mathfrak{G}\right]\right)_+$ with its square, and by choosing appropriate instruments (as detailed in Andrews and Shi's Section 3) to transform

the Cramér-von Mises (CvM) statistics introduced by Andrews and Shi (2009, see their equations (3.6), (3.7), (3.8), and their Section 9). When the assumptions imposed by Andrews and Shi are satisfied, one can obtain confidence sets that have correct uniform asymptotic coverage probability for the true parameter vector, by inverting the KS or the CvM tests. Under mild regularity conditions, these assumptions are satisfied using our characterization, because our moment function in expression (2.4) is Lipschitz in u. In Appendix A.3 we formally establish this for the models in Sections 3, 4 and 5.[13]

3 Application I: Finite Games of Complete Information

3.1 Model Set-Up

We consider simultaneous-move games of complete information (normal form games) in which each player has a finite set of actions (pure strategies) $\mathcal{Y}_j$, $j = 1, \ldots, J$, with J the number of players. Let $y = (y_1, \ldots, y_J) \in \mathcal{Y}$ denote a generic vector specifying an action for each player, with $\mathcal{Y} = \times_{j=1}^{J} \mathcal{Y}_j$ and $\mathcal{Y}_{-j} = \times_{i \neq j} \mathcal{Y}_i$. Let $\pi_j\left(y_j, y_{-j}, x_j, \varepsilon_j, \theta\right)$ denote the payoff function for player j, where y_{-j} is the vector of player j's opponents' actions, $x_j \in \mathcal{X}$ is a vector of observable payoff shifters, ε_j is a payoff shifter observed by the players but unobserved by the econometrician, and $\theta \in \Theta \subset \Re^p$ is a vector of parameters of interest, with Θ the parameter space. Let $\sigma_j : \mathcal{Y}_j \to [0, 1]$ denote the mixed strategy for player j that assigns to each action $y_j \in \mathcal{Y}_j$ a probability $\sigma_j\left(y_j\right) \geq 0$ that it is played, with $\sum_{y_j \in \mathcal{Y}_j} \sigma_j\left(y_j\right) = 1$ for each $j = 1, \ldots, J$. Let $\Sigma\left(\mathcal{Y}_j\right)$ denote the mixed extension of $\mathcal{Y}_j$, and $\Sigma\left(\mathcal{Y}\right) = \times_{j=1}^{J} \Sigma\left(\mathcal{Y}_j\right)$. With the usual slight abuse of notation, denote by $\pi_j\left(\sigma_j, \sigma_{-j}, x_j, \varepsilon_j, \theta\right)$ the expected payoff associated with the mixed strategy profile $\sigma = \left(\sigma_1, \ldots, \sigma_J\right)$. With respect to the general notation used in Section 2, $z = (y, \underline{x})$, $\xi = \varepsilon$, $\mathfrak{F}$ is the σ-algebra generated by $\left(y, \underline{x}, \varepsilon\right)$, and $\mathfrak{G}$ is the σ-algebra generated by $\underline{x}$. We formalize our assumptions on the games and sampling processes as follows. These assumptions are fairly standard in the literature.[14]

---

the conditional moment inequalities in unconditional ones. See Appendix A.3 for details.

[13] Imbens and Manski (2004, see also Stoye (2009)) discuss the difference between confidence sets that uniformly cover the true parameter vector with a prespecified asymptotic probability, and confidence sets that uniformly cover Θ_I. Providing methodologies to obtain asymptotically valid confidence sets of either type when the conditioning variables have a continuous distribution, is a developing area of research, to which the method of Andrews and Shi (2009) belongs. In certain empirically relevant models (see for example Section 4 and Beresteanu, Molchanov, and Molinari (2010b, Appendix B and Appendix D)) the characterization in Theorem 2.1 yields a finite number of (conditional) moment inequalities. In such cases, the methods of Chernozhukov, Hong, and Tamer (2007) and Romano and Shaikh (2010) can be applied after discretizing the conditioning variables, to obtain confidence sets which cover Θ_I with a prespecified asymptotic probability, uniformly in the case of Romano and Shaikh (2010). Ciliberto and Tamer (2009) verify the required regularity conditions for finite games of complete information.

[14] We assume that players' actions and the outcomes observable by the econometrician coincide. This is a standard assumption in the literature, see e.g. ABJ, CT, Berry and Tamer (2007) and Bajari, Hong, and Ryan (2009). Our

Assumption 3.1 *(i) The set of outcomes of the game $\mathcal{Y}$ is finite. Each player j has $\kappa_{\mathcal{Y}_j} \geq 2$ pure strategies to choose from. The number of players is $J \geq 2$.*
(ii) The observed outcome of the game results from static, simultaneous move, Nash play.
(iii) The parametric form of the payoff functions $\pi_j(y_j, y_{-j}, x_j, \varepsilon_j, \theta)$, $j = 1, \ldots, J$, is known, and for a known action $\bar{y}$ it is normalized to $\pi_j(\bar{y}_j, \bar{y}_{-j}, x_j, \varepsilon_j, \theta) = 0$ for each j. The payoff functions are continuous in x_j and ε_j. The parameter space Θ is compact.[15]

Assumption 3.2 *The econometrician observes data that identify $\mathbf{P}(y|\underline{x})$. The observed matrix of payoff shifters $\underline{x}$ is comprised of the non-redundant elements of x_j, $j = 1, \ldots, J$. The unobserved random vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J)$ has a continuous conditional distribution function $F_\theta(\varepsilon|\underline{x})$ that is known up to a finite dimensional parameter vector that is part of θ.*

Remark 3.1 Under Assumption 3.2, Assumption 2.1 is satisfied.

It is well known that the games and sampling processes satisfying Assumptions 3.1-3.2 may lead to multiple MSNE and partial identification of the model parameters, see for example Berry and Tamer (2007) for a thorough discussion of this problem. To achieve point identification, Bjorn and Vuong (1985), Bresnahan and Reiss (1988, 1990, 1991), Berry (1992), Mazzeo (2002), Tamer (2003), and Bajari, Hong, and Ryan (2009), for example, add assumptions concerning the nature of competition, heterogeneity of firms, availability of covariates with sufficiently large support and/or instrumental variables, and restrictions on the selection mechanism which, in the data generating process, determines the equilibrium played in the regions of multiplicity.[16]

We show that the models considered in this Section satisfy Assumptions 2.1-2.5, and therefore our methodology gives a computationally feasible characterization of Θ_I. Our approach does not impose any assumption on the nature of competition, on the form of heterogeneity across players, or on the selection mechanism. It does not require availability of covariates with large support or instruments, but fully exploits their identifying power if they are present.

---

results, however, apply to the more general case that the strategy profiles determine the outcomes observable by the econometrician through an outcome rule known by the econometrician, as we illustrate with a simple example in Beresteanu, Molchanov, and Molinari (2010b, Appendix B.1). Of course, the outcome rule needs to satisfy assumptions guaranteeing that it conveys some information about players actions.

[15] Continuity is needed to establish measurability and closedness of certain sets. A location normalization is needed because if we add a constant to the payoff of each action, the set of equilibria does not change.

[16] Tamer (2003) also suggests an approach to partially identify the model's parameters when no additional assumptions are imposed.

3.2 The Sharp Identification Region

For a given realization of $(\underline{x},\varepsilon)$, the mixed strategy profile $\sigma=(\sigma_1,\ldots,\sigma_J)$ constitutes a Nash equilibrium if $\pi_j(\sigma_j,\sigma_{-j},x_j,\varepsilon_j,\theta)\geq\pi_j(\tilde{\sigma}_j,\sigma_{-j},x_j,\varepsilon_j,\theta)\ \forall\tilde{\sigma}_j\in\Sigma(\mathcal{Y}_j),\ \forall j=1,...,J.$ Hence, for a given realization of $(\underline{x},\varepsilon)$ we define the θ-dependent set of MSNE as

$$S_\theta(\underline{x},\varepsilon)=\{\sigma\in\Sigma(\mathcal{Y}):\pi_j(\sigma_j,\sigma_{-j},x_j,\varepsilon_j,\theta)\geq\pi_j(\tilde{\sigma}_j,\sigma_{-j},x_j,\varepsilon_j,\theta)\ \forall\tilde{\sigma}_j\in\Sigma(\mathcal{Y}_j)\ \forall j\}. \tag{3.1}$$

Example 3.1 *Consider a simple two player entry game similar to the one in Tamer (2003), omit the covariates, assume that players' payoffs are given by* $\pi_j=y_j(y_{-j}\theta_j+\varepsilon_j)$, *where* $y_j\in\{0,1\}$ *and* $\theta_j<0$, $j=1,2$. *Let* $\sigma_j\in[0,1]$ *denote the probability that player* j *enters the market, with* $1-\sigma_j$ *the probability that he does not. Figure 1-(a) plots the set of mixed strategy equilibrium profiles* $S_\theta(\varepsilon)$ *resulting from the possible realizations of* $\varepsilon_1,\varepsilon_2$. $\square$

For ease of notation we write the set $S_\theta(\underline{x},\varepsilon)$ and its realizations, respectively, as S_θ and $S_\theta(\omega)\equiv S_\theta(\underline{x}(\omega),\varepsilon(\omega))$, $\omega\in\Omega$, omitting the explicit reference to $\underline{x}$ and ε. Proposition 3.1 establishes that the set S_θ is a random closed set in $\Sigma(\mathcal{Y})$.

Proposition 3.1 *Let Assumption 3.1 hold. Then the set* S_θ *is a random closed set in* $\Sigma(\mathcal{Y})$ *as per Definition A.1 in Appendix A.*

For a given $\theta\in\Theta$ and $\omega\in\Omega$, each element $\sigma(\omega)\equiv(\sigma_1(\omega),\ldots,\sigma_J(\omega))\in S_\theta(\omega)$ is one of the admissible mixed strategy Nash equilibrium profiles associated with the realizations $\underline{x}(\omega)$ and $\varepsilon(\omega)$, and it takes values in $\Sigma(\mathcal{Y})$. The resulting random elements $\sigma=\{\sigma(\omega),\omega\in\Omega\}$ are the selections of S_θ, denoted $\mathrm{Sel}(S_\theta)$, see Definition A.3 in Appendix A.

Example 3.1 (Cont.) *Consider the set* S_θ *plotted in Figure 1-(a). Let* $\Omega^M=\{\omega\in\Omega:\varepsilon(\omega)\in[0,-\theta_1]\times[0,-\theta_2]\}$. *Then for* $\omega\notin\Omega^M$ *the set* S_θ *has only one selection, since the equilibrium is unique. For* $\omega\in\Omega^M$, S_θ *contains a rich set of selections, which can be obtained as*

$$\sigma(\omega)=(\sigma_1(\omega),\sigma_2(\omega))=\begin{cases}(1,0) & \text{if } \omega\in\Omega_1^M,\\ \left(\frac{\varepsilon_2(\omega)}{-\theta_2},\frac{\varepsilon_1(\omega)}{-\theta_1}\right) & \text{if } \omega\in\Omega_2^M,\\ (0,1) & \text{if } \omega\in\Omega_3^M,\end{cases}$$

for all measurable disjoint $\Omega_i^M\subset\Omega^M$, $i=1,2,3$, *such that* $\Omega_1^M\cup\Omega_2^M\cup\Omega_3^M=\Omega^M$. $\square$

By definition of a mixed strategy profile, $\sigma_j(\omega):\mathcal{Y}_j\to[0,1]$ assigns to each action $t_j\in\mathcal{Y}_j$ a probability $\sigma_j(\omega,t_j)\geq0$ that it is played, with $\sum_{t_j\in\mathcal{Y}_j}\sigma_j(\omega,t_j)=1$, $j=1,\ldots,J$. Index the set $\mathcal{Y}=\times_{j=1}^J\mathcal{Y}_j$ in some (arbitrary) way, such that $\mathcal{Y}=\{t^1,\ldots,t^{\kappa_\mathcal{Y}}\}$ and $t^k\equiv(t_1^k,\ldots,t_J^k)$,

$k = 1, \ldots, \kappa_{\mathcal{Y}}$. Then for a given parameter value $\theta \in \Theta$ and realization $\sigma(\omega)$, $\omega \in \Omega$, of a selection $\sigma \in \mathrm{Sel}(S_\theta)$, the implied probability that y is equal to t^k is given by $\prod_{j=1}^{J} \sigma_j\left(\omega, t_j^k\right)$. Hence, we can use a selection $\sigma \in \mathrm{Sel}(S_\theta)$ to define a random vector $q(\sigma)$ whose realizations have coordinates

$$\left([q(\sigma(\omega))]_k = \prod_{j=1}^{J} \sigma_j\left(\omega, t_j^k\right), \quad k = 1, \ldots, \kappa_{\mathcal{Y}}\right). \tag{3.2}$$

By construction, the random point $q(\sigma)$ is an element of $\Delta^{\kappa_{\mathcal{Y}}-1}$. For given $\omega \in \Omega$, each vector $\left([q(\sigma(\omega))]_k, k = 1, \ldots, \kappa_{\mathcal{Y}}\right)$ is the multinomial distribution over outcomes of the game (a J-tuple of actions) determined by the mixed strategy equilibrium $\sigma(\omega)$. Repeating the above construction for each $\sigma \in \mathrm{Sel}(S_\theta)$, we obtain

$$Q_\theta = \left\{\left([q(\sigma)]_k, k = 1, \ldots, \kappa_{\mathcal{Y}}\right) : \sigma \in \mathrm{Sel}(S_\theta)\right\}. \tag{3.3}$$

Remark 3.2 The set $Q_\theta \equiv Q_\theta(\underline{x}, \varepsilon)$ satisfies Assumption 2.2 by construction. By Proposition 3.1, Q_θ is a random closed set in $\Delta^{\kappa_{\mathcal{Y}}-1}$, because it is given by a continuous map applied to the random closed set S_θ. Because every realization of $q \in \mathrm{Sel}(Q_\theta)$ is contained in $\Delta^{\kappa_{\mathcal{Y}}-1}$, Q_θ is integrably bounded. Hence, Assumption 2.3 is satisfied.

