ANNUAL REVIEWS

Annual Review of Economics

Set Identification, Moment Restrictions, and Inference

Christian Bontemps[1,2] and Thierry Magnac[1]

[1]Toulouse School of Economics, Université de Toulouse, 31000 Toulouse, France; email: christian.bontemps@tse-fr.eu, thierry.magnac@tse-fr.eu

[2]École Nationale de l'Aviation Civile, 31055 Toulouse, France

Annu. Rev. Econ. 2017. 9:103–29

First published as a Review in Advance on April 7, 2017

The *Annual Review of Economics* is online at economics.annualreviews.org

https://doi.org/10.1146/annurev-economics-063016-103658

JEL codes: C10, C40

Keywords

set identification, moment inequality, convex set, support function

Abstract

For the past 10 years, the topic of set identification has been much studied in the econometric literature. Classical inference methods have been generalized to the case in which moment inequalities and equalities define a set instead of a point. We review several instances of partial identification by focusing on examples in which the underlying economic restrictions are expressed as linear moments. This setting illustrates the fact that convex analysis helps not only for characterizing the identified set but also for inference. From this perspective, we review inference methods using convex analysis or inversion of tests and detail how geometric characterizations can be useful.

1. INTRODUCTION

The importance of the standard notion of point identification, which appears in standard econometric textbooks (for instance, Wooldridge 2010), has been questioned for the past 30 years, especially by Manski and his coauthors (initiated in Manski 1989), who reintroduce and develop the notion of set or partial identification in the literature. Many other scholars have followed up, contributing to a blossoming literature on selection models, structural models, and models of treatment effects. Seminal work has been developed by Gini (1921) and Frisch (1934) for the simple regression model with measurement errors, Reiersol (1941) and Marschak & Andrews (1944) for simultaneous equation models, Hoeffding (1940) and Fréchet (1951) for bounds on the joint distributions of variables when only marginal distributions are observed (in two different surveys, for example), and Klepper & Leamer (1984) and Leamer (1987) for the linear regression model with measurement errors on all variables. This work remained little known or used until the work of Manski, which he himself has summarized (Manski 2003). Many of Manski's students helped to develop this literature [see, in particular, the surveys by Tamer (2010) and Molchanov & Molinari (2015)].

The general reasoning that leads to partial identification is the notion of the incompleteness of data or models. First, the data may be incomplete because of censorship mechanisms, the use of two different databases, or the existence of two exclusive states of treatment. For the evaluation of public policies, observational treatment data are necessarily incomplete because individuals can never be observed simultaneously in treatment and off treatment. Second, structural models can be incomplete if they do not specify unambiguous solutions. A classic example of this scenario is provided by multiple equilibria in games (e.g., Tamer 2003). The economic model does not specify the selection mechanism (stochastic or not) of the observed equilibrium.

The most common procedure in the applied literature is to make assumptions or add new information to complete the data and obtain point-identified models. For example, we would specify additional latent variables and their distributions to supplement the data (as in models of censorship or treatment) or mechanisms that make the solution unique in economic models (as an equilibrium selection mechanism in a game). However, choosing a single completion of the model is arbitrary, and point identification becomes implausible.

Even so, this approach provides the first insight into partial identification. Data analysis could still be conducted by examining all acceptable arbitrary assumptions that are consistent with the model assumptions and by collecting all values of point-identified parameters implied by each of these assumptions. The acceptability of an hypothesis depends on the application, and these assumptions either refer to sets (e.g., a probability of equilibrium selection belongs to the interval [0, 1], or censored values are bounded) or are functional (like monotonicity or concavity). The identifying power of different assumptions may be compared in terms of the size of the set that is identified.

As presented above, partial identification seems to be very different from the traditional setting. However, when we include the other steps in empirical work of estimating and constructing confidence intervals, this notion of partial identification fits naturally, at least when the identified set is connected. At the estimation stage, we can replace point estimates of the parameter of interest with point estimates of the boundary of the set. In inference, the presentation using confidence regions does not change because results naturally express themselves in terms of intervals or confidence regions. Only their interpretation changes, as these confidence regions are not only the result of sampling variability but also of radical uncertainty about the identification of the underlying parameters.

Sections 2 and 3 are devoted to identification and Section 4 to a review of inference methods. Two important elements in this literature are the issues of sharp identification and of uniform inference, both of which we define. For simplicity, we focus mainly on partial observability settings in which the original moment restrictions are linear. This setting is very attractive because these examples illustrate that convex analysis helps in identification, estimation, and inference. The idea that the geometry of the problem might be used in partial identification has received little coverage in the literature.

It is indeed often the case [as shown by Beresteanu et al. (2011)] that the identified set is convex or that all points in the identified set can be characterized using an auxiliary convex set. This reduces the dimensionality of the problem tremendously because the space of convex sets, by being homeomorphic to their support functions (as we define below), has a much smaller dimensionality than the space of general sets. Furthermore, convex analysis helps not only in proving the efficiency of inference procedures but also, in practice, in constructing standard test statistics. This is one of the threads that we follow in this review by devoting Section 3 to what we call convex set identification. This section prepares the ground for discussing its implications for inference.

In Section 5, we briefly review the empirical literature that, although growing, still lags behind the recent expansion of the theoretical literature. We believe that the way in which empiricists use these methods will lead to further improvements in theoretical developments. We try to explore the specific challenges that the empirical literature faces when trying to apply the theoretical recommendations. Finally, Section 6 concludes.

2. POINT AND SET IDENTIFICATION

As is usual, we begin by abstracting from sampling issues and analyzing how the parameters of economic models can be recovered from the probability distribution functions of economic variables. We reserve capital letters to denote sets, e.g., $\Theta_I \subset \mathbb{R}^d$ for the identified set, and lowercase letters for elements of these sets, e.g., θ. We use lowercase bold letters to denote single or multidimensional random variables, e.g., $\mathbf{w}$. We focus on the practical and empirical issues implied by partial identification without paying much attention to the mathematical foundations (see Molchanov & Molinari 2015) or the theory of random sets (see Molchanov 2005).

This section defines, first, concepts of observational equivalence and point identification and, second, the notions of complete and incomplete models as well as of sharp identification. We then link these concepts with moment inequalities that, in most if not all of the partial identification literature, characterize the identified set. In other words, estimating equations are expressed as inequality restrictions on population moments or probabilities.

We start by presenting definitions and a simple example and then broadly refer to the literature on partial identification.

2.1. Setup and Definitions

We adopt a setting in which random variables, say $\mathbf{w}$, the interrelationships of which are described by an economic model, are defined on a probability space in which the space of elementary events is, for simplicity, a subset of the Euclidean space, $\mathbb{R}^p$, and the probability measure is a family of probabilities $P_{\theta,\eta}$. Our framework covers semiparametric models because the population probability distribution depends on the finite-dimensional parameter of interest $\theta \in \Theta \subset \mathbb{R}^d$, the true value of which is θ_0, and on other nuisance parameters η. These nuisance parameters, the true values of which are η_0, can be as general as one wants (for instance, they can be distribution functions). They are the source of partial identification, as defined below. Furthermore, some of those nuisance parameters are kept in the background and are supposed to be point identified (for instance, the marginal distributions of exogenous covariates).

When nuisance parameters η are fixed at the supposedly known true value η_0 or, more simply, when there are no such nuisance parameters, we can first define the concept of observational equivalence. Parameters θ and θ' are said to be observationally equivalent if and only if

$$\Pr(\mathbf{w} \leq w; \theta, \eta_0) = \Pr(\mathbf{w} \leq w; \theta', \eta_0) \text{ almost surely.}$$

Second, the parameter θ is said to be point identified if there is no $\theta \in \Theta$ that is observationally equivalent to the true parameter θ_0, holding η_0 fixed. This definition can be global or local depending on the assumptions about the range of variation of θ. We could also be interested in a subset of parameters θ.

In this definition, we maintain that the specification is correct and unique, i.e., that the population probability measure is given by a unique set of parameters (θ_0, η_0), and η_0 is known. It is easy to extend this concept to misspecified models by using a notion of distance between the population probability measure and the family of semiparametric probability distributions generated by (θ, η). In particular, if the assumption that $\eta = \eta_0$ is incorrect, the model is misspecified and, even if point identified, will generally generically deliver an incorrect parameter θ. We see below that the partial identification technique is a way to protect oneself against this type of misspecification. However, the notion of partial identification lends itself less well to cases of misspecification of the family $P_{\theta,\eta}$ (Ponomareva & Tamer 2011).

Point identification can break down if the nuisance parameters η_0 are not known or cannot be point identified using the relationship between the population probability distribution P_{θ_0,η_0} and the family $P_{\theta,\eta}$. The model is said to be incomplete when it delivers several probability measures $P_{\theta,\eta}$ that are all compatible with the population distribution function. In contrast, a model is said to be complete when its parameters are point identified.

2.1.1. Completing the model or the data. Completing a model might require some ingenuity on the part of the researcher. There are two ways to make a model complete. First, we can specify the unobserved parameter η as above and set it to η_0 (sometimes by augmenting θ with a few parameters). For instance, assuming normality of the error term completes a binary model into a probit model. Second, we may adopt a completion process by augmenting the data with a random variable $\mathbf{t}$ so that observables are now $(\mathbf{w}, \mathbf{t})$. For example, an interval-censored variable can be completed by an arbitrary but compatible random variable t, which describes the true unknown position of the variable within the interval. This additional variable can also describe the selected equilibrium in games with multiple equilibria. This completion fixes the value of the unknown nuisance parameter η, which is now interpreted in the most general sense as the distribution of variable $\mathbf{t}$ conditional on observable $\mathbf{w}$.[1]

This dual presentation makes clear that incompleteness is related to both the data and the model. Completing the data can make the model complete. Completing the model can make the data informative about the model. In this deeper sense, partial identification is related to the credibility of models and their assumptions and to the exploration of the impact of these assumptions (Manski 2003). It distinguishes the core economic variables, $\mathbf{w}$, from auxiliary variables, $\mathbf{t}$, and aims to study the impact of the specification of the distribution function of variables $\mathbf{t}$ on the parameter θ. This is expressed in the literature by saying that the nuisance parameter η is not specified by economic theory or by the statistical model, even if some restrictions might apply to it.

[1]The population probability function is the marginal distribution function of the complete function.