Example 3.1 (Cont.) *Consider the set S_θ plotted in Figure 1-(a). Index the set $\mathcal{Y}$ so that $\mathcal{Y} = \{(0,0), (1,0), (0,1), (1,1)\}$. Then*

$$Q_\theta = \left\{ q(\sigma) = \begin{bmatrix} (1-\sigma_1)(1-\sigma_2) \\ \sigma_1(1-\sigma_2) \\ (1-\sigma_1)\sigma_2 \\ \sigma_1\sigma_2 \end{bmatrix} : \sigma \in \mathrm{Sel}(S_\theta) \right\}.$$

Figure 1-(b) plots the set Q_θ resulting from the possible realizations of $\varepsilon_1, \varepsilon_2$. $\square$

Because Q_θ is an integrably bounded random closed set, all its selections are integrable and its conditional Aumann expectation is

$$\begin{aligned} \mathbb{E}(Q_\theta | \underline{x}) &= \left\{\mathbf{E}(q | \underline{x}) : q \in \mathrm{Sel}(Q_\theta)\right\} \\ &= \left\{\left(\mathbf{E}\left([q(\sigma)]_k | \underline{x}\right), k = 1, \ldots, \kappa_{\mathcal{Y}}\right) : \sigma \in \mathrm{Sel}(S_\theta)\right\}. \end{aligned}$$

Example 3.1 (Cont.) *Consider the set Q_θ plotted in Figure 1-(b). Let $\Omega^M = \{\omega \in \Omega : \varepsilon(\omega) \in [0, -\theta_1] \times [0, -\theta_2]\}$. Then for $\omega \notin \Omega^M$ the set Q_θ has only one selection, since the equilibrium is unique. For $\omega \in \Omega^M$, the selections of Q_θ are:*

$$q(\sigma(\omega)) = \begin{cases} [0\ 1\ 0\ 0]' & \text{if } \omega \in \Omega_1^M, \\ q\left(\frac{\varepsilon_2(\omega)}{-\theta_2}, \frac{\varepsilon_1(\omega)}{-\theta_1}\right) & \text{if } \omega \in \Omega_2^M, \\ [0\ 0\ 1\ 0]' & \text{if } \omega \in \Omega_3^M, \end{cases}$$

for all measurable partitions $\left\{\Omega_i^M\right\}_{i=1}^3$ *of* Ω^M. *In the above expression,*

$$q\left(\frac{\varepsilon_2(\omega)}{-\theta_2}, \frac{\varepsilon_1(\omega)}{-\theta_1}\right) = \left[\left(1-\frac{\varepsilon_2(\omega)}{-\theta_2}\right)\left(1-\frac{\varepsilon_1(\omega)}{-\theta_1}\right) \quad \frac{\varepsilon_2(\omega)}{-\theta_2}\left(1-\frac{\varepsilon_1(\omega)}{-\theta_1}\right) \quad \left(1-\frac{\varepsilon_2(\omega)}{-\theta_2}\right)\frac{\varepsilon_1(\omega)}{-\theta_1} \quad \frac{\varepsilon_2(\omega)}{-\theta_2}\frac{\varepsilon_1(\omega)}{-\theta_1}\right]'$$

The expectations of the selections of Q_θ *build the set* $\mathbb{E}(Q_\theta)$, *which is a convex subset of* Δ^3 *with infinitely many extreme points.* $\square$

The set $\mathbb{E}(Q_\theta|\underline{x})$ collects vectors of probabilities with which each outcome of the game can be observed. It is obtained by integrating the probability distribution over outcomes of the game implied by each mixed strategy equilibrium σ given $\underline{x}$ and ε (that is, by integrating each element of $\text{Sel}(Q_\theta)$), against the probability measure of $\varepsilon|\underline{x}$. We emphasize that in case of multiplicity, a different mixed strategy equilibrium $\sigma(\omega) \in S_\theta(\omega)$ may be selected (with different probability) for each ω. By construction, $\mathbb{E}(Q_\theta|\underline{x})$ is the set of probability distributions over action profiles conditional on $\underline{x}$ which are consistent with the maintained modeling assumptions, i.e., with *all* the model's implications. In other words, it is the convex set of moment predictions.

If the model is correctly specified, there exists at least one value of $\theta \in \Theta$ such that the observed conditional distribution of y given $\underline{x}$, $\mathbf{P}(y|\underline{x})$, is a point in the set $\mathbb{E}(Q_\theta|\underline{x})$ for $\underline{x} - a.s.$, where $\mathbf{P}(y|\underline{x}) \equiv \left[\mathbf{P}\left(y = t^k\middle|\underline{x}\right), k = 1, \ldots, \kappa_{\mathcal{Y}}\right]$.[17] Hence, the set of observationally equivalent parameter values which form the sharp identification region is given by

$$\Theta_I = \{\theta \in \Theta : \mathbf{P}(y|\underline{x}) \in \mathbb{E}(Q_\theta|\underline{x}) \ \ \underline{x} - a.s.\} \tag{3.4}$$

Theorem 3.2 *Let Assumptions 3.1-3.2 hold, and no other information be available. Then*

$$\begin{aligned} \Theta_I &= \left\{\theta \in \Theta : \max_{u \in B}\left(u'\mathbf{P}(y|\underline{x}) - \mathbf{E}\left[h(Q_\theta, u)|\underline{x}\right]\right) = 0 \ \underline{x} - a.s.\right\} && (3.5)\\ &= \left\{\theta \in \Theta : \int \left(u'\mathbf{P}(y|\underline{x}) - \mathbf{E}\left[h(Q_\theta, u)|\underline{x}\right]\right)_+ d\mathcal{U} = 0 \ \underline{x} - a.s.\right\} && (3.6)\end{aligned}$$

where $h(Q_\theta, u) = \max_{q \in Q_\theta} u'q = \max_{\sigma \in S_\theta} \sum_{k=1}^{\kappa_{\mathcal{Y}}} u_k \prod_{j=1}^J \sigma_j\left(t_j^k\right)$ *and* $u' = [u_1 \ u_2 \ \ldots \ u_{\kappa_{\mathcal{Y}}}]$.[18]

Theorem 3.2 follows immediately from Theorem 2.1, because Assumptions 2.1-2.5 are satisfied for this application, as summarized in Remarks 3.1, 3.2, and 3.3 (the latter given below).

By Wilson's (1971) result, the realizations of the set of MSNE, S_θ, are almost surely finite sets. Therefore, the same holds for Q_θ. Hence, for given $\omega \in \Omega$, $h(Q_\theta(\omega), u)$ is given by the

---

[17] By the definition of $\mathbb{E}(Q_\theta|\underline{x})$, $\mathbf{P}(y|\underline{x}) \in \mathbb{E}(Q_\theta|\underline{x})$ if and only if $\exists\, q \in \text{Sel}(Q_\theta) : \mathbf{E}(q|\underline{x}) = \mathbf{P}(y|\underline{x})$.

[18] Recall that B is the unit ball in $\Re^{\kappa_{\mathcal{Y}}}$ and $\mathcal{U}$ is a probability measure on B with support equal to B. Recall also that $\mathcal{Y} = \{t^1, t^2, ..., t^{\kappa_{\mathcal{Y}}}\}$ is the set of possible outcomes of the game, and $t^k \equiv \left(t_1^k, \ldots, t_J^k\right)$ is a J-tuple specifying one action in $\mathcal{Y}_j$ for each player $j = 1, ..., J$.

maximum among the inner product of u with a finite number of vectors, the elements of $Q_\theta(\omega)$. These elements are known functions of $(\underline{x}(\omega), \varepsilon(\omega))$. Hence, given Q_θ, the expectation of $h(Q_\theta, u)$ is easy to compute.

Example 3.1 (Cont.) *Consider the set* Q_θ *plotted in Figure 1-(b). Pick a direction* $u \equiv [u_1\ u_2\ u_3\ u_4]' \in B$. *Then, for* $\omega \in \Omega$ *such that* $\varepsilon(\omega) \in (-\infty, 0] \times (-\infty, 0]$, *we have* $Q_\theta(\omega) = \left\{[1\ \ 0\ \ 0\ \ 0]'\right\}$, *and* $h(Q_\theta(\omega), u) = u_1$. *For* $\omega \in \Omega$ *such that* $\varepsilon(\omega) \in [0, -\theta_1] \times [0, -\theta_2]$, *we have* $Q_\theta(\omega) = \left\{[0\ \ 1\ \ 0\ \ 0]', q\left(\frac{\varepsilon_2(\omega)}{-\theta_2}, \frac{\varepsilon_1(\omega)}{-\theta_1}\right), [0\ \ 0\ \ 1\ \ 0]'\right\}$, *and* $h(Q_\theta(\omega), u) = \max\left(u_2, u'q\left(\frac{\varepsilon_2(\omega)}{-\theta_2}, \frac{\varepsilon_1(\omega)}{-\theta_1}\right), u_3\right)$. *Figure 1-(c) plots* $h(Q_\theta(\omega), u)$ *against the possible realizations of* $\varepsilon_1, \varepsilon_2$. $\square$

By a way of comparison with the previous literature, and to show how Assumptions 2.4-2.5 can be verified, we provide the abstract definition of Θ_I given by Berry and Tamer (2007, equation (2.21), page 67) for the case of a two player entry game, extending it to finite games with potentially more than two players and two actions. A finite game with multiple equilibria can be completed by a random vector which has almost surely non-negative entries that sum up to one, and which gives the probability with which each equilibrium in the regions of multiplicity is played when the game is defined by $(\underline{x}, \varepsilon, \theta)$. Denote such (random) discrete distribution by $\lambda(\cdot; \underline{x}, \varepsilon, \theta) : S_\theta \to \Delta^{\kappa_{S_\theta}-1}$. Notice that $\lambda(\cdot; \underline{x}, \varepsilon, \theta)$ is left unspecified and can depend on market unobservables even after conditioning on market observables. By definition, the sharp identification region includes all the parameter values for which one can find a random vector $\lambda(\cdot; \underline{x}, \varepsilon, \theta)$ satisfying the above conditions, such that the model augmented with this selection mechanism generates the joint distribution of the observed variables. Hence,

(3.7)

$$\Theta_I = \left\{ \theta \in \Theta : \begin{array}{l} \exists\ \lambda(\cdot; \underline{x}, \varepsilon, \theta) : S_\theta \to \Delta^{\kappa_{S_\theta}-1} \text{ for } (\underline{x}, \varepsilon) - a.s., \text{ such that } \forall k = 1, \ldots, \kappa_{\mathcal{Y}}, \\ \mathbf{P}\left(y = t^k \middle| \underline{x}\right) = \int \left(\sum\limits_{\sigma \in S_\theta(\underline{x}, \varepsilon)} \lambda(\sigma; \underline{x}, \varepsilon, \theta) \prod\limits_{j=1}^{J} \sigma_j\left(t_j^k\right) \right) dF(\varepsilon | \underline{x}) \ \ \underline{x} - a.s. \end{array} \right\}$$

Notice that with respect to the general notation used in Section 2, $w(z) = \left[1\left(y = t^k\right), k = 1, \ldots, \kappa_{\mathcal{Y}}\right]$. Finally, observe that using $\lambda(\cdot; \underline{x}, \varepsilon, \theta)$ one can construct a selected prediction $\psi(\underline{x}, \varepsilon, \theta)$ as a random vector whose realizations given $\underline{x}$ and ε are equal to

$$\left[\prod\nolimits_{j=1}^{J} \sigma_j\left(t_j^k\right), \quad k = 1, \ldots, \kappa_{\mathcal{Y}}\right], \quad \text{with probability } \lambda(\sigma; \underline{x}, \varepsilon, \theta), \quad \sigma \in \mathrm{Sel}(S_\theta).$$

Remark 3.3 The random vector $\psi(\underline{x}, \varepsilon, \theta)$ is a selected prediction satisfying Assumption 2.4. Observing that $\mathbf{E}(\psi(\underline{x}, \varepsilon, \theta) | \underline{x}) = \int \left[\sum_{\sigma \in S_\theta(\underline{x}, \varepsilon)} \lambda(\sigma; \underline{x}, \varepsilon, \theta) \prod_{j=1}^{J} \sigma_j\left(t_j^k\right), \ k = 1, \ldots, \kappa_{\mathcal{Y}}\right] dF(\varepsilon | \underline{x})$, where the integral is taken coordinate-wise, Assumption 2.5 is verified.

Remark 3.4 Appendix A.3 verifies Andrews and Shi's (2009) regularity conditions for models satisfying Assumptions 3.1-3.2, under the additional assumption that the researcher observes an i.i.d. sequence of equilibrium outcomes and observable payoff shifters $\{y_i, \underline{x}_i\}_{i=1}^n$. Andrews and Shi's (2009) generalized moment selection procedure with infinitely many conditional moment inequalities can therefore be applied, to obtain confidence sets that have correct uniform asymptotic coverage.

3.3 Comparison with the Outer Regions of ABJ and CT

While ABJ and CT discuss only the case that players are restricted to use pure strategies, it is clear and explained in Berry and Tamer (2007, pp. 65-70) that their insights can be extended to the case that players are allowed to randomize over their strategies. Here we discuss the relationship between such extensions, and the methodology that we propose.[19]

In the presence of multiple equilibria, ABJ observe that an implication of the model is that for a given $t^k \in \mathcal{Y}$, $\mathbf{P}\left(y = t^k \middle| \underline{x}\right)$ cannot be larger than the probability that t^k is *a possible* equilibrium outcome of the game. This is because for given $\theta \in \Theta$ and realization of $(\underline{x}, \varepsilon)$ such that t^k is a possible equilibrium outcome of the game, there can be another outcome $t^l \in \mathcal{Y}$ which is also a possible equilibrium outcome of the game, and when both are possible t^k is selected only part of the time. CT point out that additional information can be learned from the model. In particular, $\mathbf{P}\left(y = t^k \middle| \underline{x}\right)$ cannot be smaller than the probability that t^k is the *unique* equilibrium outcome of the game. This is because t^k is certainly realized whenever it is the only possible equilibrium outcome, but it can additionally be realized when it belongs to a set of multiple equilibrium outcomes.

The following Proposition rewrites the outer regions originally proposed by ABJ and CT, denoted Θ_O^{ABJ} and Θ_O^{CT}, using our notation. It then establishes their connection with Θ_I.

Proposition 3.3 *Let Assumptions 3.1-3.2 hold, and no other information be available. Then the outer regions proposed by ABJ and CT are respectively*

$$\Theta_O^{ABJ} = \left\{\theta \in \Theta : \mathbf{P}\left(y = t^k \middle| \underline{x}\right) \leq \max\left(\int [q(\sigma)]_k \, dF_\theta(\varepsilon | \underline{x}) : \sigma \in \operatorname{Sel}(S_\theta)\right), \ \textit{for } k = 1, ..., \kappa_{\mathcal{Y}}, \ \underline{x} - a.s.\right\} \tag{3.8}$$

$$\Theta_O^{CT} = \left\{\theta \in \Theta : \begin{array}{l} \min\left(\int [q(\sigma)]_k \, dF_\theta(\varepsilon | \underline{x}) : \sigma \in \operatorname{Sel}(S_\theta)\right) \leq \mathbf{P}\left(y = t^k \middle| \underline{x}\right) \leq \\ \max\left(\int [q(\sigma)]_k \, dF_\theta(\varepsilon | \underline{x}) : \sigma \in \operatorname{Sel}(S_\theta)\right), \ \textit{for } k = 1, ..., \kappa_{\mathcal{Y}}, \ \underline{x} - a.s. \end{array}\right\} \tag{3.9}$$

Θ_O^{ABJ} *can be obtained by solving the maximization problem in equation (3.5) over the restricted set of u's equal to the canonical basis vectors in* $\Re^{\kappa_{\mathcal{Y}}}$. Θ_O^{CT} *can be obtained by solving the maximization*

---

[19] Beresteanu, Molchanov, and Molinari (2009, Section 3.3) revisit Example 3.1 in light of this comparison.

problem in equation (3.5) over the restricted set of u's equal to the canonical basis vectors in $\Re^{\kappa_\mathcal{Y}}$ *and each of these vectors multiplied by* -1.

Hence, the approaches of ABJ and CT can be interpreted on the base of our analysis as follows. For each $\theta \in \Theta$, ABJ's inequalities give the closed half spaces delimited by hyperplanes that are parallel to the axis and that support $\mathbb{E}\left(Q_\theta|\, \underline{x}\right)$. Θ_O^{ABJ} is the collection of θ's such that $\mathbf{P}\left(y|\, \underline{x}\right)$ is contained in the non-negative part of such closed half spaces $\underline{x} - a.s.$ CT use a more refined approach, and for each $\theta \in \Theta$ their inequalities give the smallest hypercube containing $\mathbb{E}\left(Q_\theta|\, \underline{x}\right)$. Θ_O^{CT} is the collection of θ's such that $\mathbf{P}\left(y|\, \underline{x}\right)$ is contained in such hypercube $\underline{x} - a.s.$ The more $\mathbb{E}\left(Q_\theta|\, \underline{x}\right)$ differs from the hypercubes used by ABJ and CT, the more likely it is that a candidate value θ belongs to Θ_O^{ABJ} and Θ_O^{CT}, but not to Θ_I. A graphical intuition for this relationship is given in Figure 2.