2.1.2. Set identification.

If we define the identified set as the set of all possible values of the point-identified parameter when the completion is described by a value of η belonging to a set $\mathcal{E}$, which is specific to each application, we find that

$$\begin{aligned}\Theta_I &= \{\theta; \exists \eta \in \mathcal{E}, \Pr(\mathbf{w} \le w, \mathbf{t} \le t; \theta, \eta) = \Pr(\mathbf{w} \le w, \mathbf{t} \le t; \theta_0, \eta_0), \forall (w,t) \in \mathcal{W} \times \mathcal{T}\} \\ &= \bigcup_{\eta \in \mathcal{E}} \{\theta; \Pr(\mathbf{w} \le w, \mathbf{t} \le t; \theta, \eta) = \Pr(\mathbf{w} \le w, \mathbf{t} \le t; \theta_0, \eta_0), \forall (w,t) \in \mathcal{W} \times \mathcal{T}\}. \qquad 1.\end{aligned}$$

In other words, the identified set contains all values of the parameter of interest that can be reconciled with the data for at least one value of the parameter that completes the data or model.

In the absence of other restrictions on set $\mathcal{E}$, it is unlikely that Θ_I is different from the whole possible set Θ. First, there can be restrictions on the support, say $\mathcal{T}$, of the random variable $\mathbf{t}$, for example, an interval in the case of interval censoring or in the case in which the conditional probability (on exogenous variables) of the equilibrium selection, in a game with multiple equilibria, is bounded between 0 and 1. Further restrictions can be imposed, e.g., on the shape or monotonicity of functional forms or by excluding variables. All of these restrictions are written as restrictions on the parameter $\eta \in \mathcal{E}$ that can be analyzed according to their degree of credibility:

$$\eta \in \mathcal{E}_1, \eta \in \mathcal{E}_2 \subset \mathcal{E}_1, \ldots, \eta \in \{\eta_0\} \subset \mathcal{E}_m.$$

Finally, the sharp identification of a set is defined as asserting that all points in the identified set, Θ_I, correspond to an acceptable or credible assumption that completes the partially identified model.

The following simple example serves to illustrate this construction and to introduce the concepts. This example is developed further in Section 3.

2.2. Example 1: Interval Censoring and Best Single-Dimensional Linear Prediction

Stoye (2007) analyzes the case of linear prediction,

$$\mathbf{y}^* = \beta_0 + \beta_1 \mathbf{x} + \mathbf{u}, \qquad E(\mathbf{u}) = E(\mathbf{ux}) = 0,$$

in which the dependent variable is interval censored:

$$\mathbf{y}^* \in [\mathbf{y}_L, \mathbf{y}_L + \mathbf{d}]. \qquad 2.$$

Only the lower bound $\mathbf{y}_L$, the length $\mathbf{d}$, and a single covariate $\mathbf{x}$ are observed.[2] The extension to linear prediction in a multivariate model is presented in the next section. To simplify further, we assume that $E\mathbf{x} = 0$ and $E\mathbf{x}^2 > 0$ and suppose that there are no other restrictions.

This is a fairly common scenario when using data on income or household wealth because many surveys proceed through a two-stage approach. First, researchers ask households or individuals the exact level of their income or assets, and second, if households do not want to answer for privacy reasons, they ask the same question but in the form of intervals (e.g., "Is your income between \$0 and \$500? Or between \$500 and \$1,000?"). Even if \$0 is a natural lower bound for income, an upper bound is not clearly defined. Most studies then make an arbitrary assumption about the

[2]We consider the closure of the interval, although it would be natural to opt for a cadlag assumption for the true interval and open it on the right. These topological distinctions are neglected in this literature or are treated in technical appendices. In this example, it is clearly legitimate if the distribution of $\mathbf{y}^*$ is continuous because the right end is of measure 0. This is the purpose of the assumption stating the nonatomicity of probability distribution functions that is introduced below.

maximum amount of the dependent variable, such as the highest observed income (e.g., Lee 2009, in which the most conservative bounds are used).

We focus on parameter β_1 alone and proceed first by completing the data to obtain point identification. Using the bounds in Equation 2 on the unobserved outcome $\mathbf{y}^*$, we augment the data by choosing random $\mathbf{t}$ on support $\mathcal{T} = [0, 1]$, so that we can write

$$\mathbf{y}^* = \mathbf{y}_L + \mathbf{td}.$$

Note that the unknown parameter η is the conditional distribution of $\mathbf{t}$, $\eta = F(\mathbf{t} \leq t \mid \mathbf{y}_L, \mathbf{d}, \mathbf{x})$, so that it covers cases of parametric completion. If we were to use ordered probit or logit, for instance, the unobserved variable, $\mathbf{y}^*$, would be specified as normal or logistic conditionally on $\mathbf{x}$, and η would be set to an interval-truncated normal or logistic distribution. Under any of these assumptions, the parameter of interest β_1 is generically identified if there are more than three intervals.[3] If we do not want to adopt such parametric assumptions, the identified set is much larger than the singletons identified by ordered probit or logit. This set includes point-identified parameters derived by considering all possible distribution functions η of the variable $\mathbf{t}$.

The analysis with nonparametric completion proceeds as in the general definition. First, identify parameter β_1 in every completed model. Second, consider the union of all point-identified values.

As we can write

$$\mathbf{y}^* = \mathbf{y}_L + \mathbf{td} = \beta_0 + \beta_1 \mathbf{x} + \mathbf{u},$$

we can derive the value of parameter β_1 as[4]

$$\beta_1 = \frac{E((\mathbf{y}_L + \mathbf{td})\mathbf{x})}{E(\mathbf{x}^2)} = \frac{E(\mathbf{y}_L\mathbf{x})}{E(\mathbf{x}^2)} + \frac{E(\mathbf{tdx})}{E(\mathbf{x}^2)}.$$

As $\mathbf{d} \geq 0$ and $\mathbf{t} \in [0, 1]$, we find that

$$\begin{aligned} E(\mathbf{tdx}) &= E(\mathbf{tdx1}\{\mathbf{x} > 0\}) + E(\mathbf{tdx1}\{\mathbf{x} < 0\}) \\ &\leq E(\mathbf{dx1}\{\mathbf{x} > 0\}), \end{aligned}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function of the bracket. Symmetrically, we obtain

$$E(\mathbf{tdx}) \geq E(\mathbf{dx1}\{\mathbf{x} < 0\}).$$

The identified interval for β_1 is then the union of all possible values,

$$\beta_1 \in \Theta_I = \left[\frac{E(\mathbf{y}_L\mathbf{x}) + E(\mathbf{dx1}\{\mathbf{x} < 0\})}{E(\mathbf{x}^2)}, \frac{E(\mathbf{y}_L\mathbf{x}) + E(\mathbf{dx1}\{\mathbf{x} > 0\})}{E(\mathbf{x}^2)} \right]. \qquad 3.$$

Its length is always positive if

$$\frac{E(\mathbf{d}\,|\mathbf{x}|)}{E(\mathbf{x}^2)} > 0$$

and, specifically, when the interval length $\mathbf{d}$ is a nonnegative random variable not always equal to zero, so that both exact and interval-censored values are observed.

Conversely, one can show through a constructive argument that any point in this range corresponds to a possible distribution of $\mathbf{t}$ over its support $[0, 1]$. This shows that the interval is identified sharply (e.g., Stoye 2007, Magnac & Maurin 2008).

[3]The parameter β_0 and the variance of $\mathbf{u}$ are also identified. This requires having a strictly negative definite Hessian of the log-likelihood in a neighborhood of $\theta_0 = (\beta_0, \beta_1, \sigma^2)$ (see Rothenberg 1971).

[4]Recall that $E\mathbf{x} = 0$ and the variance of $\mathbf{x}$ is $E(\mathbf{x}^2) > 0$.

Finally, note that parameter β_1 can also be defined as the solution to two unconditional moment inequalities,

$$\begin{aligned} E(\mathbf{y}_L\mathbf{x}) + E(\mathbf{dx1}\{\mathbf{x} < 0\}) - \beta_1 E(\mathbf{x}^2) &\leq 0, \\ \beta_1 E(\mathbf{x}^2) - E(\mathbf{y}_L\mathbf{x}) - E(\mathbf{dx1}\{\mathbf{x} > 0\}) &\leq 0. \end{aligned} \tag{4}$$

Alternatively, the assumption of uncorrelated errors could be strengthened into mean-independent errors. This yields conditional moment inequalities of the form

$$\begin{aligned} E(\beta_0 + \beta_1\mathbf{x} - \mathbf{y}_L \mid \mathbf{x}) &\leq 0, \\ E(\mathbf{y}_L + \mathbf{d} - \beta_0 - \beta_1\mathbf{x} \mid \mathbf{x}) &\leq 0. \end{aligned}$$

2.3. Discussion

The resulting setup of moment restrictions obtained in the above example extends to many partially identified economic models. They express the identifying restrictions on parameters as inequality constraints on expectations of linear or nonlinear functions of variables and parameters and therefore lead to a finite or infinite number of moment inequalities. A more difficult issue is sharp identification, in which the characterization by moment inequalities is equivalent to the characterization of the set by the completeness restrictions. If this is not the case, what is identified is a so-called outer set, which is generally much easier to determine because the number of restrictions is smaller (see Ciliberto & Tamer 2009 for such an empirical strategy).

These extensions require more sophisticated tools than those we used in the very simple example above. In the case of structural models, in particular those derived from game theory, Galichon & Henry (2009, 2011) explain how to use tools from optimal transport methods to solve the issue of sharp identification and derive moment inequalities that are necessary and sufficient for characterizing the identified set. Alternatively, Beresteanu et al. (2012) explain how the theory of random sets also enables one to find a solution to these issues. We briefly summarize the tools of random set theory in the next section.

These tools are applicable to models analyzing censorship, such as those developed by Horowitz & Manski (1995), Manski & Pepper (2000), or, more generally, all the works reviewed by Manski (2003). Many topics are connected with the framework of partial identification. Ridder & Moffitt (2007) offer a comprehensive overview of models of data coming from multiple sources, such as two surveys or two mutually exclusive states of the world, and Pacini (2017) develops a particular case. Models with discrete variation within a framework of simultaneous equations are investigated by Chesher (2005, 2010). Polytomous discrete models are treated by Chesher & Smolinski (2012), and Chesher & Rosen (2015a) generalize instrumental variable models. Binary models with a "very exogenous" regressor, observations of which are interval censored, are analyzed by Magnac & Maurin (2008). Davezies & d'Haultfoeuille (2012) deal with attrition and departures from the missing at random assumption. Partial identification of variance and covariance parameters is studied by Horowitz & Manski (2006), Fan & Park (2010), Fan & Wu (2010), and Gomez & Pacini (2013). Nevo & Rosen (2012) and Conley et al. (2012) introduce what they call "imperfect instruments," which are variables that are not excluded from the equation of interest but are less correlated with the error term than the endogenous variable they are supposed to instrument.

3. DIRECT AND INDIRECT USES OF CONVEXITY ARGUMENTS

Some reminder of random set theory is useful to make this review self-contained, and we start by borrowing notation from the survey of Molchanov & Molinari (2015). We turn next to the definition of the support function of a convex set. More substantially for our topic, we develop three cases in which direct and indirect approaches with these tools help in deriving conditional or unconditional moment inequalities.

3.1. Random Sets and Random Selections

In Section 2, we saw the importance of defining the completion of data by a random variable $\mathbf{t}$, the support of which is restricted or to which other restrictions are applicable. In the theory of random sets, a specific random variable satisfying these restrictions is called a selection, and this selection is from a random set that gathers all possible random variables that satisfy these restrictions, say a random set $\mathbf{T}$.[5] We assume that the random set $\mathbf{T}$ is closed and bounded, and therefore compact, if the support of $\mathbf{t}$ is included in a finite-dimensional Euclidean space.

As parameter $\theta = \theta(\mathbf{t})$ is point identified when the completion is given by $\mathbf{t}$, the definition of the sharply identified set can be rephrased as

$$\Theta_I = \{\theta; \theta = \theta(\mathbf{t}); \mathbf{t} \in \mathbf{T}\}.$$

Random set theory helps us to do two things. First, it relates what are called the capacity and containment functionals of a random set $\mathbf{T}$ to the distribution function of observed variables, $\mathbf{w}$. Second, it relates the identified set Θ_I with the so-called Aumann expectation of specific random sets either directly or indirectly, as shown below. As explained by Beresteanu et al. (2011), the choice between these two methods depends on the type of restrictions that are imposed in each economic application. Best linear prediction or, more generally, mean independence restrictions are usually easier to deal with using Aumann expectations. In contrast, games of complete or incomplete information or independence restrictions are easier to deal with using capacity functionals (see also Chesher & Rosen 2015a).

3.2. Aumann Expectations and Support Functions

As we focus our survey on partial identification derived from moment restrictions, we concentrate on the use of Aumann expectations, although Section 4, which deals with inference, generally encompasses both frameworks. Defining the concept of Aumann expectation comes first. As in the work of Molchanov (2005), the Aumann expectation of a random set $\mathbf{T}$ is the set formed by the expectations of all its selections:

$$\mathbb{E}(\mathbf{T}) = \{E(\mathbf{t}); \mathbf{t} \in \mathbf{T}\}.$$

A key property of this expectation is that the resulting set is closed and convex in $\mathbb{R}^p$ under weak conditions. This opens up the possibility of using standard tools of convex analysis (Rockafellar 1970). Although a convex set can be uniquely characterized by several functions, the literature has focused on using support functions because the most commonly used distance between two convex sets, the Hausdorff distance, is the supremum of the difference of their respective support functions.[6]

[5] In this review, we identify a random set with the set of its random selections and write $\mathbf{t} \in \mathbf{T}$. A rigorous approach is given by Molchanov (2005).

[6] The embedding theorem of Hörmander (see Molchanov 2005) between convex sets and support functions is also an important motivation.

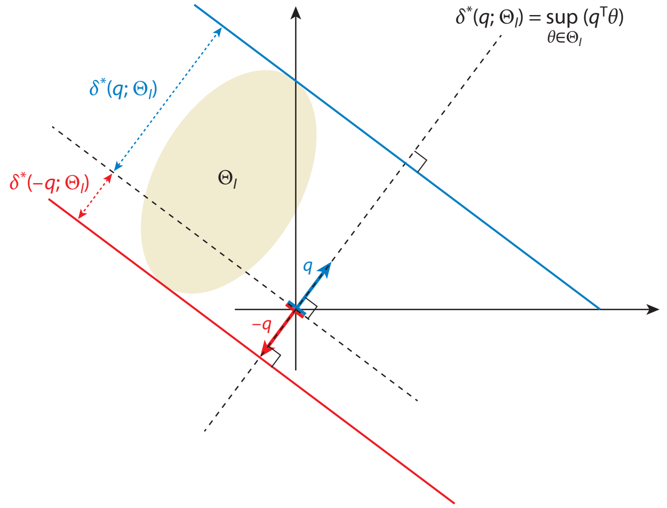

Figure 1

The support function of Θ_I, a convex set, in the direction q is the distance between the origin and the supporting hyperplane orthogonal to vector q. It is positive if both the convex set and the origin belong to the same half space bounded by this hyperplane and negative otherwise. We report the support function values for q and $-q$.

The support function of a convex set Θ is defined as

$$\delta^*(q;\Theta) = \sup_{\theta\in\Theta}(q^\top\theta)$$

for all directions $q \in \mathbb{R}^d$, which uniquely characterizes the convex set Θ (e.g., Rockafellar 1970):

$$\theta \in \Theta \Leftrightarrow \forall q \in \mathbb{R}^d, q^\top\theta \leq \delta^*(q;\Theta). \qquad 5.$$

This construction is illustrated in **Figure 1**. The support function of a convex set is defined by the location of its supporting hyperplanes in all directions.

Furthermore, support functions are sublinear functions, i.e., positive homogeneous and convex. The previous characterization can therefore be equivalently written for directions on the unit sphere $\mathbb{S}^{d-1} = \{q \in \mathbb{R}^d; \|q\| = 1\}$:

$$\theta \in \Theta \Leftrightarrow \forall q \in \mathbb{S}^{d-1},\ q^\top\theta \leq \delta^*(q;\Theta).$$

The same property also leads to theorem 2.1.22 of Molchanov (2005), which says that the support function of an Aumann expectation is equal to the expectation of the support function of the underlying random set:[7]

$$E\,[\delta(q;\mathbf{T})] = \delta(q;\mathbb{E}(\mathbf{T})). \qquad 6.$$

There are various uses of these results in the literature. First, a direct approach using Aumann expectations is developed by Beresteanu & Molinari (2008) in the case of best linear prediction with interval-censored outcomes studied by Stoye (2007). In this case, the identified set is a function of the Aumann expectation of a random set $\mathbf{T}$, and realizations of this random set are observed quantities in the sample. Second, another direct approach is used by Bontemps et al. (2012) in

[7] Its conditions of validity are that $\mathbf{T}$ is integrably bounded and convex and that the underlying probability space is nonatomic. This does not seem restrictive in most economic applications (Beresteanu et al. 2011).

the same case of best linear prediction with censored-by-interval outcomes, although the number of moment conditions in this case is larger than the number of parameters. The identified set is the intersection between two convex sets, one of which is a transform of an Aumann expectation. Finally, an indirect approach is proposed by Beresteanu et al. (2011). In this case, for any value of θ, there exists a convex set $M(\theta)$ that is itself an Aumann expectation of a random set, and the identified set can be characterized as

$$\theta \in \Theta_I \iff 0 \in M(\theta).$$

We next review these approaches through simple examples and derive the moment inequalities that each of them imply.

3.3. Convex Identified Sets: A Direct Approach

Example 1 in Section 2.2 can be extended to a multidimensional framework starting from the same linear prediction (e.g., Stoye 2007):

$$\mathbf{y}^* = \mathbf{x}\beta + \mathbf{u}, \mathbf{y}^* \in [\mathbf{y}_L,\ \mathbf{y}_L + \mathbf{d}],$$

where $E(\mathbf{x}^\top \mathbf{u}) = 0$. If we complete the data, parameter β belongs to the identified set, $\Theta_I \subset \mathbb{R}^d$, if and only if there exists a variable $\mathbf{t}$ whose distribution function is $\eta = F(\cdot \mid \mathbf{x}, \mathbf{d}, \mathbf{y}_L)$ on $[0, 1]$ such that

$$\mathbf{y}^* = \mathbf{y}_L + \mathbf{t}\mathbf{d}.$$

The point-identified parameter β using complete data is

$$\beta = \left[E(\mathbf{x}^\top \mathbf{x})\right]^{-1} E[\mathbf{x}^\top(\mathbf{y}_L + \mathbf{t}\mathbf{d})], \tag{7.}$$

and the identified set is the collection of such expressions. It is convex because the support $[0, 1]$ of $\mathbf{t}$ is convex.

The random set of interest is defined by

$$\mathbf{M} = \{\mathbf{x}^\top(\mathbf{y}_L + \mathbf{t}\mathbf{d}); \mathbf{t} \in \mathbf{T}\},$$

which is also convex with Aumann expectation $\mathbb{E}(\mathbf{M})$. Its estimation and the construction of confidence intervals are derived by Beresteanu & Molinari (2008) using laws of large numbers and central limit theorems for random sets. They also deal with the complication that the identified set is a transformation of this Aumann expectation [i.e., premultiplying by $\left(E(\mathbf{x}^\top \mathbf{x})\right)^{-1}$].

In addition, Equation 7 allows us to write that for all $q \in \mathbb{S}^{d-1}$,

$$\delta^*(q; \Theta_I) = \sup_{\beta \in \Theta_I} q^\top \beta = \sup_{\mathbf{t} \in \mathbf{T}} q^\top \left[E(\mathbf{x}^\top \mathbf{x})\right]^{-1} E[\mathbf{x}^\top(\mathbf{y}_L + \mathbf{t}\mathbf{d})].$$

Simple calculations by Stoye (2007) yield the support function as a function of population moments,

$$\delta^*(q; \Theta_I) = q^\top \left(E(\mathbf{x}^\top \mathbf{x})\right)^{-1} E\left(\mathbf{x}^\top(\mathbf{y}_L + \mathbf{1}\left\{q^\top \left(E(\mathbf{x}^\top \mathbf{x})\right)^{-1} \mathbf{x}^\top > 0\right\} \mathbf{d})\right),$$

and the estimation of the identified set can be equivalently achieved by estimating support functions.

The geometry of the set Θ_I has consequences for inference, as developed by Bontemps et al. (2012). Specifically, two important characteristics of frontiers of convex sets are exposed faces, which are nontrivial (i.e., not reduced to singletons), and kinks (or corner points). An exposed face is the intersection between a supporting hyperplane, defined by its outer normal vector q_0, and convex set Θ_I:

$$B(q_0) = \{\beta \in \Theta_I; q_0^\top \beta = \delta^*(q_0; \Theta_I)\}.$$

When Θ_I is strictly convex, the previous set is trivial because it is reduced to a singleton in every direction.

The second characteristic is that convex sets can have kinks (or corner points) when the set of supporting hyperplanes orthogonal to q at a point β_0 of the frontier of Θ_I,

$$C(\beta_0) = \{q \in \mathbb{S}^{d-1}, q^\top \beta_0 = \delta^*(q; \Theta_I)\},$$

is not reduced to a singleton.

In the specific example of best linear prediction, the first characteristic arises when at least one covariate has a mass point, and the second characteristic arises when the density function of covariates is not positive everywhere on its support (Bontemps et al. 2012).