3.4 Two Player Entry Game – An Implementation

This section presents an implementation of our method, and a series of numerical illustrations of the identification gains that it affords, in the two player entry game in Example 3.1, both with and without covariates in the payoff functions. The set S_θ for this example (omitting $\underline{x}$) is plotted in Figure 1. Appendix A.4 provides details on the method used to compute Θ_O^{ABJ}, Θ_O^{CT} and Θ_I.

For all the data generating processes (DGPs) we let $(\varepsilon_1, \varepsilon_2) \overset{iid}{\sim} N(0,1)$. The DGPs without covariates are designed as follows. We build a grid of 36 equally spaced values for $\theta_1^\star, \theta_2^\star$ on $[-1.8, -0.8] \times [-1.7, -0.7]$, yielding multiple equilibria with a probability that ranges from substantial (0.21), to small (0.07). We match each point on the $\theta_1^\star, \theta_2^\star$ grid, with each point on a grid of 10 values for $\lambda^\star$, the probability distribution over equilibria in the region of multiplicity.[20] This results in 360 distinct DGPs, each with a corresponding vector $[\mathbf{P}(y = t),\ t \in \{(0,0),(1,0),(0,1),(1,1)\}]$. We compute Θ_I, Θ_O^{CT} and Θ_O^{ABJ} for each DGP, letting the parameter space be $\Theta = [-4.995, -0.005]^2$. We then rank the results, according to $\frac{L(\mathrm{Proj}(\Theta_I|1))+L(\mathrm{Proj}(\Theta_I|2))}{L(\mathrm{Proj}(\Theta_O^{CT}|1))+L(\mathrm{Proj}(\Theta_O^{CT}|2))}$, where $\mathrm{Proj}(\cdot|i)$ is the projection of the set in parentheses on dimension i, and $L(\mathrm{Proj}(\cdot|i))$ is the length of such projection. To conserve space, in Table 1 we report only the results of our "top 15% reduction," "median reduction," and "bottom 15% reduction."[21] Figure 3 plots Θ_I, Θ_O^{CT} and Θ_O^{ABJ} for each of these DGPs.

---

[20] The grid on $\lambda^\star$ is constructed by letting $\lambda^\star\left((0,1)\right)$ take values 0, 0.25, 0.5, 0.75, $\lambda^\star\left((1,0)\right)$ take values in $[0, 0.75 - \lambda^\star\left((0,1)\right)]$ with step size 0.25, and by letting $\lambda^\star\left(\left(\frac{\varepsilon_2}{-\theta_2}, \frac{\varepsilon_1}{-\theta_1}\right)\right) = 1 - \lambda^\star\left((0,1)\right) - \lambda^\star\left((1,0)\right)$.

[21] The full set of results is available from the authors upon request. Our best result has a 97% reduction in size of Θ_I compared to Θ_O^{CT}. Our worst result has a 20% reduction in size of Θ_I compared to Θ_O^{CT}. Only 6% of the DGPs yield a reduction in size of Θ_I compared to Θ_O^{CT} of less than 25%.

To further illustrate the computational feasibility of our methodology, we allow for covariates in the payoff functions. Specifically, we let $\pi_j = y_j \left(y_{-j}\theta_j + \beta_{0j} + x_{1j}\beta_{1j} + x_{2j}\beta_{2j} + \varepsilon_j \right)$, $j = 1, 2$, where $[x_{11} \;\; x_{21}]$, the covariates for player 1, take four different values, $\{[-2 \;\; 1], [1 \;\; -1.5], [0 \;\; 0.75], [-1.5 \;\; -1]\}$ and $[x_{12} \;\; x_{22}]$, the covariates for player 2, take five different values, $\{[1 \;\; -1.75], [-1.25 \;\; 1], [0 \;\; 0], [0.6 \;\; 0.5], [0.5 \;\; -0.5]\}$. The parameter vector of interest is $\theta = [(\theta_j \; \beta_{0j} \; \beta_{1j} \; \beta_{2j})_{j=1,2}]$. In generating $\mathbf{P}(y|\underline{x})$, we use the values of $\lambda^{\star}$ and $\theta_1^{\star}, \theta_2^{\star}$ which yield the "top 15% reduction," "median reduction," and "bottom 15% reduction" in the DGPs with no x variables, and pair them with $[\beta_{01}^{\star} \; \beta_{11}^{\star} \; \beta_{21}^{\star}] = [0 \;\; 1/2 \;\; 1/3]$ and $[\beta_{02}^{\star} \; \beta_{12}^{\star} \; \beta_{22}^{\star}] = [0 \;\; -1/3 \;\; -1/2]$. This results in three different DGPs. Compared to the case with no covariates, for each of these DGPs the computational time required to verify whether a candidate θ is in Θ_I is linear in the number of values that $\underline{x}$ can take. The reductions in size of Θ_I compared to the outer regions of ABJ and CT is of similar magnitude to the case with no covariates. Table 2 reports the results.

4 Application II: Entry Games of Incomplete Information

We now consider the case that players have incomplete information (see, e.g. Aradillas-López (2010), Brock and Durlauf (2001, 2007), Seim (2006), Sweeting (2009)). We retain the notation introduced in Section 3, but we substitute Assumption 3.1 with the following one, which is fairly standard in the literature.[22] We continue to maintain Assumption 3.2.

Assumption 4.1 *(i) The set of outcomes of the game $\mathcal{Y}$ is finite. The observed outcome of the game results from simultaneous move, pure strategy Bayesian Nash play.*
(ii) All players and the researcher observe payoff shifters x_j, $j = 1, ..., J$. The payoff shifter ε_j is private information to player $j = 1, ..., J$, and unobservable to the researcher. Conditional on $\{x_j, j = 1, ..., J\}$, ε_j is independent of $\{\varepsilon_i\}_{i \neq j}$. Players have correct common prior $F_\theta(\varepsilon|\underline{x})$.
(iii) The payoffs are additively separable in ε : $\pi_j(y_j, y_{-j}, x_j, \varepsilon_j; \theta) = \tilde{\pi}_j(y_j, y_{-j}, x_j; \theta) + \varepsilon_j$. Assumption 3.1-(iii) holds.

For the sake of brevity, we restrict attention to two player entry games. However, this restriction is not necessary. Our results easily extend, with appropriate modifications to the notation and the definition of the set of pure strategy Bayesian Nash Equilibria (BNE), to the case of $J \geq 2$ players each with $2 \leq \kappa_{\mathcal{Y}_j} < \infty$ strategies. In what follows, we characterize the set of BNE of the game,

---

[22] The independence condition in Assumption 4.1-(iii) substantially simplifies the task of calculating the set of BNE. Conceptually, however, our methodology applies also when players' types are correlated. The resulting difficulties associated with calculating the set of BNE are to be faced with any methodology for inference in this class of games. The correct-common-prior condition in Assumption 4.1-(iii) can be relaxed, but we maintain it here for simplicity.

borrowing from the treatment in Grieco (2009, Section 4), and then apply our methodology to this set.[23] To conserve space, we do not explicitly verify Assumptions 2.1-2.5. Assumptions 2.1-2.3 follow by similar arguments as in Section 3. Assumptions 2.4-2.5 follow by the same construction that we provide at the end of Section 3, replacing equation (3.7) with equation (8) in Grieco (2009, Theorem 4).

With incomplete information, players' strategies are decision rules $y_j : \mathcal{E} \rightarrow \{0,1\}$, with $\mathcal{E}$ the support of ε. The set of outcomes of the game is $\mathcal{Y} = \{(0,0),(1,0),(0,1),(1,1)\}$. Given $\theta \in \Theta$ and a realization of $\underline{x}$ and ε_j, player j enters the market if and only if his expected payoff is non-negative. Therefore, equilibrium mappings (decision rules) are step functions determined by a threshold: $y_j(\varepsilon_j) = 1(\varepsilon_j \geq t_j)$, $j = 1,2$. As a result, player j's beliefs about player $-j$'s probability of entry under the common prior assumption is $\int y_{-j}(\varepsilon_{-j})\, dF_\theta(\varepsilon_{-j}|\underline{x}) = 1 - F_\theta(t_{-j}|\underline{x})$, and therefore player j's best response cutoff is[24]

$$t_j^b(t_{-j},\underline{x};\theta) = -\tilde{\pi}_j(1,0,x_j;\theta)\, F_\theta(t_{-j}|\underline{x}) - \tilde{\pi}_j(1,1,x_j;\theta)\,(1 - F_\theta(t_{-j}|\underline{x})).$$

Hence, the set of equilibria can be defined as the set of cutoff rules:

$$T_\theta(\underline{x}) = \left\{(t_1,t_2):\ t_j = t_j^b(t_{-j},\underline{x};\theta)\ \ \forall\, j = 1,2\right\}.$$

Note that the equilibrium thresholds are functions of $\underline{x}$ only. The set $T_\theta(\underline{x})$ might contain a finite number of equilibria (e.g., if the common prior is the Normal distribution), or a continuum of equilibria. For ease of notation we write the set $T_\theta(\underline{x})$ and its realizations, respectively, as T_θ and $T_\theta(\omega) \equiv T_\theta(\underline{x}(\omega))$, $\omega \in \Omega$.

For given realization of the random variables characterizing the model, i.e., for given $\omega \in \Omega$, we need to map the set of equilibrium decision rules of each player, into outcomes of the game. Consider the realization $t(\omega)$ of $t \in \mathrm{Sel}(T_\theta)$. Through the threshold decision rule, such realization implies the following action profile:

$$q(t(\omega)) = \begin{bmatrix} 1(\varepsilon_1(\omega) \leq t_1(\omega), \varepsilon_2(\omega) \leq t_2(\omega)) \\ 1(\varepsilon_1(\omega) \geq t_1(\omega), \varepsilon_2(\omega) \leq t_2(\omega)) \\ 1(\varepsilon_1(\omega) \leq t_1(\omega), \varepsilon_2(\omega) \geq t_2(\omega)) \\ 1(\varepsilon_1(\omega) \geq t_1(\omega), \varepsilon_2(\omega) \geq t_2(\omega)) \end{bmatrix} \in \Delta^3,$$

with Δ^3 the simplex in $\Re^4$. The vector $q(t(\omega))$ indicates which of the four possible pairs of actions is played with probability 1, when the realization of $(\underline{x},\varepsilon)$ is $(\underline{x}(\omega),\varepsilon(\omega))$ and the equilibrium

---

[23] We refer to Grieco (2009) for a thorough discussion of the related literature and of identification problems in games of incomplete information with multiple BNE. See also Berry and Tamer (2007, Section 3).

[24] For example, with payoffs linear in $\underline{x}$ and given by $\pi(y_j, y_{-j}, \underline{x}, \varepsilon_j;\theta) = y_j(y_{-j}\theta_{1j} + x_j\theta_{2j} + \varepsilon_j)$, we have that player 1 enters if and only if $(\varepsilon_1 + x_1\theta_{21})\, F_\theta(t_2|\underline{x}) + (\varepsilon_1 + x_1\theta_{21} + \theta_{11})(1 - F_\theta(t_2|\underline{x})) \geq 0$. Therefore the cutoff is $t_j^b(t_{-j},\underline{x};\theta) = -x_1\theta_{21}F_\theta(t_2|\underline{x}) - (x_1\theta_{21} + \theta_{11})(1 - F_\theta(t_2|\underline{x})) = -x_1\theta_{21} - \theta_{11}(1 - F_\theta(t_2|\underline{x}))$.

threshold is $t(\omega) \in T_\theta(\underline{x}(\omega))$. Applying this construction to all measurable selections of T_θ, we construct a random closed set in Δ^3 :

$$Q_\theta = \{q(t) : t \in \mathrm{Sel}(T_\theta)\}.$$

For given $\underline{x}$ and $\theta \in \Theta$, define the conditional Aumann expectation

$$\mathbb{E}(Q_\theta|\underline{x}) = \{\mathbf{E}(q(t)|\underline{x}) : t \in \mathrm{Sel}(T_\theta)\}.$$

Notice that for a specific selection $t \in \mathrm{Sel}(T_\theta)$, given the independence assumption on $\varepsilon_1, \varepsilon_2$, the first entry of the vector $\mathbf{E}(q(t)|\underline{x})$ is

$$\mathbf{E}(1(\varepsilon_1 \le t_1, \varepsilon_2 \le t_2)|\underline{x}) = (1 - F_\theta(t_1|\underline{x}))(1 - F_\theta(t_2|\underline{x})),$$

and similarly for the other entries of $\mathbf{E}(q(t)|\underline{x})$. This yields the multinomial distribution over outcome profiles determined by equilibrium threshold $t \in \mathrm{Sel}(T_\theta)$. By the same logic as in Section 3, $\mathbb{E}(Q_\theta|\underline{x})$ is the set of probability distributions over action profiles conditional on $\underline{x}$ which are consistent with the maintained modeling assumptions, i.e., with *all* the model's implications. By the same results that we applied in Section 3, the set $\mathbb{E}(Q_\theta|\underline{x})$ is closed and convex.

Observe that regardless of whether T_θ contains a finite number of equilibria or a continuum, Q_θ can take on only a finite number of realizations,[25] corresponding to each of the vertices of Δ^3. As we show in the proof of Theorem 4.1, this implies that $\mathbb{E}(Q_\theta|\underline{x})$ is a closed convex polytope $\underline{x} - a.s.$, fully characterized by a finite number of supporting hyperplanes. In turn, this allows us to characterize Θ_I through a finite number of moment inequalities, and to compute it using efficient algorithms in linear programming.

Theorem 4.1 *Let Assumptions 4.1 and 3.2 hold, and no other information be available. Then*

$$\begin{aligned} \Theta_I &= \left\{\theta \in \Theta : \max_{u \in B} \left(u'\mathbf{P}(y|\underline{x}) - \mathbf{E}[h(Q_\theta, u)|\underline{x}]\right) = 0 \ \underline{x} - a.s.\right\} \\ &= \left\{\theta \in \Theta : u'\mathbf{P}(y|\underline{x}) \le \mathbf{E}[h(Q_\theta, u)|\underline{x}] \ \forall\, u \in D,\ \underline{x} - a.s.\right\} \end{aligned}$$

where $D = \left\{u = \begin{bmatrix} u_1 & \dots & u_{\kappa_\mathcal{Y}} \end{bmatrix}' : u_i \in \{0,1\},\ i = 1, ..., \kappa_\mathcal{Y}\right\}$.

Remark 4.1 Grieco (2009) introduces an important model, where each player has a vector of payoff shifters unobservable by the researcher. Some of the elements of this vector are private information to the player, while the others are known to all players. Our results in Section 2 apply to this set-up as well, by the same arguments as in Section 3 and in this Section.

---

[25] Hence, the set Q_θ is a "simple" random closed set in Δ^3, see Definition A.4 in Appendix A.

Remark 4.2 Appendix A.3 verifies Andrews and Shi's (2009) regularity conditions for models satisfying Assumptions 4.1 and 3.2, under the additional assumption that the researcher observes an i.i.d. sequence of equilibrium outcomes and observable payoff shifters $\{y_i, \underline{x}_i\}_{i=1}^n$.

5 Application III: Best Linear Prediction with Interval Outcome and Covariate Data

Here we consider the problem of best linear prediction under square loss, when both outcome and covariate data are interval valued.[26] HMPS studied the related problem of identification of the BLP parameters with missing data on outcome and covariates, and provided a characterization of the identification region of each component of the vector θ. While their characterization is sharp, the computational complexity of the problem in the HMPS formulation grows with the number of points in the support of the outcome and covariate variables, and becomes essentially unfeasible if these variables are continuous, unless one discretizes their support quite coarsely. Using the same approach as in the previous part of the paper, we provide a characterization of Θ_I which remains computationally feasible regardless of the support of outcome and covariate variables.