The existence of nontrivial exposed faces has an impact on the asymptotic distribution of estimates, which we review in the next section. The existence of kink points affects the number of moment inequalities that are binding, an important point in inference also developed in the next section. Indeed, as Equation 5 makes clear, necessary and sufficient moment inequalities at any point of the identified set are

$$\forall \beta \in \Theta_I, \forall q \in \mathbb{S}^{d-1}; q^\top \beta - \delta^*(q; \Theta_I) \leq 0.$$

Consequently, for interior points of Θ_I, no inequality restriction is binding. For frontier points of Θ_I that are not kinks, a single inequality is binding. And finally, for any kink frontier point of Θ_I, many inequalities indexed by directions q in the nonsingular cone, $C(\beta_0)$, are binding.

3.4. Convex Identified Sets: A Two-Step Approach

In this section, we consider another extension of Example 1, with a single covariate, $\mathbf{x}$, such that $E\mathbf{x} = 0$. Restrictions $E(\mathbf{xu}) = E(\mathbf{u}) = 0$ are now completed by another restriction $E(\mathbf{zu}) = 0$, and we analyze how this additional information restricts the information set. For the sake of exposition, we also suppose that the covariate $\mathbf{z}$ is single dimensional, $E(\mathbf{z}) = 0$, and uncorrelated with $\mathbf{x}$.[8]

To begin, observe that the presence of this instrument imposes a restriction on the random selection parameter $\mathbf{t}$. For instance, consider the selection $\mathbf{t} = \mathbf{1}\{\mathbf{x} > 0\}$ that leads to the largest value for β_1 in Equation 3, say β_1^U. It corresponds to the true outcome

$$\mathbf{y}^* = \mathbf{y}_L + \mathbf{1}\{\mathbf{x} > 0\}\mathbf{d} = \beta_1^U \mathbf{x} + \mathbf{u}^U$$

and

$$\begin{aligned} E(\mathbf{zu}^U) &= E[\mathbf{z}(\mathbf{y}_L + \mathbf{1}\{\mathbf{x} > 0\}\mathbf{d} - \beta_1^U \mathbf{x})] \\ &= E(\mathbf{zy}_L) + E(\mathbf{zd1}\{\mathbf{x} > \mathbf{0}\}). \end{aligned}$$

Note that $\pi = E(\mathbf{zy}_L) + E(\mathbf{zd1}\{\mathbf{x} > \mathbf{0}\})$ is observable and that, if $\pi \neq 0$, the orthogonality condition $E(\mathbf{zu}^U) = 0$ is not satisfied. Selection $\mathbf{t}$ is no longer admissible. This is also true for other selections, and the interval of identified slopes, β_1, shrinks because some random selections are ruled out.

The general geometric construction of the sharp identified set (Bontemps et al. 2012) is readily adapted to this simple example. We augment the regression by adding $\mathbf{z}$ as an additional explanatory variable,

$$\mathbf{y}^* = \beta_1 \mathbf{x} + \gamma \mathbf{z} + \mathbf{u},$$

[8]By relabeling $\mathbf{z}$ as the residual of the linear prediction of $\mathbf{z}$ on $\mathbf{x}$, this is without loss of generality.

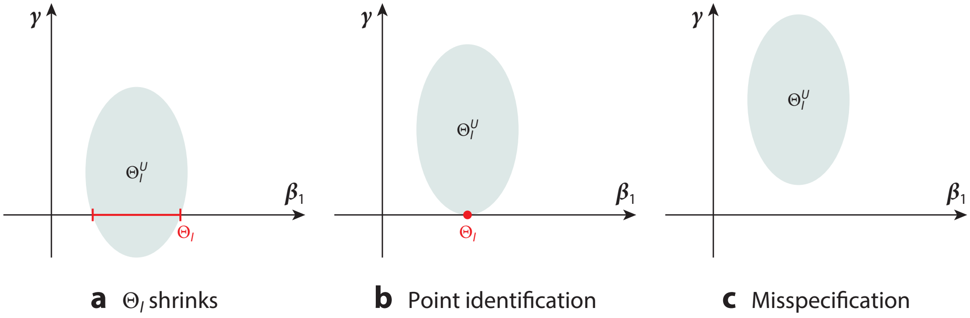

Figure 2

Geometry of different cases in which the straight line $\gamma = 0$ (*a*) can cross the interior of the unrestricted identified set, Θ_I^U, resulting in interval identification of β_1; (*b*) can be tangent to this set, restoring point identification of β_1; and (*c*) can have no intersection with Θ_I^U, a case of misspecification.

and note that parameter γ should be zero under the above assumptions.

Without restrictions on γ, there are as many parameters, (β, γ), as restrictions: $E(\mathbf{xu}) = 0$, $E(\mathbf{zu}) = 0$. Therefore, the unrestricted identified set, say Θ_I^U, is obtained, as in Section 3.3, by deriving its support function.[9] If we reconsider restriction $\gamma = 0$, the restricted identified set is the intersection of the two convex sets Θ_I^U and the hyperplane $\gamma = 0$. A standard formula (Rockafellar 1970) for the support function of the intersection of two convex sets is given by

$$\forall q_\beta \in \mathbb{S}^{d-1}; \delta^*(q_\beta; \Theta_I) = \inf_{q_\gamma \in \mathbb{R}} \delta^*\left((q_\beta, q_\gamma); \Theta_I^U\right), \qquad 8.$$

where q_β is associated with β_1 and q_γ with γ.

Moreover, the usual Sargan condition of the validity of moment restrictions, in this case $E(\mathbf{zu}) = 0$, is satisfied if this intersection is not empty, i.e., when $\gamma = 0$ is an acceptable restriction. Let us call B_{Sargan} the orthogonal projection of Θ_I^U on the space of parameter γ. The Sargan set is an interval $[\gamma_L, \gamma_U]$ in which

$$\gamma_L = E[\mathbf{z}(\mathbf{y}_L + \mathbf{1}\{\mathbf{z} < 0\}\mathbf{d})] \text{ and } \gamma_U = E[\mathbf{z}(\mathbf{y}_L + \mathbf{1}\{\mathbf{z} > 0\}\mathbf{d})],$$

and the Sargan condition can be written as

$$0 \in [\gamma_L, \gamma_U].$$

If it is not verified, the model is misspecified, and moment restrictions are incompatible with the data (see **Figure 2**).

3.5. Nonconvex Identified Sets: An Indirect Convexity Approach

There are other cases in which direct approaches cannot be used. A third extension of our original example was originally developed by Horowitz et al. (2003) and revisited by Beresteanu et al. (2011). We return to the single-dimensional best linear prediction

$$\mathbf{y}^* = \beta_0 + \beta_1 \mathbf{x}^* + \mathbf{u}, \qquad E(\mathbf{u}) = E(\mathbf{ux}^*) = 0$$

[9] As $\mathbf{z}$ is orthogonal to $\mathbf{x}$, Θ_I^U is the collection of points in $\mathbb{R}^2$ whose coordinates are, respectively, $\beta_1 = E(\mathbf{x}^2)^{-1}E[\mathbf{x}(\mathbf{y}_L + \mathbf{td})]$ and $\gamma = E(\mathbf{z}^2)^{-1}E[\mathbf{z}(\mathbf{y}_L + \mathbf{td})]$ for the same $\mathbf{t}$.

and assume now that both outcome and covariate are censored by interval:

$$\mathbf{y}^* \in [\mathbf{y}_L, \mathbf{y}_L + \mathbf{d}^y], \mathbf{x}^* \in [\mathbf{x}_L, \mathbf{x}_L + \mathbf{d}^x].$$

We complete the data by associating $\mathbf{t}^y$ on $[0, 1]$ with $\mathbf{y}^* = \mathbf{y}_L + \mathbf{t}^y \mathbf{d}^y$ and $\mathbf{t}^x$ on $[0, 1]$ with $\mathbf{x}^* = \mathbf{x}_L + \mathbf{t}^x \mathbf{d}^x$. The identified set can still be characterized as

$$\Theta_I = \left\{\beta; \beta = \left[E\left((\mathbf{x}_L + \mathbf{t}^x\mathbf{d}^x)^\top(\mathbf{x}_L + \mathbf{t}^x\mathbf{d}^x)\right)\right]^{-1} E\left((\mathbf{x}_L + \mathbf{t}^x\mathbf{d}^x)^\top(\mathbf{y}_L + \mathbf{t}^y\mathbf{d}^y)\right); (\mathbf{t}^y, \mathbf{t}^x) \in \mathbf{T}\right\},$$

but it is not necessarily convex because of the first term.

The alternative is to proceed as follows. First, fix $\theta = (\beta_0, \beta_1) \in \Theta$. Consider the random set

$$\begin{aligned}\mathbf{M}(\theta) &= \left\{\mathbf{m}_\theta = \begin{pmatrix}\mathbf{u}\\ \mathbf{u}\mathbf{x}^*\end{pmatrix}; (\mathbf{t}^y, \mathbf{t}^x) \in \mathbf{T}\right\},\\ &= \left\{\mathbf{m}_\theta = \begin{pmatrix}\mathbf{y}_L + \mathbf{t}^y\mathbf{d}^y - \beta_0 - \beta_1(\mathbf{x}_L + \mathbf{t}^x\mathbf{d}^x)\\ (\mathbf{y}_L + \mathbf{t}^y\mathbf{d}^y - \beta_0 - \beta_1(\mathbf{x}_L + \mathbf{t}^x\mathbf{d}^x))(\mathbf{x}_L + \mathbf{t}^x\mathbf{d}^x)\end{pmatrix}; (\mathbf{t}^y, \mathbf{t}^x) \in \mathbf{T}\right\}.\end{aligned}$$

Its Aumann expectation, $\mathbb{E}(\mathbf{M}(\theta))$, is convex even though the random set itself might not be, and its support function, $\delta^*(q; \mathbb{E}(\mathbf{M}(\theta)))$, characterizes $\mathbb{E}(\mathbf{M}(\theta))$. Furthermore, if $\theta \in \Theta_I$, defined by the moment restrictions $E(\mathbf{u}) = E(\mathbf{u}\mathbf{x}^*) = 0$, there exists a random selection in $\mathbf{M}(\theta)$ whose expectation is equal to 0. Therefore, we have

$$\theta \in \Theta_I \iff 0 \in \mathbb{E}(\mathbf{M}(\theta)) \iff 0 \leq \min_{q \in \mathbb{S}^1} \delta^*(q; \mathbb{E}(\mathbf{M}(\theta)))$$

because of Equation 5. As in Equation 6, $E\left[\delta(q^*; \mathbf{M}(\theta))\right] = \delta(q^*; \mathbb{E}(\mathbf{M}(\theta)))$, we can write

$$\theta \in \Theta_I \iff 0 \leq \min_{q \in \mathbb{S}^1} E(\delta^*(q; \mathbf{M}(\theta))).$$

This provides a set of moment inequalities. The support function, $\delta^*(q; \mathbf{M}(\theta))$, is easy to evaluate and can be minimized by standard techniques, although this has to be done for any candidate value of θ. In this sense, this case is significantly more costly than the cases reviewed in the previous two sections.