We let $y^\star, x^\star$ denote the unobservable outcome and covariate variables. To simplify the exposition, we let $x^\star$ be scalar, though this assumption can be relaxed and is not essential for our methodology. We maintain the following assumption:

Assumption 5.1 *The researcher does not observe the realizations of* $(y^\star, x^\star)$, *but rather the realizations of real valued random variables* y_L, y_U, x_L, x_U *such that* $\mathbf{P}(y_L \leq y^\star \leq y_U) = 1$ *and* $\mathbf{P}(x_L \leq x^\star \leq x_U) = 1$. $\mathbf{E}(|y_i|)$, $\mathbf{E}(|x_j|)$, $\mathbf{E}(|y_i x_j|)$, *and* $\mathbf{E}\left(x_j^2\right)$ *are all finite, for each* $i, j = L, U$. *One of the following holds: (i) at least one of* $y_L, y_U, x_L, x_U, y^\star, x^\star$ *has a continuous distribution; or (ii)* $(\Omega, \mathfrak{F}, \mathbf{P})$ *is a non-atomic probability space.*[27]

With respect to the general notation used in Section 2, $z = (y_L, y_U, x_L, x_U)$, $\xi = (y^\star, x^\star)$, $\mathfrak{F}$ is the σ-algebra generated by $(y_L, y_U, x_L, x_U, y^\star, x^\star)$. The researcher works with unconditional moments.

Remark 5.1 Under Assumption 5.1, Assumption 2.1 is satisfied.

---

[26] Beresteanu and Molinari (2008) study identification and statistical inference for the parameters $\theta \in \Theta$ of the BLP parameters when only the outcome variable is interval valued. See also Bontemps, Magnac, and Maurin (2008) for related results. Here we significantly generalize their identification results by allowing also for interval valued covariates. This greatly complicates computation of Θ_I and inference, because Θ_I is no longer a linear transformation of an Aumann expectation.

[27] Clearly, case (i) guarantees that $(\Omega, \mathfrak{F}, \mathbf{P})$ is non-aotmic.

When $y^\star$ and $x^\star$ are perfectly observed, it is well known that the BLP problem can be expressed through a linear projection model, where the prediction error associated with the BLP parameters $\theta^\star$ and given by $\varepsilon^\star = y^\star - \theta_1^\star - \theta_2^\star x^\star$ satisfies $\mathbf{E}(\varepsilon^\star) = 0$ and $\mathbf{E}(\varepsilon^\star x^\star) = 0$. For any candidate $\theta \in \Theta$, we extend the construction of the prediction error to the case of interval valued data. We let $Y = [y_L, y_U]$ and $X = [x_L, x_U]$. It is easy to show that these are random closed sets in $\Re$ as per Definition A.1 (see Beresteanu and Molinari (2008, Lemma A.3)). We build the set

$$Q_\theta = \left\{ q = \left[\begin{array}{c} y - \theta_1 - \theta_2 x \\ (y - \theta_1 - \theta_2 x)\, x \end{array} \right] : \ (y, x) \in \mathrm{Sel}(Y \times X) \right\}. \tag{5.1}$$

This is the not necessarily convex θ-dependent set of prediction errors and prediction errors multiplied by covariate which are implied by the intervals Y and X.

Remark 5.2 The set Q_θ satisfies Assumption 2.2 by construction. Because it is given by a continuous map applied to the random closed sets Y and X, Q_θ is a random closed set in $\Re^2$. By Assumption 5.1, the set Q_θ is integrably bounded, see Beresteanu and Molinari (2008, proof of Theorem 4.2). By the Fundamental Selection Theorem (Molchanov (2005, Theorem 1.2.13)) and by Lemma 2.1, there exist selected predictions $\psi(y_L, y_U, x_L, x_U, y^\star, x^\star, \theta)$ that satisfy Assumption 2.4. The last step in the proof of Theorem 5.1, given in Appendix A, establishes that Assumption 2.5 holds.

Given the set Q_θ, one can relate conceptually our approach in Section 2 to the problem that we study here, as follows. For a candidate $\theta \in \Theta$, each selection (y, x) from the random intervals Y and X yields a moment for the prediction error $\varepsilon = y - \theta_1 - \theta_2 x$ and its product with the covariate x. The collection of such moments for all $(y, x) \in \mathrm{Sel}(Y \times X)$ is equal to the (unconditional) Aumann expectation $\mathbb{E}(Q_\theta) = \{\mathbf{E}(q) : q \in \mathrm{Sel}(Q_\theta)\}$. Because the probability space is non-atomic and Q_θ belongs to a finite dimensional space, $\mathbb{E}(Q_\theta)$ is a closed convex set. If $\mathbb{E}(Q_\theta)$ contains the vector $[0 \;\; 0]'$ as one of its elements, then the candidate value of θ is one of the observationally equivalent parameters of the BLP of $y^\star$ given $x^\star$. This is because if the condition just mentioned is satisfied, then for the candidate $\theta \in \Theta$ there exists a selection in $\mathrm{Sel}(Y \times X)$, that is, a pair of admissible random variables y and x, which implies a prediction error that has mean zero and is uncorrelated with x, hence satisfying the requirements for the BLP prediction error.[28] This intuition is formalized in Theorem 5.1.

---

[28] Notice that with respect to the general notation used in Section 2, $w(z) = [0 \;\; 0]'$.

Theorem 5.1 *Let Assumption 5.1 hold, and no other information be available. Then*

$$\begin{aligned}\Theta_I &= \left\{\theta \in \Theta : \max_{u \in B} \left(-\mathbf{E}\left[h\left(Q_\theta, u\right)\right]\right) = 0\right\} \\ &= \left\{\theta \in \Theta : \int \left(\mathbf{E}\left[h\left(Q_\theta, u\right)\right]\right)_{-} d\mathcal{U} = 0\right\}.\end{aligned}$$

The support function of Q_θ can be easily calculated. In particular, for any $u = [u_1 \;\; u_2]' \in B$,

$$h\left(Q_\theta, u\right) = \max_{q \in Q_\theta} u'q = \max_{y \in Y, x \in X} \left[u_1\left(y - \theta_1 - \theta_2 x\right) + u_2\left(yx - \theta_1 x - \theta_2 x^2\right)\right]. \tag{5.2}$$

For given $\theta \in \Theta$ and $u \in B$, this maximization problem can be efficiently solved using the gradient method, regardless of whether $\left(y_i, x_i\right)_{i=L,U}$, $\left(y^\star, x^\star\right)$ are continuous or discrete random variables. Hence, $h\left(Q_\theta, u\right)$ is an easy to calculate continuous-valued convex sublinear function of u. Membership of a candidate θ to the set Θ_I can be verified using efficient algorithms in convex programming, or taking integrals of concave functions.

Remark 5.3 Appendix A.3 verifies Andrews and Shi's (2009) regularity conditions for models satisfying Assumption 5.1, under the additional assumption that the researcher observes an i.i.d. sequence $\{y_{iL}, y_{iU}, x_{iL}, x_{iU}\}_{i=1}^n$ and that these have finite fourth moments.

6 Conclusions

This paper introduces a computationally feasible characterization for the sharp identification region Θ_I of the parameters of incomplete econometric models with convex moment predictions. Our approach is based on characterizing, for each $\theta \in \Theta$, the set of moments which are consistent with *all* the model's implications, as the (conditional) Aumann expectation of a properly defined random set. If the model is correctly specified, one can then build Θ_I as follows. A candidate θ is in Θ_I if and only if it yields a conditional Aumann expectation which, for $\underline{x} - a.s.$, contains the relevant expectations of random variables observed in the data. Because in general, for each $\theta \in \Theta$, the conditional Aumann expectation may have infinitely many extreme points, characterizing the set Θ_I entails checking that an infinite number of moment inequalities are satisfied. However, we show that this computational hardship can be avoided, and the sharp identification region can be characterized as the set of parameter values for which the maximum of an easy-to-compute superlinear (hence concave) function over the unit ball is equal to zero. We show that finite sample inference can be carried out adopting the generalized moment selection procedure with infinitely many conditional moment inequalities recently proposed by Andrews and Shi (2009). We exemplify our methodology by applying it to empirically relevant models, for which a feasible characterization of Θ_I was absent in the literature.

We acknowledge that the method proposed in this paper may, for some models, be computationally more intensive than existing methods (e.g., ABJ and CT in the analysis of finite games of complete information with multiple equilibria). However, advanced computational methods in convex programming made available in recent years, along with the use of parallel processing, can substantially alleviate this computational burden. On the other hand, the benefits in terms of identification yielded by our methodology may be substantial, as illustrated in our examples.

A Proofs and Auxiliary Results for Sections 3-5

A.1 Definitions

The theory of random closed sets generally applies to the space of closed subsets of a locally compact Hausdorff second countable topological space $\mathbb{F}$ (e.g., Molchanov (2005)). For the purposes of this paper it suffices to consider $\mathbb{F} = \Re^d$, which simplifies the exposition. Denote by $\mathcal{F}$ the family of closed subsets of $\Re^d$.

Definition A.1 *A map $Z : \Omega \rightarrow \mathcal{F}$ is called a* ***random closed set****, also known as a closed set valued random variable, if for every compact set K in $\Re^d$, $\{\omega \in \Omega : Z(\omega) \cap K \neq \emptyset\} \in \mathfrak{F}$.*

Definition A.2 *A random closed set $Z : \Omega \rightarrow \mathcal{F}$ is called* ***integrably bounded*** *if $\|Z\|_H = \sup\{\|z\| : z \in Z\}$ has a finite expectation.*

Definition A.3 *Let Z be a random closed set in $\Re^d$. A random element z with values in $\Re^d$ is called a (measurable)* ***selection*** *of Z if $z(\omega) \in Z(\omega)$ for almost all $\omega \in \Omega$. The family of all selections of Z is denoted by* $\mathrm{Sel}(Z)$.

Definition A.4 *A random closed set Z in $\mathcal{F}$ is called* ***simple*** *if it assumes at most a finite number of values, so that there exists a finite measurable partition $\Omega_1, \ldots, \Omega_m$ of Ω and sets $K_1, ..., K_m \in \mathcal{F}$ such that $Z(\omega) = K_i$ for all $\omega \in \Omega_i$, $1 \leq i \leq m$.*

A.2 Proofs

Proof of Lemma 2.1. For any given $\theta \in \Theta$, if $\psi(z, \xi, \theta)$ is a selected prediction, then $\psi(z, \xi, \theta)$ is a random element as a composition of measurable functions, and it belongs to Q_θ for almost all $\omega \in \Omega$ by Assumption 2.4-(i). Conversely, for any given $\theta \in \Theta$ let $q \in \mathrm{Sel}(Q_\theta)$. Because q is $\mathfrak{F}$-measurable, by the Doob-Dynkin Lemma (see, e.g., Rao and Swift (2006, Proposition 3, Chapter 1)) q can be represented as a measurable function of z and ξ, which is then the selected prediction and satisfies conditions (i)-(ii) in Assumption 2.4. This selected prediction can also be obtained using a selection mechanism which picks a prediction equal to $q(\omega)$ for each $\omega \in \Omega$. ■

Proof of Lemma 2.2. For any given $\theta \in \Theta$, let $\mu \in \mathbb{E}(Q_\theta | \mathfrak{G})$. Then by the definition of the conditional Aumann expectation, there exists a $q \in \mathrm{Sel}(Q_\theta)$ such that $\mathbf{E}(q | \mathfrak{G}) = \mu$. By Lemma 2.1 there exists a $\psi(z, \xi, \theta)$ satisfying Assumption 2.4 such that $q(\omega) = \psi(z(\omega), \xi(\omega), \theta)$ for almost all $\omega \in \Omega$, and therefore $\mu \in \{\mathbf{E}(\psi(z, \xi, \theta) | \mathfrak{G}) : \psi(z, \xi, \theta)$ satisfies Assumption 2.4$\}$. A similar argument yields the reverse conclusion. ■

Proof of Proposition 3.1. Write the set S_θ as follows:

$$S_\theta = \bigcap_{j=1}^{J} \left\{\sigma \in \Sigma(\mathcal{Y}) : \pi_j(\sigma_j, \sigma_{-j}, x_j, \varepsilon_j, \theta) \geq \tilde{\pi}_j(\sigma_{-j}, x_j, \varepsilon_j, \theta)\right\},$$

where $\tilde{\pi}_j(\sigma_{-j}, x_j, \varepsilon_j, \theta) = \sup_{\tilde{\sigma}_j \in \Sigma(\mathcal{Y}_j)} \pi_j(\tilde{\sigma}_j, \sigma_{-j}, x_j, \varepsilon_j, \theta)$. Since $\pi_j(\sigma_j, \sigma_{-j}, x_j, \varepsilon_j, \theta)$ is a continuous function of $\sigma, x_j, \varepsilon_j$, its supremum $\tilde{\pi}_j(\sigma_{-j}, x_j, \varepsilon_j, \theta)$ is a continuous function. Continuity in x_j, ε_j follows from

Assumption 3.1-(iii). Continuity in σ follows because by definition

$$\pi_j\left(\sigma, x_j, \varepsilon_j, \theta\right) \equiv \sum_{t^k \in \mathcal{Y}} \left[\prod_{j=1}^{J} \sigma_j\left(t_j^k\right)\right] \pi_j\left(t^k, x_j, \varepsilon_j, \theta\right),$$

where $t^k \equiv \left(t_1^k, \ldots, t_J^k\right)$, $k = 1, \ldots, \kappa_{\mathcal{Y}}$ and $\mathcal{Y}$ can be ordered arbitrarily so that $\mathcal{Y} = \left\{t^1, \ldots, t^{\kappa_{\mathcal{Y}}}\right\}$. Therefore S_θ is the finite intersection of sets defined as solutions of inequalities for continuous (random) functions. Thus, S_θ is a random closed set, see Molchanov (2005, Section 1.1). ■

Proof of Proposition 3.3. To see that the expression in equation (3.8) is the outer region proposed by ABJ, observe that $\max\left(\int \left[q\left(\sigma\right)\right]_k dF\left(\varepsilon \middle| \underline{x}\right) : \sigma \in \operatorname{Sel}\left(S_\theta\right)\right)$ gives the probability that t^k is a possible equilibrium outcome of the game according to the model. It is obtained by selecting with probability one, in each region of multiplicity, the mixed strategy profile which yields the highest probability that t^k is the outcome of the game. To see that the expression in equation (3.9) is the outer region proposed by CT, observe that $\min\left(\int \left[q\left(\sigma\right)\right]_k dF\left(\varepsilon \middle| \underline{x}\right) : \sigma \in \operatorname{Sel}\left(S_\theta\right)\right)$ gives the probability that t^k is the unique equilibrium outcome of the game according to the model. It is obtained by selecting with probability one, in each region of multiplicity, the mixed strategy profile which yields the lowest probability that t^k is the outcome of the game.