4. INFERENCE METHODS

Inference principles for parameters in set-identified models follow closely from those used in point-identified models. For example, estimating an interval, as in Example 1, consists of estimating its upper and lower bounds (e.g., Imbens & Manski 2004). In higher-dimensional spaces, this is somewhat more difficult unless this set is convex. These constructions are the object of this section.

In the literature, confidence sets are derived using two alternative routes. The classical approach consists, first, of estimating the identified set and, second, of constructing the confidence region as the set of points that are close to this estimate. What differs from the point-identified case is that the distribution of the distance between the confidence set and the estimated set is generally nonstandard.

In a seminal article, Chernozhukov et al. (2007) first estimate the identified set as the collection of points defined by values close to zero of a nonnegative criterion function. For a given level of confidence, they similarly define confidence regions as the set of points, the criterion value of which is smaller than a critical value that is adjusted by subsampling techniques. Alternatively, Beresteanu & Molinari (2008) and Bontemps et al. (2012) estimate the support function of the convex identified set Θ_I, as defined in Section 3.3, using empirical counterparts of population moments. Next, they construct confidence regions using the estimated sampling variability of those empirical moments.

The second approach consists of inverting a test statistic. This method has been widely used in models that are characterized by moment inequalities (see, in particular, Romano & Shaikh 2008, Andrews & Soares 2010, Andrews & Shi 2012). For any given value of θ, a test of level α of

$$H_0 : \theta \in \Theta_I, \qquad H_a : \theta \notin \Theta_I,$$

is inverted by gathering all nonrejected values of the parameter θ in the confidence region of level $1-\alpha$. Note that the classical approach detailed above also depends on the inversion of a test but is made easier by the estimation of the identified set.

We study these approaches in this section. As a preliminary, we discuss two issues that format the debates. First, it often seems reasonable to require that the inference is robust to changes in the actual, albeit unknown, probability distribution of the data. Many authors consider that this distribution varies in a wide range of probability distributions, and the inference procedure is constructed to be robust to this variation. In this case, it will be said that the inference is uniform (with respect to the considered set of probability distributions). Second, researchers must consider whether the confidence region should cover a single true value or the true identified set.

Next, we detail the general inference techniques in a moment inequality setup. In the case of the test-inversion approach, we pay attention to the issue of selecting relevant moment inequalities (Andrews & Soares 2010, Andrews & Barwick 2012). We also present techniques adapted to the convexity arguments developed in Section 3. We also review the interesting case of intersections of bounds (Chernozhukov et al. 2013), which occurs when parameters are bounded by an infinity of moments. We end this section by turning to the recently investigated issue of inference on a subvector of parameters and to a brief review of Bayesian methods.

4.1. Coverage of a Point or a Set and Uniformity

We begin with the issue of coverage of a point or a set. Suppose that the distribution function of the data is denoted P and let $\Theta_I(P)$ be the identified set, that is, all values compatible with P and structural restrictions. If we want to cover a single point θ by an interval or a confidence region I_n using an asymptotic level of confidence at least equal to $1-\alpha$, we have to find I_n as a solution of

$$\liminf_{n\to\infty} \left(\inf_{\theta\in\Theta_I(P)} \Pr(\theta \in I_n) \right) \geq 1-\alpha. \qquad 9.$$

We see that the consequence of partial identification is to replace the true value of the point-identified parameter $\theta_0(P)$ in this expression by all values in the identified set $\Theta_I(P)$.

In Example 1, developed above, of censorship of the dependent variable interval, the confidence interval will take the form $I_n = [\hat{\beta}^L_{1,n} - \hat{c}^L_n, \hat{\beta}^U_{1,n} + \hat{c}^U_n]$, where $\hat{\beta}^L_{1,n}$ and $\hat{\beta}^U_{1,n}$ are the estimators of the lower and upper bounds of the quantities defined by Equation 3 and $\hat{c}^L_n$ and $\hat{c}^U_n$ are estimators depending on the joint distribution of the estimators of the bounds and a critical value that is adjusted using Equation 9. This adjustment is made for all possible values of θ in the identified interval defined in Equation 3 and is exposed in detail by, for instance, Imbens & Manski (2004).

This construction covers a point, the supposedly single true value of the parameters. But now that the identified set has some thickness, one might want to cover regions or intervals I instead of singletons $\{\theta\}$. This is why we could search for regions I_n that satisfy the asymptotic level of confidence of at least $1-\alpha$,

$$\liminf_{n\to\infty} \left(\inf_{I\subset\Theta_I(P)} \Pr(I \subset I_n) \right) \geq 1-\alpha.$$

Most econometric applications aim to cover a point, but there are differing opinions, and the two presentations are common in the literature. For example, Romano & Shaikh (2008, 2010) study both presentations in two different articles. Note, however, that the second condition is more restrictive than the first because singletons are degenerate regions (e.g., Henry & Onatski 2012). Confidence regions covering sets are therefore generally larger than those covering points, $\theta \in \Theta_I(P)$.

In addition, the issue of uniformity can be approached in the single-dimension inference framework of Example 1. The identified interval is described by a lower and an upper bound, as in Equation 3, and these bounds are functions of population moments, which are estimated by empirical counterparts. If the two bounds are far from each other, in the sense of the metrics induced by their covariance matrix, the confidence intervals for each bound do not intersect. Consequently, a confidence region for the set (or any point of the set) is then defined by the lower value of the confidence interval of the lower bound on one side and the higher value of the confidence interval of the upper bound on the other side. However, it is clear that this construction no longer holds when the true set is small, the limiting case being a singleton. The solution to this problem is proposed by Imbens & Manski (2004) and extended by Stoye (2009). The authors construct confidence intervals, the statistical properties of which are robust to the true diameter of the identified set, that are particularly attractive when the model is point identified or close to being point identified.

Returning to the general case of covering a point uniformly, we suppose that the true data-generating process belongs to a family $\mathcal{P}$. In Example 1, which describes an interval-censored dependent variable, this family includes the case in which there is no censorship, so that the width of the observed interval is zero, $\mathbf{d} = 0$, and the parameter β_1 is identified. To accommodate this case, we would then search for a confidence interval I_n at an asymptotic level at least equal to $1-\alpha$ that satisfies

$$\liminf_{n\to\infty} \left(\inf_{P\in\mathcal{P},\theta\in\Theta_I(P)} \Pr(\theta \in I_n) \right) \geq 1-\alpha. \qquad 10.$$

In this case, too, the condition is more stringent than in the nonuniform case, and uniform confidence regions are larger than those that have been previously defined. Yet this case seems the most interesting because researchers seldom have clear ideas about the true distribution P and its range of variation. Uniformity, however, is as varied as the class of distributions, $\mathcal{P}$.

4.2. Inference in Moment Inequality Models

In this section, we describe inference techniques proposed in the recent literature mainly dealing with microeconometric models. This is why we assume, in this section, that observations are independent and identically distributed. Most of the literature has focused on sets that are defined by moment inequalities—possibly combined with moment equalities, each of which are treated as two opposite moment inequalities—and has started with a finite number of inequality conditions. Next, it was extended to the case of an infinite number of moment inequalities derived, in particular, from conditional moment inequalities.

4.2.1. Moment inequalities in finite number. Suppose that the identified set is defined by a finite number of moment inequalities:

$$\theta \in \Theta_I \iff E(h_j(\mathbf{y}, \mathbf{x}, \theta)) \leq 0 \text{ for } j = 1, \ldots, J.$$

For example, in Example 1, the arguments in Equation 4 are the two functions

$$h_1(\mathbf{y}_L, \mathbf{x}, \delta, \beta_1) = \mathbf{y}_L\mathbf{x} + \delta\mathbf{x}\mathbf{1}\{\mathbf{x} < 0\} - \beta_1\mathbf{x}^2,$$
$$h_2(\mathbf{y}_L, \mathbf{x}, \delta, \beta_1) = \beta_1\mathbf{x}^2 - \mathbf{y}_L\mathbf{x} - \delta\mathbf{x}\mathbf{1}\{\mathbf{x} > 0\},$$

the expectations of which are nonpositive.

Chernozhukov et al. (2007) were the first to consider inference for a set defined by a nonnegative criterion function that takes value zero at each point of the set. This aggregator of inequality restrictions generalizes the usual generalized method of moments (GMM) criterion for moment equalities. These authors were followed by Rosen (2008), Romano & Shaikh (2008), Andrews & Soares (2010), and many others.

Chernozhukov et al. (2007) propose the criterion

$$Q(\theta) = \sum_{j=1}^{J} a_j(\theta) \left[Eh_j(\mathbf{y}, \mathbf{x}, \theta) \right]^2 \mathbf{1}\{Eh_j(\mathbf{y}, \mathbf{x}, \theta) > 0\} \tag{11.}$$

for a positive sequence of weights $a_j(\theta)$. The value of the criterion for all points outside set Θ_I is thus quadratic in the distance to 0 of moments at this point, and therefore, we find that

$$\theta \in \Theta_I \Longleftrightarrow Q(\theta) = 0.$$

An estimate of this criterion function for a sample of size n and observations $(y_i, x_i)_{i=1,\dots,n}$ is derived from the empirical counterparts of the moments. For example, $h_{jn}(\theta) = \frac{1}{n} \sum_{i=1}^{n} h_j(y_i, x_i, \theta)$ and the population criterion $Q(\theta)$ above is replaced by its empirical analog

$$Q_n(\theta) = \sum_{j=1}^{J} a_j(\theta) \left[h_{jn}(\theta) \right]^2 \mathbf{1}\{h_{jn}(\theta) > 0\}. \tag{12.}$$

Chernozhukov et al. (2007) propose to estimate the identified set by

$$\hat{\Theta}_n = \{\theta; Q_n(\theta) < \tau_n\},$$

where τ_n is a smoothing parameter that satisfies the limit conditions[10]

$$\tau_n/\sqrt{n} \to 0, \sqrt{\ln \ln n}/\ \tau_n \to 0.$$

They also propose a direct estimation of the confidence region as

$$\hat{\Theta}_n^C = \{\theta; Q_n(\theta) < c_n^{(1-\alpha)}\}$$

at level $1 - \alpha$. We can see this construction as the inversion of a test of the hypothesis that the estimated set covers the true set. As noted by Chernozhukov et al. (2015), the test statistic $Q_n(\cdot)$ can be interpreted as a likelihood ratio test statistic.