To obtain Θ_O^{ABJ} by solving the maximization problem in equation (3.5) over the restricted set of u's equal to the canonical basis vectors in $\Re^{\kappa_{\mathcal{Y}}}$, take the vector $u^k \in \Re^{\kappa_{\mathcal{Y}}}$ to have all entries equal to zero except entry k which is equal to one. Then

$$\mathbf{P}\left(y = t^k \middle| \underline{x}\right) = u^{k\prime} \mathbf{P}\left(y \middle| \underline{x}\right) \leq h\left(\mathbb{E}\left(Q_\theta \middle| \underline{x}\right), u^k\right) = \max\left(\mathbf{E}\left(\left[q\left(\sigma\right)\right]_k \middle| \underline{x}\right) : \sigma \in \operatorname{Sel}\left(S_\theta\right)\right).$$

To obtain Θ_O^{CT} by solving the maximization problem in equation (3.5) over the restricted set of u's equal to the canonical basis vectors in $\Re^{\kappa_{\mathcal{Y}}}$ and each of these vectors multiplied by -1, observe that the statement for the upper bound follows by the argument given above when considering Θ_O^{ABJ}. To verify the statement for the lower bound, take the vector $-u^k \in \Re^{\kappa_{\mathcal{Y}}}$ to have all entries equal to zero except entry k which is equal to minus one. Then

$$\begin{aligned} -\mathbf{P}\left(y = t^k \middle| \underline{x}\right) &= -u^{k\prime} \mathbf{P}\left(y \middle| \underline{x}\right) \\ &\leq h\left(\mathbb{E}\left(Q_\theta \middle| \underline{x}\right), \left(-u^k\right)\right) = h\left(-\mathbb{E}\left(Q_\theta \middle| \underline{x}\right), u^k\right) \\ &= -\min\left(\int \left[q\left(\sigma\right)\right]_k dF\left(\varepsilon \middle| \underline{x}\right) : \sigma \in \operatorname{Sel}\left(S_\theta\right)\right). \end{aligned}$$

Equivalently, taking u to be a vector with each entry equal to 1, except entry k which is set to 0, one has that

$$\begin{aligned} 1 - \mathbf{P}\left(y = t^k \middle| \underline{x}\right) &= u' \mathbf{P}\left(y \middle| \underline{x}\right) \leq h\left(\mathbb{E}\left(Q_\theta \middle| \underline{x}\right), u\right) = \max\left(\textstyle\sum_{i \neq k} \int \left[q\left(\sigma\right)\right]_i dF\left(\varepsilon \middle| \underline{x}\right) : \sigma \in \operatorname{Sel}\left(S_\theta\right)\right) \\ &= \max\left(1 - \int \left[q\left(\sigma\right)\right]_k dF\left(\varepsilon \middle| \underline{x}\right) : \sigma \in \operatorname{Sel}\left(S_\theta\right)\right) = 1 - \min\left(\int \left[q\left(\sigma\right)\right]_k dF\left(\varepsilon \middle| \underline{x}\right) : \sigma \in \operatorname{Sel}\left(S_\theta\right)\right). \end{aligned}$$

■

Proof of Theorem 4.1. By the same argument as in the proof of Theorem 2.1,

$$\begin{aligned} \Theta_I &= \left\{\theta \in \Theta : \mathbf{P}\left(y \middle| \underline{x}\right) \in \mathbb{E}\left(Q_\theta \middle| \underline{x}\right),\ \underline{x} - a.s.\right\} \\ &= \left\{\theta \in \Theta : \max_{u \in B}\left(u' \mathbf{P}\left(y \middle| \underline{x}\right) - \mathbf{E}\left[h\left(Q_\theta, u\right) \middle| \underline{x}\right]\right) = 0\ \underline{x} - a.s.\right\} \\ &= \left\{\theta \in \Theta : u' \mathbf{P}\left(y \middle| \underline{x}\right) \leq \mathbf{E}\left[h\left(Q_\theta, u\right) \middle| \underline{x}\right]\ \forall\ u \in B,\ \underline{x} - a.s.\right\} \end{aligned}$$

It remains to show equivalence of the conditions

$$\begin{array}{rlll} (i) & u'\mathbf{P}\left(y|\,\underline{x}\right) & \leq & \mathbf{E}\left[h\left(Q_\theta,u\right)|\,\underline{x}\right]\ \forall u\in B \\ (ii) & u'\mathbf{P}\left(y|\,\underline{x}\right) & \leq & \mathbf{E}\left[h\left(Q_\theta,u\right)|\,\underline{x}\right]\ \forall u\in D. \end{array}$$

By the positive homogeneity of the support function, condition (i) is equivalent to $u'\mathbf{P}\left(y|\,\underline{x}\right)\leq\mathbf{E}\left[h\left(Q_\theta,u\right)|\,\underline{x}\right]$ $\forall u\in\Re^{\kappa_{\mathcal{Y}}}$. It is obvious that this condition implies condition (ii). To see why condition (ii) implies condition (i), observe that because the set Q_θ and the set $\mathrm{co}\left[Q_\theta\right]$ are simple, one can find a finite measurable partition $\Omega_1,\ldots,\Omega_m$ of Ω and convex sets $K_1,...,K_m\in\Delta^{\kappa_{\mathcal{Y}}-1}$, such that by Theorem 2.1.21 in Molchanov (2005)

$$\mathbb{E}\left(Q_\theta|\,\underline{x}\right)=K_1\mathbf{P}\left(\Omega_1|\,\underline{x}\right)\oplus K_2\mathbf{P}\left(\Omega_2|\,\underline{x}\right)\oplus...\oplus K_m\mathbf{P}\left(\Omega_m|\,\underline{x}\right),$$

with K_i the value that $\mathrm{co}\left[Q_\theta\left(\omega\right)\right]$ takes for $\omega\in\Omega_i$, $i=1,...,m$ (see Definition A.4). By the properties of the support function, see Schneider (1993, Theorem 1.7.5),

$$h\left(\mathbb{E}\left(Q_\theta|\,\underline{x}\right),u\right)=\sum_{i=1}^{m}\mathbf{P}\left(\Omega_i|\,\underline{x}\right)h\left(K_i,u\right).$$

Finally, for each $i=1,...,m$, the vertices of K_i are a subset of the vertices of $\Delta^{\kappa_{\mathcal{Y}}-1}$. Hence the supporting hyperplanes of K_i, $i=1,...,m$, are a subset of the supporting hyperplanes of the simplex $\Delta^{\kappa_{\mathcal{Y}}-1}$, which in turn are obtained through its support function evaluated in directions $u\in D$. Therefore the supporting hyperplanes of $\mathbb{E}\left(Q_\theta|\,\underline{x}\right)$ are a subset of the supporting hyperplanes of $\Delta^{\kappa_{\mathcal{Y}}-1}$. ■

Proof of Theorem 5.1. It follows from our discussion in Section 2 that $\min_{u\in B}\mathbf{E}\left[h\left(Q_\theta,u\right)\right]=0$ if and only if $0\leq h\left(\mathbb{E}\left(Q_\theta\right),u\right)\ \forall\ u\in B$, which in turn holds if and only if $[0\ \ 0]'\in\mathbb{E}\left(Q_\theta\right)$. By the definition of the Aumann expectation, this holds if and only if $\exists\ q\in\mathrm{Sel}\left(Q_\theta\right):\mathbf{E}\left(q\right)=[0\ \ 0]'$. This is equivalent to saying that a candidate θ belongs to Θ_I if and only if $\exists\ (y,x)\in\mathrm{Sel}(Y\times X)$ which yields, together with θ, a prediction error $\varepsilon=y-\theta_1-\theta_2x$ such that $\mathbf{E}\left(\varepsilon\right)=0$ and $\mathbf{E}\left(\varepsilon x\right)=0$. By Theorem 2.1 in Artstein (1983) and Lemma A.2 in Beresteanu, Molchanov, and Molinari (2010a), $(y,x)\in\mathrm{Sel}(Y\times X)$ if and only if $\mathbf{P}\left((y,x)\in K\times L\right)\geq\mathbf{P}\left((Y\times X)\subset K\times L\right)=\mathbf{P}\left(y_L>\inf K,y_U<\sup K,x_L>\inf L,x_U<\sup L\right)$ for all compact intervals $K,L\subset\Re$. Hence, the above condition is equivalent to being able to find a pair of random variables (y,x) with a joint distribution $\mathbf{P}\left(y,x\right)$ that belongs to the (sharp) identification region of $\mathbf{P}\left(y^\star,x^\star\right)$ as defined by Manski (2003, Chapter 3), such that $\theta=\arg\min_{\vartheta\in\Theta}\int(y-\vartheta_1-\vartheta_2x)^2d\mathbf{P}\left(y,x\right)$. It then follows that the set Θ_I is equivalent to the sharp identification region characterized by Manski (2003, Complement 3B, pp. 56-58). The previous step and Lemma 2.1 also verify Assumption 2.5 ■

A.3 Applicability of Andrews and Shi's (2009) GMS Procedure[29]

A.3.1 Finite Games of Complete and Incomplete Information

AS (Section 9) consider conditional moment inequality problems of the form $\mathbf{E}\left(m_d\left(y,\underline{x},\theta,u\right)|\underline{x}\right)\geq0\ \forall u\in B$ $\underline{x}-a.s.$, $d=1,...,D$. They show that the conditional moment inequalities can be transformed into equivalent unconditional moment inequalities, by choosing appropriate weighting functions (instruments) $g\in\mathcal{G}$, with $\mathcal{G}$ a collection of instruments and g that depend on $\underline{x}$. This yields $\mathbf{E}\left(m_d\left(y,\underline{x},\theta,g,u\right)\right)\geq0$, $\forall u\in B$, $\forall g=$

---

[29] We are grateful to Xiaoxia Shi for several discussions that helped us develop this Section.

$[g_1, ..., g_D]' \in \mathcal{G}$, $d = 1, ..., D$, where $m_d(y, \underline{x}, \theta, g, u) = m_d(y, \underline{x}, \theta, u) g(\underline{x})$. In the models that we analyze in Sections 3-4, the conditional moment inequalities are of the " $\leq$ " type, and

$$\begin{aligned} m(y, \underline{x}, \theta, u) &= u' \left[1 \left(y = t^k \right),\ k = 1, ..., \kappa_{\mathcal{Y}} \right] - \mathbf{E}\left[h(Q_\theta, u) | \underline{x} \right], \\ m(y, \underline{x}, \theta, g, u) &= \left(u' \left[1 \left(y = t^k \right),\ k = 1, ..., \kappa_{\mathcal{Y}} \right] - \mathbf{E}\left[h(Q_\theta, u) | \underline{x} \right] \right) g(\underline{x}). \end{aligned}$$

Notice that $\mathbf{E}[h(Q_\theta, u) | \underline{x}]$ is a known (or simulated) function of θ, u and $\underline{x}$, and that for each $u \in B$, we have only one inequality. Notice also that by the positive homogeneity of the support function, our moment inequalities can be written equivalently as $\mathbf{E}(m(y, \underline{x}, \theta, g, u)) \leq 0$, $\forall g \in \mathcal{G}$, $\forall u \in S \equiv \{u \in \Re^{\kappa_{\mathcal{Y}}} : \|u\| = 1\}$. Hence, they are invariant to rescaling of the moment function, which is important for finite sample inference (see, e.g., Andrews and Soares (2010)).

In all that follows, to simplify the exposition, we abstract from the choice of $\mathcal{G}$. Once we establish that our problem fits into the general framework of AS, one can choose instruments g as detailed in Section 3 of AS. To avoid ambiguity, in this Section we denote $F(y|\underline{x}) \equiv \left[\mathbf{P}\left(y = t^k | \underline{x} \right),\ k = 1, ..., \kappa_{\mathcal{Y}} \right]$. We first establish that Θ_I can be equivalently defined using only the first $\kappa_{\mathcal{Y}} - 1$ entries of $\mathcal{Y}$, thereby avoiding the problems for inference associated with linear dependence among the entries of $F(y|\underline{x})$ and also lowering the dimension over which the maximization is performed. Let $\widetilde{F}(y|\underline{x})$ denote the first $\kappa_{\mathcal{Y}} - 1$ rows of $F(y|\underline{x})$, $B^{\kappa_{\mathcal{Y}}-1} = \{u \in \Re^{\kappa_{\mathcal{Y}}-1} : \|u\| \leq 1\}$, $S^{\kappa_{\mathcal{Y}}-1} = \{u \in \Re^{\kappa_{\mathcal{Y}}-1} : \|u\| = 1\}$, and

$$\tilde{Q}_\theta = \{\tilde{q} = [[q(\sigma)]_k,\ k = 1, ..., \kappa_{\mathcal{Y}} - 1],\ q \in \mathrm{Sel}(Q_\theta)\}.$$

Theorem A.1 *Let Assumptions 3.1 (or 4.1) and 3.2 hold, and no other information be available. Then*

$$\begin{aligned} \text{(A.1)} \qquad \tilde{\Theta}_I &\equiv \left\{ \theta \in \Theta : \max_{u \in B^{\kappa_{\mathcal{Y}}-1}} \left(u' \widetilde{F}(y|\underline{x}) - \mathbf{E}\left[h\left(\tilde{Q}_\theta, u \right) | \underline{x} \right] \right) = 0 \ \ \underline{x} - a.s. \right\} \\ \text{(A.2)} \qquad &= \left\{ \theta \in \Theta : \left[\max_{u \in S^{\kappa_{\mathcal{Y}}-1}} \left(u' \widetilde{F}(y|\underline{x}) - \mathbf{E}\left[h\left(\tilde{Q}_\theta, u \right) | \underline{x} \right] \right) \right]_+ = 0 \ \ \underline{x} - a.s. \right\} \\ &= \Theta_I \end{aligned}$$

Proof. The equality between equations (A.1) and (A.2) follows by standard arguments, see, e.g., Beresteanu and Molinari (2008, Lemma A.1). To establish that $\tilde{\Theta}_I = \Theta_I$, observe that $\theta \in \tilde{\Theta}_I$ if and only if $\widetilde{F}(y|\underline{x}) \in \mathbb{E}\left(\tilde{Q}_\theta | \underline{x} \right)$. Pick $\theta \in \Theta_I$. Then there exists a $q \in \mathrm{Sel}(Q_\theta) : F(y|\underline{x}) = \mathbf{E}(q|\underline{x})$. Notice that this implies $\widetilde{F}(y|\underline{x}) = \mathbf{E}(\tilde{q}|\underline{x})$ for $\tilde{q} \in \mathrm{Sel}\left(\tilde{Q}_\theta \right)$, hence, $\theta \in \tilde{\Theta}_I$. Conversely, pick $\theta \in \tilde{\Theta}_I$. Then there exists a $\tilde{q} \in \mathrm{Sel}\left(\tilde{Q}_\theta \right) : \widetilde{F}(y|\underline{x}) = \mathbf{E}(\tilde{q}|\underline{x})$, which in turn implies that $q = \left[\tilde{q}; 1 - \sum_{k=1}^{\kappa_{\mathcal{Y}}-1} \tilde{q} \right] \in \mathrm{Sel}(Q_\theta)$ and $F(y|\underline{x}) = \mathbf{E}(q|\underline{x})$. Hence, $\theta \in \Theta_I$. ■

AS propose a confidence set with nominal value $1 - \alpha$ for the true parameter vector, as follows:

$$CS_n = \{\theta \in \Theta : T_n(\theta) \leq c_{n,1-\alpha}(\theta)\},$$

where $T_n(\theta)$ is a test statistic and $c_{n,1-\alpha}(\theta)$ is a corresponding critical value for a test with nominal significance level α. AS establish that, under certain assumptions, this confidence set has correct uniform asymptotic size. In order to apply the construction in AS, we maintain the following:

Assumption A.1 *The researcher observes an i.i.d. sequence of equilibrium outcomes and observable payoff shifters* $\{y_i, \underline{x}_i\}_{i=1}^n$. *Define* $\tilde{\Sigma}_{\underline{x}} = \operatorname{diag}\left(\widetilde{F}\left(y|\underline{x}\right)\right) - \widetilde{F}\left(y|\underline{x}\right)\widetilde{F}\left(y|\underline{x}\right)'$, *and let* $\tilde{\Sigma}_{\underline{x}}$ *be non-singular with* $a < \left\|\tilde{\Sigma}_{\underline{x}}\right\| < b$ *for some constants* $0 < a < b < \infty$, $\underline{x} - a.s.$, *where* $\left\|\tilde{\Sigma}_{\underline{x}}\right\|$ *is a matrix norm for* $\tilde{\Sigma}_{\underline{x}}$ *compatible with the Euclidean norm.*