The difficult part is to determine the critical value c_n; this is done by subsampling (Chernozhukov et al. 2007, Romano & Shaikh 2010) or by bootstrap (Bugni 2010). Canay (2010) adopts an empirical likelihood approach and also proposes an adapted bootstrap method. The authors listed in this paragraph show that the confidence region thus constructed respects the asymptotic coverage condition given by Equation 9 or by Equation 10. Nonetheless, subsampling techniques are notoriously costly in terms of computations and could perform badly in small samples.

4.2.2. Generalized moment selection. Chernozhukov et al. (2007) do not exploit the particular structure given by moment inequalities. Following the literature on inequality testing (see Silvapulle & Sen 2005), Andrews & Soares (2010) propose a method—which they call generalized

[10]Because of sampling variability, points may belong to the estimate of the identified set even if the criterion is slightly above zero. It is generally not recommended to take $\tau_n = 0$ as it might lead to an empty estimated set when the true identified set is small. This is similar in spirit to the GMM point-identified case. A positive value of the criterion is admissible for the estimated point.

moment selection (GMS)—for calculating the critical value in an effective manner. This method is derived from the observation that the asymptotic distribution of $Q_n(\theta)$ depends only on the moments that are binding. However, because of sampling, the identity of the binding moments is unknown. One solution consists of considering that they are all binding. This solution controls for size but is too conservative when only a few moments are binding because it increases the critical value.

In contrast, GMS is a data-driven selection of which moments matter and is based on the distance of the empirical moments to zero. Andrews & Soares (2010) construct a critical value c_n associated with such a procedure and compare its performance with different resampling techniques. First, the naive bootstrap does not work (Andrews & Guggenberger 2009). Second, GMS procedures have better finite distance properties than the subsampling techniques proposed by Chernozhukov et al. (2007) or Romano & Shaikh (2010). Sophisticated bootstraps are also available in the work of Bugni (2010) and Henry et al. (2015).

Another way to get better finite distance behavior is to redefine the criterion $Q(\theta)$. Andrews & Barwick (2012) compare different criteria. First, criterion $Q(\theta)$ in Equation 12 is of a Cramér–von Mises type, as it sums the squared positive deviations from zero. Alternatively, a Kolmogorov–Smirnov (KS)-type statistic constructed as the maximum of those deviations could be retained. Andrews & Barwick (2012) illustrate the better performance of the KS-type statistic using simulations. Also, weighting each moment condition by the inverse of its variance is also recommended by the authors, as in a GMM approach using moment equality restrictions.

Moreover, Andrews & Barwick (2012) refine the selection procedure of Andrews & Soares (2010) to ensure better finite sample behavior. Moments are selected using a more flexible criterion that does not vary with the number of observations while still correcting the size of the test. Romano et al. (2014) offer a simplification of this very computationally intensive method, particularly when the number of moments is large, at the price of a possibly conservative procedure and, therefore, a slight loss of power.

4.2.3. Infinitely many moment inequalities. The most recent literature extends this topic to an infinite number of moment inequalities. Andrews & Shi (2013) are specifically interested in the transformation of a finite number of conditional moment inequalities into unconditional moment inequalities, the number of which grows with the sample size. This is also the case in the work of Lee et al. (2014), who propose a test of functional inequalities or conditional moments.

Armstrong (2014, 2015) also considers conditional moment inequalities and analyzes simultaneously the optimality of the chosen statistic and that of the chosen instruments that are used to transform conditional moment inequalities into unconditional ones. He proves that, as in the work of Andrews & Barwick (2012), a KS statistic is more powerful than a Cramér–von Mises one. Additionally, kernel-based instruments outperform bounded ones in terms of rates of convergence, and Armstrong proposes a method for selecting the optimal bandwidth.

Other authors, such as Menzel (2014) and Ponomareva (2010), study the case of many moments and the way in which they should be selected and used. In particular, Chernozhukov et al. (2016) use large deviation theory to provide simple yet reasonably efficient critical values for testing many moment inequalities.

4.3. Estimation and Inference of Convex Sets

In the examples in Section 3, convex analysis enables inference from a different perspective, as in the work of Beresteanu & Molinari (2008). In regular cases, what makes this approach attractive is that it avoids computationally costly resampling procedures because the distribution of the test statistic is standard. Specifically, an estimator of the support function in each direction q, as developed in Section 3.3, can be expressed as an ordinary least squares (OLS) estimator.

Namely, the expression of a point on the boundary of the identified set, the supporting hyperplane of which is perpendicular to the direction q (see **Figure 1**), is written

$$\beta_q = (E(\mathbf{x}^\top \mathbf{x}))^{-1} E(\mathbf{x}^\top (\mathbf{1}\{\mathbf{x}_q > 0\}\mathbf{d} + \mathbf{y}_L)).$$

Its estimate, $\hat{\beta}_q$, is obtained by OLS when the dependent variable is constructed as

$$\mathbf{1}\{x_{n,qi} > 0\}d_i + y_{Li}, \text{ with } x_{n,qi} = q^\top \left(\frac{1}{n}\sum_{i=1}^{n} x_i^\top x_i\right)^{-1} x_i^\top$$

and the covariates are x_i. The estimate of the support function is then derived as $\hat{\delta}_n^*(q;\Theta_I) = q^\top \hat{\beta}_q$. Inference methods are developed by Beresteanu & Molinari (2008) and Bontemps et al. (2012). These methods exploit the convex structure of the identified set, and under certain technical conditions, inference is efficient (Kaido & Santos 2014).

As discussed in Section 3.3, the method is based on the fact that the process on the unit sphere

$$T(q) = q^\top \theta_0 - \delta^*(q;\Theta_I)$$

is always nonpositive when the point tested, θ_0, belongs to the identified set. In practice, $T(q)$ is estimated by its empirical counterpart $T_n(q)$, derived by plugging in the previous estimate of the support function. Chernozhukov et al. (2015) interpret this test statistic as a Wald statistic because it measures the distance between θ_0 and Θ_I. Note that this test statistic can be studentized because the variance of $\hat{\delta}_n^*(q)$ has a closed form.

Bontemps et al. (2012) prove that the identified set is smooth and strictly convex and therefore has no exposed faces and no kinks when covariates x are continuously distributed and have a probability density function (PDF) positive everywhere. In this case, $\sqrt{n}(T_n(q) - T(q))$ tends uniformly in distribution, as n approaches infinity, to a Gaussian stochastic process, and the argument of its maximum is asymptotically unique. This is the direction q for which $\beta_q = \theta_0$. Note that we can interpret the search for a maximizer of $T(q)$ as a moment selection procedure that fully exploits the geometry of the set. A direct application of the moment inequality literature to this issue would lead to the selection of too many moments around the true one and would cause efficiency losses.

Furthermore, the test statistic $\max_q \sqrt{n}T_n(q)$ is asymptotically normally distributed, and a plug-in estimate of the variance is proposed by Bontemps et al. (2012) based on OLS residuals.[11]

When the PDF of covariates $\mathbf{x}$ is not strictly positive, the identified set might have a kink at the tested point, θ_0. The argument of the maximum of $T(q)$ is no longer unique and belongs to a nontrivial cone. The asymptotic distribution of $\max_q \sqrt{n}T_n(q)$ is no longer standard. This result is similar to what is found in the maximum likelihood (ML) literature (see Redner 1981). What is key is that the ML estimator is valid even under loss of identification (because, even if the argument of the maximum, $\theta_{\max}$, is not unique, the likelihood function of $\theta_{\max}$ is unique), although the implied likelihood-ratio (LR) test statistic is no longer a χ^2 distribution. Liu & Shao (2003) derive the asymptotic distribution of the test in this case. Conceptually, this is akin to a moment inequality setup when we do not know the identity of binding moments, although we deal in this case with a connected continuum of such inequalities.

[11] Beresteanu & Molinari (2008) also propose estimates of the covariance operator of the support function estimator, but they are slightly more complicated to compute than with Bontemps et al.'s (2012) method.

Bontemps et al. (2012) propose to add a perturbation in $T_n(q)$ that makes the limit of the sequence of the argument unique and the asymptotic distribution of the statistic standard. Alternatively, Chandrasekhar et al. (2012) propose to smooth the variable of interest by adding a small continuous noise, the support of which is infinite, to recover a smooth and convex identified set and therefore a unique maximizer.

This latter method is also helpful when $\mathbf{x}$ is composed of discrete random variables because, in this case, the identified set has exposed faces not reduced to singletons and the empirical process is no longer asymptotically normal.[12]

Finally, the estimates of the support function of the convex set $\mathbb{E}\mathbf{M}(\theta_0)$ used in the indirect method developed in Section 3.3 are asymptotically normally distributed even when variables are discrete. All the technology developed by Beresteanu & Molinari (2008) or Bontemps et al. (2012) can be brought in for these models, in which the identified set may be nonconvex but for which the indirect approach works.

4.4. Intersection of Bounds

Section 3.4 develops an example in which the slope parameter β_1 is bounded by an infinite number of moments. From Equation 8, we have that, for $q_\beta = 1$,

$$\beta_1 \leq \inf_{q_\gamma \in \mathbb{R}} E\left(\left(E(\mathbf{x}^2)^{-1}\mathbf{x} + q_\gamma E(\mathbf{z}^2)^{-1}\mathbf{z}\right)\left(\mathbf{y}_L + \mathbf{x}\mathbf{1}\{E(\mathbf{x}^2)^{-1}\mathbf{x} + q_\gamma E(\mathbf{z}^2)^{-1}\mathbf{z} > 0\}\mathbf{d}\right)\right).$$

In other examples, bounds result from independence restrictions (e.g., Manski & Pepper 2000) and satisfy conditions as

$$\theta \leq \inf_z \left[E(h(\mathbf{y},\mathbf{x},\mathbf{z}) \mid \mathbf{z} = z)\right].$$

If we designate $h_n(z)$ an estimator for a sample of size n, for example, a nonparametric estimator of $E(h(\mathbf{y},\mathbf{x},\mathbf{z}) \mid \mathbf{z} = z)$, the estimation of this bound by the quantity

$$\inf_z \left(h_n(z)\right) \qquad 13.$$

is severely biased downwards in small samples because sampling variability, and specifically the variation of the variance of $h_n(z)$ as a function of z, is not controlled. The argument of the infimum of the estimated function in Equation 13 has a strong tendency to be a point z at which the estimate is very noisy.

Chernozhukov et al. (2013) propose to solve this inference problem by using the estimator

$$\inf_z \left[h_n(z) + c_n v_n(z)\right],$$

where $v_n(z)$ is an estimator of the variance of the empirical counterpart $h_n(z)$ at point z. The addition of this term to the objective function penalizes regions in which the conditional variances of the objective function are large. Again, the difficulty is the calculation of the critical value c_n.