AS propose various criterion functions T_n, some of the Cramér-von Mises type, some of the Kolmogorov-Smirnov type. Here, we work with a mix of Cramér-von Mises and Kolmogorov-Smirnov statistic, using a modification of the function S_1 on page 10 of AS. Specifically, we use

(A.3)

$$T_n\left(\theta\right) = \int\left(\max_{u\in B^{\kappa_{\mathcal{Y}}-1}}\sqrt{n}\bar{m}_n\left(\theta,g,u\right)\right)^2 d\Gamma = \int\left(\max_{u\in S^{\kappa_{\mathcal{Y}}-1}}\sqrt{n}\bar{m}_n\left(\theta,g,u\right)\right)_+^2 d\Gamma = \int\max_{u\in S^{\kappa_{\mathcal{Y}}-1}}\left(\sqrt{n}\bar{m}_n\left(\theta,g,u\right)\right)_+^2 d\Gamma$$

where Γ denotes a probability measure on $\mathcal{G}$ whose support is $\mathcal{G}$ as detailed in Section 3 of AS, the second equality follows from the proof of Theorem A.1, and

$$\begin{aligned}\bar{m}_n\left(\theta,g,u\right) &= \frac{1}{n}\textstyle\sum_{i=1}^n\left(u'w\left(y_i\right) - f\left(\underline{x}_i,\theta,u\right)\right)g\left(\underline{x}_i\right), \\ f\left(\underline{x}_i,\theta,u\right) &= \mathbf{E}\left[h\left(\tilde{Q}_\theta,u\right)|\underline{x}_i\right] \\ w\left(y_i\right) &= \left[1\left(y_i = t^k\right),\ k = 1,...,\kappa_{\mathcal{Y}} - 1\right],\end{aligned}$$

so that $\bar{m}_n\left(\theta,g,u\right)$ is the sample analog of a version of $\mathbf{E}\left(m\left(y,\underline{x},\theta,g,u\right)\right)$ which is based on the first $\kappa_{\mathcal{Y}} - 1$ entries of $\mathcal{Y}$ and on $\tilde{Q}_\theta$. Note that by the same argument which follows, our problem specified as in equation (3.6) corresponds to the Cramér-von Mises test statistic of AS, with modified function S_1.

We conclude by showing that our modified function S_1 satisfies Assumptions S1-S4 of AS, and that Assumption M2 of AS is also satisfied. This establishes that their generalized moment selection procedure with infinitely many conditional moment inequalities is applicable. We note that one can take the confidence set CS_n applied with confidence level 1/2 to obtain half-median-unbiased estimated sets, see AS and Chernozhukov, Lee, and Rosen (2009). Finally, one can also take equation (A.1), replace there $\widetilde{F}(y|\underline{x})$ with its sample analog, and construct an Hausdorff-consistent estimator of Θ_I using the methodology proposed by Chernozhukov, Hong, and Tamer (2007, equation 3.2 and Theorem 3.1).[30]

Theorem A.2 *Let Assumption A.1 hold. Then Assumptions S1-S4 and M2 of AS are satisfied.*

---

[30] Assume that the payoff functions are continuous in θ. Then the Nash equilibrium correspondence has a closed graph, see Fudenberg and Tirole (1991, Section 1.3.2). This implies that Q_θ has a closed graph, and therefore the same is true for $\mathbb{E}(Q_\theta|\underline{x})$, see Aumann (1965, Corollary 5.2). In turn, this yields $\limsup_{\theta_n\to\theta}\mathbb{E}\left(Q_{\theta_n}|\underline{x}\right) \subseteq \mathbb{E}\left(Q_\theta|\underline{x}\right)$. Observe that

$$\max_{u\in B^{\kappa_{\mathcal{Y}}-1}}\left(u'\widetilde{F}\left(y|\underline{x}\right) - \mathbf{E}\left[h\left(\tilde{Q}_\theta,u\right)|\underline{x}\right]\right) = d_H\left(\widetilde{F}\left(y|\underline{x}\right),\mathbb{E}\left(\tilde{Q}_\theta|\underline{x}\right)\right).$$

The criterion function $s\left(\theta\right) \equiv \int d_H\left(\widetilde{F}\left(y|\underline{x}\right),\mathbb{E}\left(\tilde{Q}_\theta|\underline{x}\right)\right)dF_{\underline{x}}$, with $F_{\underline{x}}$ the probability distribution of $\underline{x}$ (or a probability measure which dominates it), is therefore lower semicontinuous in θ, because

$$\begin{aligned}\liminf_{\theta_n\to\theta} s(\theta_n) &\geq \int\liminf_{\theta_n\to\theta} d_H\left(\widetilde{F}\left(y|\underline{x}\right),\mathbb{E}\left(\tilde{Q}_{\theta_n}|\underline{x}\right)\right)dF_{\underline{x}} \geq \int d_H\left(\widetilde{F}\left(y|\underline{x}\right),\limsup\mathbb{E}\left(\tilde{Q}_{\theta_n}|\underline{x}\right)\right)dF_{\underline{x}} \\ &\geq \int d_H\left(\widetilde{F}\left(y|\underline{x}\right),\mathbb{E}\left(\tilde{Q}_\theta|\underline{x}\right)\right)dF_{\underline{x}} = s(\theta).\end{aligned}$$

Conditions (c-e) in Assumption C1 of Chernozhukov, Hong, and Tamer (2007) are verified by standard arguments.

Proof. Assumption S1-a follows because the moment inequalities are defined for $u \in S^{\kappa_{\mathcal{Y}}-1}$, hence any rescaling of the moment function is absorbed by a corresponding rescaling in u. The rest of Assumption S1 and Assumptions S2-S4 are verified by AS. To verify Assumption M2, observe that

$$\widetilde{m}\left(y, \underline{x}, \theta, u\right) \equiv u' w\left(y\right) - f\left(\underline{x}, \theta, u\right)$$

is given by the sum of a linear function of u and a Lipschitz function of u, with Lipschitz constant equal to 1. It is immediate that the processes $\{u' w\left(y_{in}\right), u \in S^{\kappa_{\mathcal{Y}}-1}, i \leq n, n \geq 1\}$ satisfy Assumption M2. We now show that the same holds for the processes $\{f\left(\underline{x}_{in}, \theta_n, u\right), u \in S^{\kappa_{\mathcal{Y}}-1}, i \leq n, n \geq 1\}$. Assumption M2-(a) holds because for all $u \in S^{\kappa_{\mathcal{Y}}-1}$,

$$\left|\frac{f\left(\underline{x}, \theta, u\right)}{Var\left(\widetilde{m}\left(y, \underline{x}, \theta, u\right)\right)}\right| \leq \left|\frac{f\left(\underline{x}, \theta, u\right)}{\mathbf{E}\left(u' \tilde{\Sigma}_{\underline{x}} u\right)}\right| \leq c \left|\mathbf{E}\left[h\left(\tilde{Q}_\theta, u\right) |\underline{x}\right]\right| \leq c \mathbf{E}\left(\left\|\tilde{Q}_\theta\right\|_H |\underline{x}\right) \leq c \quad \underline{x} - a.s.,$$

where the first inequality follows from the variance decomposition formula, c is a constant that depends on a and b from Assumption A.1, and the last inequality follows recalling that $\tilde{Q}_\theta$ takes its realizations in the unit simplex which is a subset of the unit ball. Assumption M2-(b) follows immediately because the envelope function is a constant. Assumption M2-(c) is verified observing that $f\left(\underline{x}, \theta, u\right)$ is Lipschitz in u, with Lipschitz constant equal to 1. By Lemma 2.13 in Pakes and Pollard (1989), the class of functions $\{f\left(\cdot, u\right),$ $u \in S^{\kappa_{\mathcal{Y}}-1}\}$ is Euclidean with envelope equal to a constant, and therefore manageable. Assumption M2 for the processes $\{\left(u' w\left(y_{in}\right) - f\left(\underline{x}_{in}, \theta_n, u\right)\right), u \in S^{\kappa_{\mathcal{Y}}-1}, i \leq n, n \geq 1\}$ then follows by Lemma E1 of AS. ■

A.3.2 BLP with Interval Outcome and Covariate Data

We maintain the following:

Assumption A.2 *The researcher observes an i.i.d. sequence of tuples* $\{y_{iL}, y_{iU}, x_{iL}, x_{iU}\}_{i=1}^n$. $\mathbf{E}\left(\left|y_i\right|^2\right)$, $\mathbf{E}\left(\left|x_j\right|^2\right)$, $\mathbf{E}\left(\left|y_i x_j\right|^2\right)$, *and* $\mathbf{E}\left(x_j^4\right)$ *are all finite, for each* $i, j = L, U$.

Let $Q_{\theta i}$ be the mapping defined as in equation (5.1) using $\left(y_{iL}, y_{iU}, x_{iL}, x_{iU}\right)$. Beresteanu and Molinari (2008, Lemmas A.4, A.5 and proof of Theorem 4.2) establish that $\{Q_{\theta i}\}_{i=1}^n$ is a sequence of i.i.d. random closed sets, such that $\mathbf{E}\left(\left\|Q_{\theta i}\right\|_H^2\right) < \infty$. Define $T_n\left(\theta\right)$ similarly to the previous Section:

$$\begin{aligned} T_n\left(\theta\right) &= \left(\max_{u \in B}\left(-\sqrt{n}\bar{m}_n\left(\theta, u\right)\right)\right)^2 = \left(\max_{u \in S} -\sqrt{n}\bar{m}_n\left(\theta, u\right)\right)_+^2 = \max_{u \in S}\left(-\sqrt{n}\bar{m}_n\left(\theta, u\right)\right)_+^2, \\ \bar{m}_n\left(\theta, u\right) &= \frac{1}{n}\sum_{i=1}^n h\left(Q_{\theta i}, u\right), \end{aligned}$$

where, again, the fact that $u \in S$ guarantees that the above test statistic is invariant to rescaling of the moment function. This preserves concavity of the objective function. We then have the following result:

Theorem A.3 *Let Assumptions 5.1 and A.2 hold. Then Assumption EP on page 37 of AS is satisfied.*

Proof. Let $m\left(y_{iL}, y_{iU}, x_{iL}, x_{iU}, \theta, u\right) = h\left(Q_{\theta i}, u\right)$. Following AS notation, define

$$\begin{aligned}
\sqrt{n}\bar{m}_n\left(\theta, u\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h\left(Q_{\theta i}, u\right), \\
\gamma_{1,n}\left(\theta, u\right) &= \sqrt{n}\mathbf{E}\left[h\left(Q_{\theta i}, u\right)\right], \\
\gamma_2\left(\theta, u, u^*\right) &= \mathbf{E}\left[h\left(Q_{\theta i}, u\right) h\left(Q_{\theta i}, u^*\right)\right] - \mathbf{E}\left[h\left(Q_{\theta i}, u\right)\right]\mathbf{E}\left[h\left(Q_{\theta i}, u^*\right)\right], \\
\nu_n\left(\theta, u\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[h\left(Q_{\theta i}, u\right) - \mathbf{E}\left(h\left(Q_{\theta i}, u\right)\right)\right].
\end{aligned}$$

Given the above definitions, we have

$$\sqrt{n}\bar{m}_n\left(\theta, u\right) = \nu_n\left(\theta, u\right) + \gamma_{1,n}\left(\theta, u\right).$$

By the Central Limit Theorem for i.i.d. sequences of random sets (Molchanov (2005, Theorem 2.2.1))

$$\nu_n\left(\theta, \cdot\right) \Longrightarrow \nu_{\gamma_2(\theta)}\left(\cdot\right),$$

a Gaussian process with mean zero, covariance kernel $\gamma_2\left(\theta, u, u^*\right)$, and continuous sample paths. It follows from the Strong Law of Large Numbers in Banach spaces of Mourier (1955) that the sample analog estimator $\hat{\gamma}_{2,n}\left(\theta, u, u^*\right)$ which replaces population moments with sample averages, satisfies $\hat{\gamma}_{2,n}\left(\theta, \cdot, \cdot\right) \to^{a.s.} \gamma_2\left(\theta, \cdot, \cdot\right)$, uniformly in u, u^*. ■

A.4 Computational Aspects of the Problem

In this Section, we focus on games of complete information. The case of games of incomplete information can be treated analogously, and we refer to Grieco (2009) for a thorough discussion of how to compute the set of Bayesian Nash equilibria. The case of BLP with interval data is straightforward.

When computing Θ_I (or Θ_O^{ABJ} and Θ_O^{CT}), one faces three challenging tasks. We describe them here, and note how each task is affected by the number of players, the number of strategy profiles, and the presence of covariates in the payoff functions. For comparison purposes, we also discuss the differences in computational costs associated with our methodology, versus those associated with ABJ's and CT's methodology.

The first step in the procedure requires one to compute the set of all MSNE for given realizations of the payoff shifters, $S_\theta\left(\underline{x}, \varepsilon\right)$. This is a computationally challenging problem, though a well studied one which can be performed using the Gambit software described by McKelvey and McLennan (1996).[31] The complexity of this task grows quickly with the number of players and the number of actions that each player can choose from. Notice, however, that this step has to be performed regardless of which features of normal form games are identified: whether sufficient conditions are imposed for point identification of the parameter vector of interest, as in Bajari, Hong, and Ryan (2009), or this vector is restricted to lie in an outer region, or its sharp identification region is characterized through the methodology proposed in this paper. For example, Bajari, Hong, and Ryan (2009) work with an empirical application which has a very large number of players, but

---

[31] The Gambit software can be freely downloaded at http://gambit.sourceforge.net/. Bajari, Hong, and Ryan (2009) recommend the use of this software to compute the set of mixed strategy Nash equilibria in finite normal form games.

group the smaller ones together in order to reduce the number of players to 4. Similarly, application of our method to games with multiple *mixed* strategies Nash equilibria requires a limited number of players.[32]

The second task involves verifying whether a candidate $\theta \in \Theta$ is in the region of interest. The difficulty of this task varies depending on whether one wants to check that $\theta \in \Theta_I$, or that $\theta \in \Theta_O^{ABJ}$ or $\theta \in \Theta_O^{CT}$. As established in Proposition 3.3, in all cases one needs to work with $\mathbf{E}\left[h\left(Q_\theta, u\right)\middle| \underline{x}\right]$, so we first describe, for a given $u \in \Re^{\kappa_{\mathcal{Y}}}$, how to obtain this quantity by simulation (see, e.g., McFadden (1989) and Pakes and Pollard (1989)). Recall that for given $\theta \in \Theta$ and realization of $\underline{x}$

$$\mathbf{E}\left[h\left(Q_\theta, u\right)\middle| \underline{x}\right] = \mathbf{E}\left[\max_{\sigma \in S_\theta(\underline{x},\varepsilon)} u' q\left(\sigma\right) \middle| \underline{x}\right] = \int \max_{\sigma \in S_\theta(\underline{x},\varepsilon)} \sum_{k=1}^{\kappa_{\mathcal{Y}}} u_k \prod_{j=1}^{J} \sigma_j\left(t_j^k\right) dF_\theta\left(\varepsilon\middle| \underline{x}\right),$$

where $u' = \left[u_1 u_2 \dots u_{\kappa_{\mathcal{Y}}}\right]$ and $\mathcal{Y} = \left\{t^1, ..., t^{\kappa_{\mathcal{Y}}}\right\}$ is the set of possible outcomes of the game. One can simulate this multidimensional integral using the following procedure.[33] Let $\mathcal{X}$ denote the support of $\underline{x}$. For any $\underline{x} \in \mathcal{X}$, draw realizations of ε, denoted ε^r, $r = 1, \dots, R$, according to the distribution $F\left(\varepsilon\middle| \underline{x}\right)$ with identity covariance matrix. These draws stay fixed throughout the remaining steps. Transform the realizations ε^r, $r = 1, \dots, R$, into draws with covariance matrix specified by θ. For each ε^r, compute the payoffs $\pi_j\left(\cdot, x_j, \varepsilon_j^r, \theta\right)$, $j = 1, \dots, J$, and obtain the set $S_\theta\left(\underline{x}, \varepsilon^r\right)$. Then compute the set $Q_\theta\left(\underline{x}, \varepsilon^r\right)$ as the set of multinomial distributions over outcome profiles implied by each element of $S_\theta\left(\underline{x}, \varepsilon^r\right)$. Pick a $u \in \Re^{\kappa_{\mathcal{Y}}}$, compute the support function $h\left(Q_\theta\left(\underline{x}, \varepsilon^r\right), u\right)$, and average it over a large number of draws of ε^r. Call the resulting average $\hat{\mathbf{E}}_R\left[h\left(Q_\theta, u\right)\middle| \underline{x}\right]$. Note that because each summand is a function of ε^r and these are i.i.d. draws from the distribution $F_\theta\left(\varepsilon\middle| \underline{x}\right)$, $\mathbf{E}_{F_\theta(\varepsilon|\underline{x})}\left(\hat{\mathbf{E}}_R\left[h\left(Q_\theta, u\right)\middle| \underline{x}\right]\right) = \mathbf{E}\left[h\left(Q_\theta, u\right)\middle| \underline{x}\right]$.

Having obtained $\hat{\mathbf{E}}_R\left[h\left(Q_\theta, u\right)\middle| \underline{x}\right]$, in order to verify whether $\theta \in \Theta_O^{ABJ}$ and $\theta \in \Theta_O^{CT}$, it suffices to verify conditional moment inequalities involving, respectively, $\kappa_{\mathcal{Y}}$ and $2\kappa_{\mathcal{Y}}$ evaluations of $\hat{\mathbf{E}}_R\left[h\left(Q_\theta, u\right)\middle| \underline{x}\right]$, corresponding to the choices of u detailed in Proposition 3.3. As illustrated in our examples, however, using only these inequalities may lead to outer regions which are much larger than Θ_I. Verifying whether $\theta \in \Theta_I$ using the method described in this paper involves solving $\max_{u \in B^{\kappa_{\mathcal{Y}}-1}}\left(u'\widetilde{F}\left(y|\underline{x}\right) - \hat{\mathbf{E}}_R\left[h\left(\tilde{Q}_\theta, u\right)|\underline{x}\right]\right)$ and checking if the resulting value function is equal to zero, for each value of $\underline{x}$. We emphasize that the dimensionality of u does *not* depend in any way on the number of equilibria of the game (just on the number of players and strategies), or on the number R of draws of ε taken to simulate $\mathbf{E}\left[h\left(\tilde{Q}_\theta, u\right)\middle| \underline{x}\right]$. As stated before, for given $\underline{x} \in \mathcal{X}$, the criterion function to be maximized is concave, and the maximization occurs over a convex subset of $\Re^{\kappa_{\mathcal{Y}}-1}$. In a two player entry game with payoffs linear in $\underline{x}$, we have experienced that efficient algorithms in convex programming, such as the CVX software for MatLab (Grant and Boyd (2010)), can solve this maximization problem with a handful of iterations, in the order of 10-25 depending on the candidate θ. We have also experienced that a simple Nelder-Mead algorithm programmed in Fortran 90 works very well, yielding the usual speed advantages of Fortran over MatLab. For each parameter candidate, the above maximization problem needs to be solved for each possible value of $\underline{x} \in \mathcal{X}$ when $\underline{x}$ is discrete and for each $g \in \mathcal{G}$ when applying AS procedure,[34] and checking whether all required conditions are satisfied.

---

[32] On the other hand, our method is applicable to models with a larger set of players, when players are restricted to playing pure strategies, or the game is one of incomplete information.

[33] The procedure described here is very similar to the one proposed by Ciliberto and Tamer (2009). When the assumptions maintained by Bajari, Hong, and Ryan (2009, Section 3) are satisfied, their algorithm can be used to significantly reduce the computational burden associated with simulating the integral.

[34] AS show that in practice, the integral over $\mathcal{G}$ can be replaced by a finite sum, see their Section 3.5.

Therefore, it is reasonable to say that the computational burden of this stage is linear in the number of values that $\underline{x}$ (respectively, g) takes.

Finally, the region of interest needs to be computed. This means that the researcher should search over the parameter space Θ and collect all the points in Θ_I or Θ_O^{CT} or Θ_O^{ABJ}. This is of course a theoretical set and in practice the researcher seeks to collect enough points that belong to the region of interest, such that it can be covered reasonably well. While easy to program, a grid search over Θ is highly inefficient, especially when Θ belongs to a high-dimensional space. CT propose to search over Θ using a method based on Simulated Annealing. In this paper we use an alternative algorithm called Differential Evolution. We give here a short description of this method focusing mainly on its complexity. We refer to Price, Storn, and Lampinen (2004) for further details. Differential Evolution (DE) is a type of Genetic Algorithm that is often used to solve optimization problems. The algorithm starts from a population of N_p points picked randomly from the set Θ. It then updates this list of points at each stage, creating a new generation of N_p points to replace the previous one. A candidate to replace a current member of the population (child) is created by combining information from members of the current population (parents). This new candidate is accepted to the population as a replacement to a current member if it satisfies a certain criterion. In our application, the criterion for being admitted into the new generation is to be a member of Θ_I (or Θ_O^{CT} or Θ_O^{ABJ}, when computing each of these regions). The process of finding a replacement for each of the current N_p points is repeated N times, yielding $N \cdot N_p$ maximizations of the criterion function (respectively, evaluation of the conditional inequalities for CT and ABJ). During this process we record the points which were found to belong to the regions of interest. In our simulations, we experienced that this method explores Θ in a very efficient way. Price, Storn, and Lampinen (2004) recommend for N_p to grow linearly with the dimensionality of Θ. The number of iterations (generations) N depends on how well one wants to cover the region of interest. For example, in a two player entry game with $\Theta \subset \Re^4$, we have found that setting $N_p = 40$ and $N = 1000$ gave satisfactory results, and when N was increased to 5000 the regions of interest seemed to be very well covered, while the projections on each component of θ remained very similar to what we obtained with $N = 1000$. Creating candidates to replace members of the population involve trivial algebraic operations whose number grows linearly with the dimensionality of Θ. These operations involve picking two tuning parameters, but satisfactory rules of thumb exist in the literature, see Price, Storn, and Lampinen (2004).

References

Andrews, D. W. K., S. T. Berry, and P. Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," mimeo.

Andrews, D. W. K., and X. Shi (2009): "Inference Based on Conditional Moment Inequalities," mimeo.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

Aradillas-Lopez, A., and E. Tamer (2008): "The Identification Power of Equilibrium in Simple Games," *Journal of Business and Economic Statistics*, 26(3), 261–310.

Aradillas-López, A. (2010): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information," *Journal of Econometrics*, 157, 409–431.

Artstein, Z. (1983): "Distributions of Random Sets and Random Selections," *Israel Journal of Mathematics*, 46(4), 313–324.

Aumann, R. J. (1965): "Integrals of Set Valued Functions," *Journal of Mathematical Analysis and Applications*, 12, 1–12.

Bajari, P., H. Hong, and S. Ryan (2009): "Identification and Estimation of a Discrete Game of Complete Information," *Econometrica*, forthcoming.

Beresteanu, A., I. Molchanov, and F. Molinari (2008): "Sharp Identification Regions in Games," CeMMAP Working Paper CWP15/08.

——— (2009): "Sharp Identification Regions in Models with Convex Predictions: Games, Individual Choice, and Incomplete Data," CeMMAP Working Paper CWP27/09.

——— (2010a): "Partial Identification Using Random Sets Theory," mimeo.

——— (2010b): "Supplement to 'Sharp Identification Regions in Models with Convex Moment Predictions'," mimeo.

Beresteanu, A., and F. Molinari (2006): "Asymptotic Properties for a Class of Partially Identified Models," *CeMMAP working papers*, CWP10/06.

——— (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814.

Berry, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60(4), 889–917.

Berry, S. T., and E. Tamer (2007): "Identification in Models of Oligopoly Entry," in *Advances in Economics and Econometrics: Theory and Application*, vol. II, chap. 2, pp. 46–85. Cambridge University Press, Ninth World Congress.

Bjorn, P. A., and Q. H. Vuong (1985): "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation," CalTech DHSS Working Paper Number 557.

Bontemps, C., T. Magnac, and E. Maurin (2008): "Set Identified Linear Models," mimeo.

Boyd, S., and L. Vandenberghe (2004): *Convex Optimization*. Cambridge University Press, New York.

Bresnahan, T. F., and P. C. Reiss (1988): "Do Entry Conditions Vary Across Markets?," *Brookings Papers on Economic Activity*, pp. 833–871.

——— (1990): "Entry in Monopoly Markets," *Review of Economic Studies*, 57, 531–553.

——— (1991): "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 99(5), 977–1009.

Brock, W., and S. Durlauf (2007): "Identification of binary choice models with social interactions," *Journal of Econometrics*, 140, 52–75.

Brock, W. A., and S. N. Durlauf (2001): *Interactions-Based Models*chap. 54, pp. 3297–3380, Handbook of Econometrics. Elsevier Science.

Bugni, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78, 735–753.

Canay, I. A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156, 408–425.

Chernozhukov, V., H. Hong, and E. Tamer (2004): "Parameter Set Inference in a Class of Econometric Models," Discussion paper, MIT, Duke University and Northwestern University.

——— (2007): "Estimation and Confidence Regions for Parameter Sets In Econometric Models," *Econometrica*, 75, 1243–1284.

Chernozhukov, V., S. Lee, and A. Rosen (2009): "Intersection Bounds: Estimation and Inference," CeMMAP Working Paper CWP19/09.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828.

Dyer, M., A. Frieze, and R. Kannan (1991): "A random polynomial-time algorithm for approximating the volume of convex bodies," *Journal of the Association for Computing Machinery*, 38, 1–17.

Dynkin, E., and I. V. Evstigneev (1976): "Regular conditional expectations of correspondences," *Theory of Probability and its Applications*, 21, 325–338.

Fudenberg, D., and J. Tirole (1991): *Game Theory*. MIT Press, Cambridge Massachusetts.

Galichon, A., and M. Henry (2006): “Inference in Incomplete Models,” mimeo.

——— (2009a): “Set Identification in Models with Multiple Equilibria,” mimeo.

——— (2009b): “A Test of Non-Identifying Restrictions and Confidence Regions for Partially Identified Parameters,” *Journal of Econometrics*, 152, 186–196.

Grant, M., and S. Boyd (2008): “Graph Implementations for Nonsmooth Convex Programs,” in *Recent Advances in Learning and Control*, ed. by V. Blondel, S. Boyd, and H. Kimura, pp. 95–110. Springer-Verlag Limited, `http://stanford.edu/~boyd/graph_dcp.html`.

——— (2010): “CVX: Matlab Software for Disciplined Convex Programming, Version 1.21,” `http://cvxr.com/cvx`.

Grieco, P. L. E. (2009): “Discrete Choice Games with Flexible Information Structure: An Application to Local Grocery Markets,” mimeo.

Horowitz, J. L., C. F. Manski, M. Ponomareva, and J. Stoye (2003): “Computation of Bounds on Population Parameters When the Data Are Incomplete,” *Reliable Computing*, 9, 419–440.

Imbens, G. W., and C. F. Manski (2004): “Confidence Intervals for Partially Identified Parameters,” *Econometrica*, 72, 1845–1857.

Kim, K.-i. (2009): “Set Estimation and Inference with Models Characterized by Conditional Moment Inequalities,” mimeo.

Lovász, L., and S. Vempala (2006): “Simulated annealing in convex bodies and an O*(n4) volume algorithm,” *Journal of Computer and System Sciences*, 72, 392–417.

Manski, C. F. (1989): “Anatomy of the Selection Problem,” *Journal of Human Resources*, 24, 343–360.

——— (2003): *Partial Identification of Probability Distributions*. Springer Verlag, New York.

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

Mazzeo, M. (2002): “Product Choice and Oligopoly Market Structure,” *RAND Journal of Economics*, 33(2), 221–242.

McFadden, D. (1989): “A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration,” *Econometrica*, 57, 995–1026.

McKelvey, R. D., and A. McLennan (1996): “Computation of Equilibria in Finite Games,” in *Handbook of Computational Economics*, vol. 1, chap. 2, pp. 87–142. Elsevier Science.

Molchanov, I. (2005): *Theory of Random Sets*. Springer Verlag, London.

Molinari, F. (2008): "Partial Identification of Probability Distributions with Misclassified Data," *Journal of Econometrics*, 144(1), 81–117.

Mourier, E. (1955): "L-Random Elements and L^*-Random Elements in Banach Spaces," *Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability*, 2, 231–242.

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2006): "Moment Inequalities and Their Application," mimeo.

Price, K. V., R. M. Storn, and J. A. Lampinen (2004): *Differential Evolution: a Practical Approach to Global Optimization*. The MIT Press, Cambridge, Massachusetts.

Rao, M. M., and R. J. Swift (2006): *Probability Theory with Applications*, Mathematics and Its Applications, Band 582. Springer-Verlag, New York.

Rockafellar, R. (1970): *Convex Analysis*. Princeton University Press.

Romano, J. P., and A. M. Shaikh (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169–211.

Rosen, A. (2008): "Confidence Sets for Partially Identified Parameters That Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146(1), 107–117.

Schneider, R. (1993): *Convex Bodies: The Brunn-Minkowski Theory*. Cambridge Univ. Press.

Seim, K. (2006): "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *RAND Journal of Economics*, 37, 619–640.

Stoye, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1299–1315.

Sweeting, A. (2009): "The Strategic Timing of Radio Commercials: An Empirical Analysis Using Multiple Equilibria," *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.

——— (2009): "Partial Identification in Econometrics," *Annual Reviews of Economics*, forthcoming.

Wilson, R. (1971): "Computing Equilibria of N-Person Games," *SIAM Journal on Applied Mathematics*, 21, 80–87.

B Tables and Figures

Table 1: Projections of Θ_O^{ABJ}, Θ_O^{CT} and Θ_I, reduction in bounds width compared to CT, and reduction in area of Θ_I compared to Θ_O^{CT} . Results reported are "top 15 percent," "median," and "bottom 15 percent" reduction in the sum of length of sharp bounds on θ_1 and θ_2 compared to the sum of length of outer bounds of CT. Two player entry game with mixed strategy Nash equilibrium as solution concept.