Observe that, even if the two examples above appear identical, the first exhibits more regularity than the second. The role played by function h in the first example is the support function of the unrestricted set Θ_I^U, the variance of which has a closed form that can be exploited (see Bontemps et al. 2012). More importantly, the control variable q_γ is not a random variable, unlike $\mathbf{z}$. The

[12] In the direct approach, this is due to the premultiplication by matrix $E[x^\top x]^{-1}$ that, because this matrix is estimated, introduces sampling variability in the directions orthogonal to the exposed faces. This is also why Kaido & Santos (2014) assume that, if the convex set has exposed faces, then these directions are known.

calibration of c_n is therefore much easier to handle in the first example. This remark applies to any convex set that is identified using the two-step approach developed in Section 3.4.

4.5. Inference for Subvectors

In many cases, empirical researchers are only interested in a subvector of parameters or in specific functionals of parameters. One solution to the problem of narrowing down the parameters to a subvector consists of projecting confidence sets on the dimensions of interest, although it is likely to be (very) conservative. It is worth noting that similar issues arise in the weak instrumental variable literature, where Anderson-Rubin-type statistics are used to test whether some given values are admissible (see, for example, Guggenberger et al. 2012 for an improvement of the projection method). Romano & Shaikh (2008) and Bugni et al. (2017) propose another approach, where the statistic of interest is concentrated out in the dimensions that are of no interest. Again, the resulting test statistics are not standard. Romano & Shaikh (2008) propose computation of the critical values by subsampling techniques, whereas Bugni et al. (2017) propose a bootstrap approach. Furthermore, Kaido et al. (2016) exploit a local linear approximation of the moment inequalities to provide an alternative method for computing the critical values.

It is worth noting that inference for subvectors or linear functions of the full vector is straightforward when the identified set is convex. Appropriately choosing one or several directions is enough for inference on the corresponding subvectors. Inference on, say, the first component of parameter θ using the direct approach requires choosing $q = (1, 0, \ldots, 0)^\top$ and $q = (-1, 0, \ldots, 0)^\top$ as the directions of interest and studying the behavior of the support function in these directions only.

4.6. Bayesian Estimation

Several authors have developed Bayesian methods for set identification. Liao & Jiang (2010) work in a standard setting of moment inequalities. An interesting aspect of their study is that the slackness of each moment inequality is assumed to be an auxiliary parameter, and some prior distribution is used for them as well as for the structural parameters. The posterior densities for the latter are obtained by integrating out the former. The authors also develop methods for moment and model selection in order to select the most parsimonious and precise model.

Moon & Schorfheide (2012) work in a setting in which some reduced-form parameters are point identified and in which partial identification is generated by the relationship between this parameter and the structural form parameters that generates partial identification. An entry game provides such an example because probabilities of simultaneous actions by agents are point identified. The posterior distribution function of structural parameters is derived from posteriors of the reduced-form parameters. The authors' main finding is that the Bernstein–von Mises theorem does not hold. Bayesian credible regions, covering a true parameter and defined by the highest posterior density, do not coincide with the corresponding frequentist confidence region and in fact, under appropriate conditions, are strictly contained in this frequentist region.

Kitagawa (2012) solves this problem by introducing a more general class of priors and using an inferior envelope of the posteriors to reconcile the Bayesian and frequentist approaches, at least asymptotically. The partial prior knowledge is modeled as a class and distinguished by whether the priors are revisable by the data. Indeed, the lack of point identification is associated with flat regions of the likelihood function, and this translates into the absence of revision of priors in this region. Usual priors are considered for identified parameters, whereas all possible priors are considered for unidentified parameters. The author then uses a posterior gamma minimax which minimizes the worst-case posterior risk over the class of all posteriors generated by this general class of priors. Another way of solving this problem is developed by Kline & Tamer (2016). They show that there is an asymptotic equivalence between Bayesian and frequentist analyses when the inference concerns the identified set rather than the partially identified parameter.

Liao & Simoni (2016) consider the estimation of closed and convex sets in a similar setting to that of Moon & Schorfheide (2012). Under some conditions, they derive a uniformly linear approximation of the support function as a function of reduced-form parameters. This result allows them to prove an analog to the Bernstein–von Mises theorem for the support function. Bayesian credible sets coincide asymptotically with frequentist regions. The intuitive reason for which the Bernstein–von Mises theorem holds is that the last three studies mentioned focus on the posterior distribution of the identified set and not on the specific value of the partially identified parameter.

5. A SAMPLE OF EMPIRICAL APPLICATIONS AND RELATED TOPICS

Even though the number of empirical applications in the literature is steadily increasing, most papers do not use these recent inference methods that we have reviewed above. A notable exception is the use of these methods for the estimation of treatment models and, more generally, the estimation of reduced-form models with selectivity. Because of the low dimensionality of the random variable that completes the model, bounds can be easily characterized and efficiently estimated. By contrast, the empirical literature on set-identified structural models has not yet reached a mature level. These models are often estimated by using moment inequalities that exploit players' rational behavior, combined with equilibrium constraints. The issue of sharpness is set apart in order to alleviate estimation costs. In addition, convexity, though promising, has not yet been fully exploited. In this section, we briefly present these streams of empirical research (for a more complete and elaborate review of empirical applications, see Ho & Rosen 2015).

Among the first authors to use partial identification concepts in an empirical framework are Hotz et al. (1997). They use a reduced-form model that sets questions of treatment evaluation in a setting where the main instrumental variable does not fully respect the usual conditions for its validity. The parameter of interest is the causal effect of early pregnancy—the age during adolescence at which the first child was born—on subsequent behavior and outcomes. The instrumental variable in question is the occurrence of a miscarriage during pregnancy. Miscarriages do indeed provide a valid instrument, but only for a subsample of the population, and are therefore contaminated in the sense of Horowitz & Manski (1995). The literature on treatment and selection also includes the work of Manski & Pepper (2000), who analyze the returns to education at all its levels (these levels are considered as multiple treatments). They use monotonicity assumptions on the effect of treatment or the existence of a variable that monotonically affects income. The same authors analyze the deterrent effects of the death penalty in the United States (Manski & Pepper 2013) and show that different assumptions lead to dramatically different conclusions.

Another example of reduced-form estimation in a model with selectivity is the work of Honoré & Lleras-Muney (2004). The authors estimate bounds on the evolution, over the past 40 years in the United States, of the two main causes of death: heart disease and cancer. These causes are treated as competing risks in a duration model, and the correlation between these risks is the parameter that is not point identifiable. The authors show that progress in the fight against cancer seems to have been hidden by the important progress against heart disease in analyses that assume independent competing risks.

The evaluation of public policies such as internships offered to certain populations takes center stage in the recent literature in applied econometrics, and some authors have used bounds. For example, Lee (2009) shows how to overcome the problems of selection in employment to assess the effects of a training program, the Job Corps in the United States. Lee uses controlled experimental data and an assumption of monotonicity of the treatment effect on employment to infer the effects of the treatment on wages conditional on employment. The framework proposed by Manski for dealing with selection issues is also applied by Blundell et al. (2007) in the case of changes in the returns to education for men and for women in the United Kingdom over the past 30 years while dealing with nonparticipation. The empirical literature in those cases of treatment and selection is quite well developed, and other references could have been given.

Most examples of set-identified structural models are borrowed from empirical industrial organization. Entry games have been used as a case study in the theoretical literature. They provide an example of a simultaneous equation model with discrete endogenous outcomes, i.e., the decisions of firms to enter or not enter into a collection of independent markets (see Berry & Reiss 2007 for a survey). An entry game may be set identified because of the existence of multiple equilibria that we do not know how to select.[13] Ciliberto & Tamer (2009) use US data and the method of Chernozhukov et al. (2007) to estimate parameters of a linear profit function in an entry game played by airlines on routes connecting two airports. They do not sharply characterize the identified set because of its complexity in the many-player case. Grieco (2014) generalizes the informational structure of Ciliberto & Tamer (2009) and allows for both complete and incomplete information. He estimates the impact of supercenters on competition in rural grocery markets.

Several contributions have also been developed in the literature on auctions. One of the first examples is presented by Haile & Tamer (2003). The authors develop a structural model for ascending auctions, for which parameters are notoriously difficult to identify because of poor observed information. The authors only exploit rationality constraints on agents' behavior and do not make any assumption on the distribution of bidders' private values. They assume that potential buyers bid up to the value that they give to the object and do not let the item be sold at a price lower than this value. In a more recent study, Chesher & Rosen (2015b) develop methods to derive sharp identification in this model. Komarova (2013) relaxes the assumption of independent private value in second-price and ascending auctions and exploits rationality constraints, as do Haile & Tamer (2003). Armstrong (2013) derives bounds in the presence of unobserved heterogeneity, and Gentry & Li (2014) consider entry costs in auctions and derive bounds for the distributions of interest.

Many studies exploit rationality constraints in games and situations. Examples are provided by Pakes (2010) and Pakes et al. (2015). Pakes et al. (2015) develop the estimation of structural models under general rationality constraints upon ordered choices (such as the number of bank ATMs) or in noncooperative games between hospitals and HMOs. Inequality constraints on the parameters of interest governing profit functions of firms are derived from the restriction that firm choices should bring them profits that are higher than they would have earned had they made other decisions.

Finally, a few applications of structural models in other subfields use set identification. Specifically, Blundell et al. (2008) exploit revealed preferences and smooth Engel curves to bound demand functions in a case where the distribution of prices is discrete. This prevents point identification of price elasticities. This work is extended by Blundell et al. (2014). Henry & Mourifié (2013) study political competition and the spatial voting model and show how to test this model despite partial identification. The authors reject the model using US data. Recent work on networks exploits pairwise stability to estimate models of network formation. De Paula et al. (2015) adapt the solution of Ciliberto & Tamer (2009) to this problem, whereas Sheng (2014), due to the curse of dimensionality because of the (exponentially) increasing number of

[13] Having regions of multiple equilibria does not preclude having point identification, as shown by Tamer (2003).

moment inequalities generated, only considers pairwise stability in subnetworks (see de Paula 2016 for a complete review of the econometrics of network models).

6. CONCLUSION

In general, we can describe the empirical strategy of an applied econometrician as a choice of implicit or explicit structural assumptions that are used in the analysis of data to estimate economic parameters. The traditional approach seeks to complete this list of assumptions so that only a single parameter value could be the result of this approach. For example, using censored data, we can readily identify parameters of interest by assuming normality of errors and using ordered probit, as in Example 1 of Section 2.2. The concept of partial identification allows us to abandon this ad hoc completion at the cost of admitting that credible structures are loose enough to lead to the identification of a set of parameter values only. Instead of a normality assumption, we could use other assumptions such as independence, mean independence, or the absence of correlation with respect to covariates. Despite this extension of the concept of identification, reporting inference results through confidence regions is conducted in a similar way to the point-identified case, and the usual empirical reasoning of applied econometricians remains the same.