True Values[1]	Projections of:			% Width Reduction[2]	% Region Reduction[3]
	Θ_O^{ABJ}	Θ_O^{CT}	Θ_I		
$\theta_1^\star = -1.0$	$[-3.22, -0.22]$	$[-3.22, -0.28]$	$[-2.21, -0.30]$	35.03%	63.81%
$\theta_2^\star = -1.3$	$[-3.22, -1.05]$	$[-3.22, -1.15]$	$[-2.32, -1.16]$	43.96%	
$\lambda^\star = \begin{bmatrix} 0 & \frac{3}{4} & \frac{1}{4} \end{bmatrix}$					
$\theta_1^\star = -0.8$	$[-1.82, -0.53]$	$[-1.82, -0.57]$	$[-1.51, -0.58]$	25.60%	56.87%
$\theta_2^\star = -1.1$	$[-1.82, -0.59]$	$[-1.82, -0.64]$	$[-1.55, -0.64]$	22.88%	
$\lambda^\star = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$					
$\theta_1^\star = -1.2$	$[-2.19, -0.75]$	$[-2.19, -0.90]$	$[-2.11, -1.05]$	17.83%	26.02%
$\theta_2^\star = -1.5$	$[-2.19, -0.75]$	$[-2.19, -0.79]$	$[-2.13, -0.90]$	12.14%	
$\lambda^\star = \begin{bmatrix} \frac{1}{4} & 0 & \frac{3}{4} \end{bmatrix}$					

[1] $\lambda^\star = [\Pr\big((0,1) \text{ is chosen} \mid \varepsilon \in \mathcal{E}^M_{\theta^\star}\big) \ \Pr\big((1,0) \text{ is chosen} \mid \varepsilon \in \mathcal{E}^M_{\theta^\star}\big) \ \Pr\big((\frac{\varepsilon_2}{-\theta_2}, \frac{\varepsilon_1}{-\theta_1}) \text{ is chosen} \mid \varepsilon \in \mathcal{E}^M_{\theta^\star}\big)]$, $\mathcal{E}^M_{\theta^\star} = [0, -\theta_1^\star] \times [0, -\theta_2^\star]$

[2] Calculated as $\frac{\text{Proj}(\Theta_O^{CT}|j)\text{-Proj}(\Theta_I|j)}{\text{Proj}(\Theta_O^{CT}|j)}$, where $\text{Proj}(\cdot|j)$ is the projection of the set in parenthesis on dimension j.

[3] Calculated as $\frac{\text{Area}(\Theta_O^{CT})\text{-Area}(\Theta_I)}{\text{Area}(\Theta_O^{CT})}$, where $\text{Area}(\cdot)$ is the area of the set in parenthesis.

Table 2: Projections of Θ_O^{ABJ}, Θ_O^{CT} and Θ_I and reduction in volume of Θ_I compared to Θ_O^{CT}. Two player entry game with mixed strategy Nash equilibrium as solution concept.

	Projections of:								
	DGP 1[a]			DGP 2[b]			DGP 3[c]		
	Θ_O^{ABJ}	Θ_O^{CT}	Θ_I	Θ_O^{ABJ}	Θ_O^{CT}	Θ_I	Θ_O^{ABJ}	Θ_O^{CT}	Θ_I
Player 1:									
θ_1	$[-3.81, -0.50]$	$[-3.78, -0.60]$	$[-3.18, -0.63]$	$[-1.51, -0.45]$	$[-1.47, -0.61]$	$[-1.41, -0.62]$	$[-2.37, -0.46]$	$[-1.99, -0.69]$	$[-1.77, -0.80]$
β_{01}	$[-0.09, 0.43]$	$[-0.02, 0.41]$	$[-0.01, 0.35]$	$[-0.16, 0.17]$	$[-0.05, 0.15]$	$[-0.04, 0.12]$	$[-0.24, 0.19]$	$[-0.13, 0.12]$	$[-0.12, 0.08]$
β_{11}	$[0.37, 0.66]$	$[0.42, 0.64]$	$[0.44, 0.62]$	$[0.41, 0.61]$	$[0.46, 0.58]$	$[0.47, 0.57]$	$[0.37, 0.57]$	$[0.44, 0.56]$	$[0.45, 0.55]$
β_{21}	$[0.20, 0.49]$	$[0.25, 0.47]$	$[0.27, 0.44]$	$[0.24, 0.43]$	$[0.29, 0.40]$	$[0.30, 0.39]$	$[0.21, 0.42]$	$[0.27, 0.40]$	$[0.29, 0.38]$
% Region Reduction[d]			64%			38%			52%
Player 2:									
θ_2	$[-3.87, -0.95]$	$[-3.86, -1.23]$	$[-3.24, -1.25]$	$[-1.65, -0.60]$	$[-1.58, -0.77]$	$[-1.47, -0.79]$	$[-2.34, -0.65]$	$[-2.25, -0.78]$	$[-2.14, -0.92]$
β_{02}	$[-0.13, 0.11]$	$[-0.06, 0.09]$	$[-0.06, 0.06]$	$[-0.06, 0.06]$	$[-0.03, 0.05]$	$[-0.03, 0.05]$	$[-0.09, 0.04]$	$[-0.05, 0.04]$	$[-0.04, 0.03]$
β_{12}	$[-0.48, -0.20]$	$[-0.45, -0.28]$	$[-0.43, -0.29]$	$[-0.41, -0.26]$	$[-0.38, -0.31]$	$[-0.37, -0.31]$	$[-0.39, -0.25]$	$[-0.38, -0.29]$	$[-0.36, -0.30]$
β_{22}	$[-0.65, -0.38]$	$[-0.64, -0.46]$	$[-0.61, -0.48]$	$[-0.57, -0.43]$	$[-0.54, -0.48]$	$[-0.54, -0.48]$	$[-0.57, -0.41]$	$[-0.55, -0.46]$	$[-0.53, -0.47]$
% Region Reduction[d]			63%			33%			46%

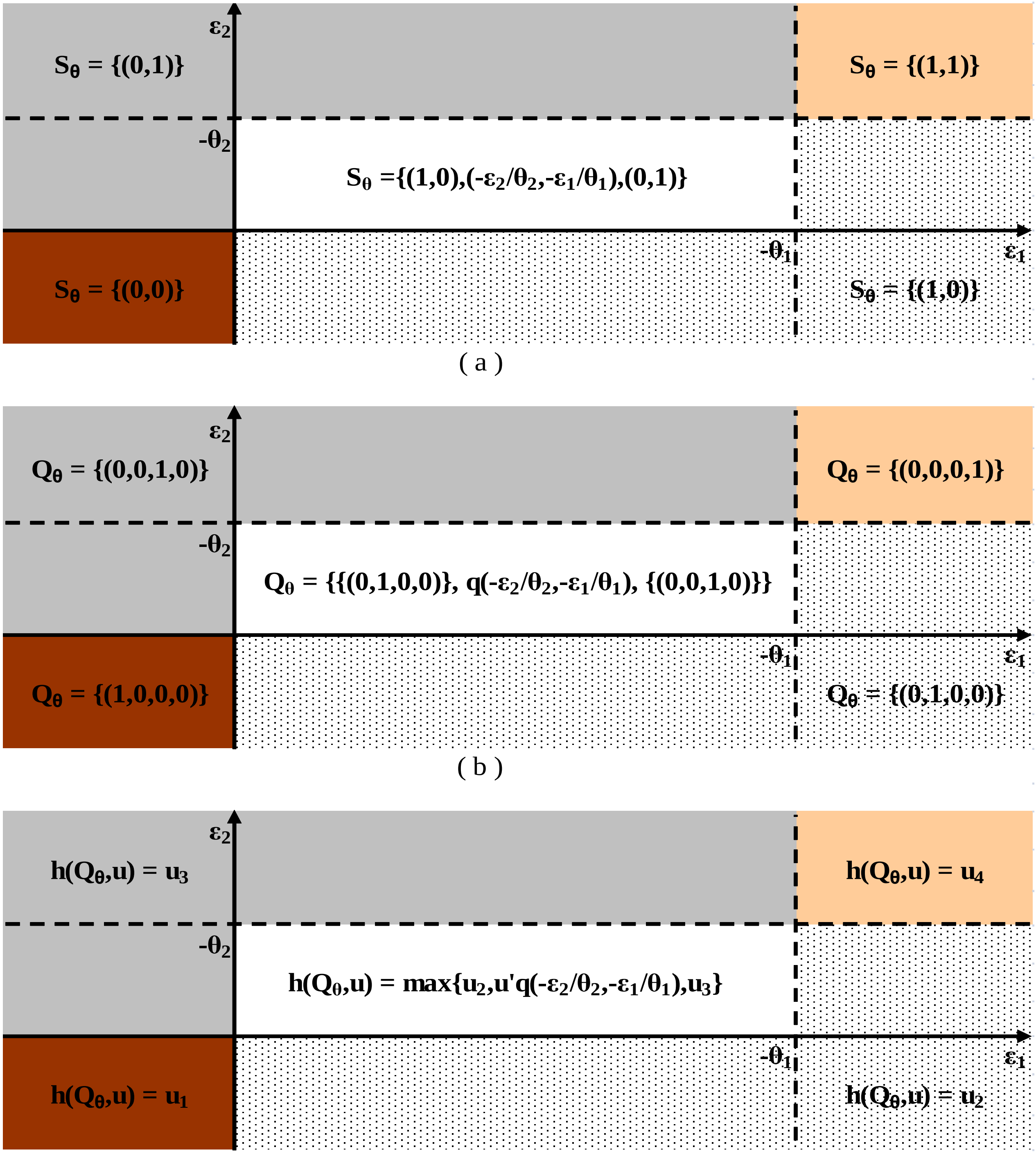
[a] $\lambda^\star = [0 \;\; \frac{3}{4} \;\; \frac{1}{4}]$, $\theta_1^\star = -1.0$, $\theta_2^\star = -1.3$, $\beta_1^\star = [0 \;\; 1/2 \;\; 1/3]$, $\beta_2^\star = [0 \;\; -1/3 \;\; -1/2]$.

[b] $\lambda^\star = [\frac{1}{2} \;\; \frac{1}{4} \;\; \frac{1}{4}]$, $\theta_1^\star = -0.8$, $\theta_2^\star = -1.1$, $\beta_1^\star = [0 \;\; 1/2 \;\; 1/3]$, $\beta_2^\star = [0 \;\; -1/3 \;\; -1/2]$.

[c] $\lambda^\star = [\frac{1}{4} \;\; 0 \;\; \frac{3}{4}]$, $\theta_1^\star = -1.2$, $\theta_2^\star = -1.5$, $\beta_1^\star = [0 \;\; 1/2 \;\; 1/3]$, $\beta_2^\star = [0 \;\; -1/3 \;\; -1/2]$.

[d] Calculated as $\frac{\text{Vol}(\Theta_O^{CT}|j)\text{-Vol}(\Theta_I|j)}{\text{Vol}(\Theta_O^{CT}|j)}$, where $\text{Vol}(\cdot|j)$ is the volume of the set in parenthesis projected on the parameters for player j (approximated by box-grid).

Figure 1: Two player entry game. Panel (a): The random set of mixed strategy NE profiles, S_θ, as a function of ε_1,ε_2. Panel (b): The random set of probability distributions over outcome profiles implied by mixed strategy NE, Q_θ, as a function of ε_1,ε_2. Panel (c): The support function in direction u of the random set of probability distributions over outcome profiles implied by mixed strategy NE, $h\left(Q_\theta, u\right)$, as a function of ε_1,ε_2.



(a)

(b)

(c)

Figure 2: A comparison between the logic behind the approaches of ABJ, CT, and this paper, obtained by projecting in $\Re^2 : \Delta^{\kappa_{\mathcal{Y}}-1}$, $\mathbb{E}\left(Q_\theta|\,\underline{x}\right)$, and the hypercubes used by ABJ and CT. A candidate $\theta \in \Theta$ is in Θ_I if $\mathbf{P}\left(y|\,\underline{x}\right)$, the white dot in the picture, belongs to the black ellipses $\mathbb{E}\left(Q_\theta|\,\underline{x}\right)$, which gives the set of probability distributions consistent with *all* the model's implications. The same θ is in Θ_O^{CT} if $\mathbf{P}\left(y|\,\underline{x}\right)$ belongs to the red region or to the black ellipses, which give the set of probability distributions consistent with the subset of model's implications used by CT. The same θ is in Θ_O^{ABJ} if $\mathbf{P}\left(y|\,\underline{x}\right)$ belongs to the yellow region or to the red region or to the black ellipses, which give the set of probability distributions consistent with the subset of model's implications used by ABJ.

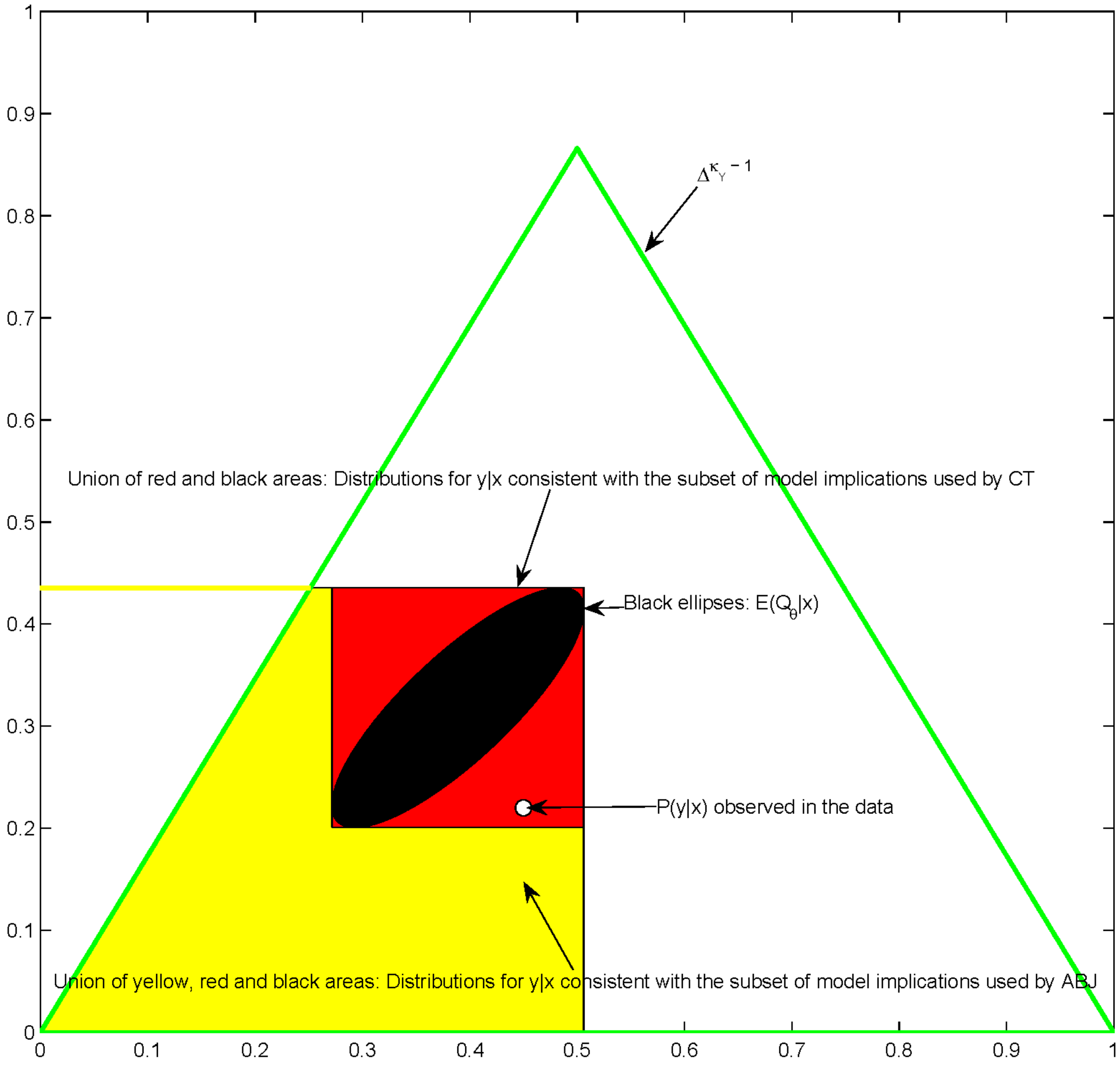


Figure 3: Identification regions in a two player entry game with mixed strategy Nash equilibrium as solution concept, for three different DGPs, see Table 1.