Note, however, that this approach seems to go in the opposite direction to the one Popper (2005) would have recommended. Popper (2005) suggests that the quality of a theory is to make sufficiently restrictive assumptions that are easy to falsify or to reject. The partial identification approach instead seems to develop a protective belt against any rejection by weakening the restrictions that are made. Easing restrictions in an unbridled way gives rise to a phenomenon of regression to infinity that is slightly discouraging because, with no restrictions, we cannot identify anything. Weak assumptions also lead to the risk of having imprecise policy recommendations at the cost of a strength that might seem extreme in other scientific fields.

This is why this approach should be interpreted otherwise. A natural direction suggested by Manski is to compare hypotheses that are increasingly binding and that reduce the size of the identified set [e.g., the empirical strategy used by Manski & Pepper (2000)]. The data will not be what justifies the credibility of research results because the data remain the same. This is the set of assumptions that researchers must justify. If the approach is open enough that readers can evaluate the credibility of stronger and stronger restrictions, readers will have the option of conducting empirical reasoning that is rich enough to determine that this assumption leads to an empirical conclusion or even to the absence of an empirical conclusion. Indeed, the bounds of identified intervals or regions may be large under weak hypotheses. This lack of conclusion should then motivate the search for new credible assumptions or the collection of new data, which would strengthen the credibility of empirical approaches in economics.

Despite a blossoming number of theoretical papers during the past 15 years, there are still too few empirical applications. Empirical researchers are reluctant to use techniques that are nonstandard and computationally challenging even though program codes are now available in standard softwares: Stata for the methods of Beresteanu & Molinari (2008) and Chernozhukov et al. (2013) and Matlab and Stata for the GMS procedure of Andrews & Shi (2013).

In general moment inequality settings, inference is conducted by inverting a test. There are two dimensionality issues with this method. First, the dimension of the parameter space increases with the number of explanatory variables. Second, in most structural models, the number of moment inequalities that characterize the sharp identified set exponentially increases with the number of control variables, as well as with other dimensions such as the number of players in games or networks.

Test inversion might seem costly to applied researchers because, for each point on a thin grid in the parameter space, a test statistic using very many moment inequalities has to be constructed and compared to a critical value that is specific to the point tested. This practical issue clearly attenuates the attractiveness of sharp identification and of efficient inference methods. One challenge for the near future is to facilitate the handling of inference techniques for reasonably large-dimensional parameter spaces and a large number of conditional moment inequalities that generate many moment inequalities.

The geometry of the identified set, and specifically its convexity, could be exploited more systematically, and the resulting simplifications in terms of the number of relevant moment inequalities are attractive. Convexity reduces the curse of dimensionality by replacing a large number of moment inequalities with the analysis of a process on the unit sphere. Many models, such as regressions with interval censoring, selection models, sample combination, or entry games, can be transformed into convex problems. This is not always easy, however, and certainly requires ingenuity on the part of the researcher.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Rohit Kumar, Steve Lawford, Nour Meddahi, David Pacini, and Anna Simoni for their comments. We gratefully acknowledge financial support from the European Research Council under the European Community's Seventh Framework Program FP7/2007-2013 grant agreement 295298 and from the Agence Nationale de Recherche grant ANR-05-BLAN-0024-01. The authors are solely responsible for remaining errors.

LITERATURE CITED

Andrews DWK, Barwick PJ. 2012. Inference for parameters defined by moment inequalities: a recommended moment selection procedure. *Econometrica* 80:2805–26

Andrews DWK, Guggenberger P. 2009. Validity of subsampling and plug-in asymptotic inference for parameters defined by moment inequalities. *Econom. Theory* 25:669–709

Andrews DWK, Shi X. 2013. Inference based on conditional moment inequalities. *Econometrica* 81:609–66

Andrews DWK, Soares G. 2010. Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* 78:119–57

Armstrong TB. 2013. Bounds in auctions with unobserved heterogeneity. *Quant. Econ.* 4:377–415

Armstrong TB. 2014. Weighted KS statistics for inference on conditional moment inequalities. *J. Econom.* 181:92–116

Armstrong TB. 2015. Asymptotically exact inference in conditional moment inequality models. *J. Econom.* 186:51–65

Beresteanu A, Molchanov I, Molinari F. 2011. Sharp identification regions in models with convex moment predictions. *Econometrica* 79:1785–821

Beresteanu A, Molchanov I, Molinari F. 2012. Partial identification using random set theory. *J. Econom.* 166:17–32

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

Berry S, Reiss P. 2007. Empirical models of entry and market structure. In *Handbook of Industrial Organization*, Vol. 3, ed. M Armstrong, R Porter, pp. 1845–86. Amsterdam: Elsevier

Blundell R, Browning M, Crawford I. 2008. Best nonparametric bounds on demand responses. *Econometrica* 76:1227–62

Blundell R, Gosling A, Ichimura H, Meghir C. 2007. Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica* 75:323–63

Blundell R, Kristensen D, Matzkin R. 2014. Bounding quantile demand functions using revealed preference inequalities. *J. Econom.* 117:112–27

Bontemps C, Magnac T, Maurin E. 2012. Set identified linear models. *Econometrica* 80:1129–55

Bugni F. 2010. Bootstrap inference in partially identified models defined by moment inequalities. *Econometrica* 78:735–54

Bugni FA, Canay IA, Shi X. 2017. Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quant. Econ.* 8:1–38

Canay I. 2010. EL inference for partially identified models: large deviations optimality and bootstrap validity. *J. Econom.* 156:408–25

Chandrasekhar A, Chernozhukov V, Molinari F, Schrimpf P. 2012. *Inference for best linear approximations to set identified functions*. Work. Pap. 43/12, Cent. Microdata Methods Pract., London

Chernozhukov V, Chetverikov D, Kato K. 2016. *Testing many moment inequalities*. Work. Pap. 42/16, Cent. Microdata Methods Pract., London

Chernozhukov V, Hong H, Tamer E. 2007. Inference on parameter sets in econometric models. *Econometrica* 75:1243–84

Chernozhukov V, Kocatulum E, Menzel K. 2015. Inferences on sets in finance. *Quant. Econ.* 6:309–58

Chernozhukov V, Lee S, Rosen AM. 2013. Intersection bounds: estimation and inference. *Econometrica* 81:667–737

Chesher A. 2005. Nonparametric identification under discrete variation. *Econometrica* 73:1525–50

Chesher A. 2010. Instrumental variable models for discrete outcomes. *Econometrica* 78:575–601

Chesher A, Rosen A. 2015a. *Characterizations of identified sets delivered by structural econometric models*. Work. Pap. 63/15, Cent. Microdata Methods Pract., London

Chesher A, Rosen A. 2015b. *Identification of the distribution of valuations in an incomplete model of English auctions*. Work. Pap. 30/15, Cent. Microdata Methods Pract., London

Chesher A, Smolinski K. 2012. IV models of ordered choice. *J. Econom.* 166(1):33–48

Ciliberto F, Tamer E. 2009. Market structure and multiple equilibria in airline markets. *Econometrica* 77:1791–828

Conley T, Hansen C, Rossi P. 2012. Plausibly exogenous. *Rev. Econ. Stat.* 94:260–72

Davezies L, d'Haultfoeuille X. 2012. *Partial identification with missing outcomes*. Unpublished manuscript, Cent. Rech. Écon. Stat., Inst. Natl. Stat. Études Écon., Paris

de Paula A. 2016. *Econometrics of network models*. Work. Pap. 06/16, Cent. Microdata Methods Pract., London

de Paula A, Richards-Shubik S, Tamer E. 2015. *Identification of preferences in network formation games*. Work. Pap. 29/15, Cent. Microdata Methods Pract., London

Fan Y, Park S. 2010. Sharp bounds on the distribution of the treatment effects and their statistical inference. *Econom. Theory* 26:931–51

Fan Y, Wu J. 2010. Partial identification of the distribution of treatment effects in switching regimes models and its confidence sets. *Rev. Econ. Stud.* 77:1002–41

Fréchet M. 1951. Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon III Sér. Sci. A* 9:53–77

Frisch R. 1934. *Statistical Confluence Analysis by Means of Complete Regression Systems*. Oslo, Nor.: Univ. Inst. Econ.

Galichon A, Henry M. 2009. A test of non-identifying restrictions and confidence regions for partially identified parameters. *J. Econom.* 152:186–96

Galichon A, Henry M. 2011. Set identification in models with multiple equilibria. *Rev. Econ. Stud.* 78:1264–98

Gentry M, Li T. 2014. Identification in auctions with selective entry. *Econometrica* 82:315–44

Gini C. 1921. Sull'interpoliazone di una retta quando i valori della variabile indipendente sono affetti da errori accidentali. *Metroeconomica* 1:63–82

Gomez F, Pacini D. 2013. *Measuring inequality from contaminated data*. Unpublished manuscript, Univ. Bristol, Bristol, UK

Grieco PLE. 2014. Discrete games with flexible information structures: an application to local grocery markets. *RAND J. Econ.* 45:303–40

Guggenberger P, Kleibergen F, Mavroeidis S, Chen L. 2012. On the asymptotic sizes of subset Anderson–Rubin and Lagrange multiplier tests in linear instrumental variables regression. *Econometrica* 80:2649–66
Haile P, Tamer E. 2003. Inference with an incomplete model of English auctions. *J. Polit. Econ.* 111:1–51
Heckman JJ. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46:931–60
Henry M, Meango R, Queyranne M. 2015. Combinatorial approach to inference in partially identified incomplete structural models. *Quant. Econ.* 6:499–529
Henry M, Mourifié I. 2013. Euclidean revealed preferences: testing the spatial voting model. *J. Appl. Econom.* 28(4):650–66
Henry M, Onatski A. 2012. Set coverage and robust policy. *Econ. Lett.* 115(2):256–57
Ho K, Rosen A. 2015. *Partial identification in applied research: benefits and challenges*. Work. Pap. 64/15, Cent. Microdata Methods Pract., London
Hoeffding W. 1940. Masstabinvariante Korrelationstheorie. *Schr. Math. Inst. Inst. Angew. Math. Univ. Berlin* 5(3):179–233
Honoré B, Lleras-Muney A. 2004. Bounds in competing risks models and the war on cancer. *Econometrica* 74:1675–98
Horowitz JL, Manski CF. 1995. Identification and robustness with contaminated and corrupted data. *Econometrica* 63:281–302
Horowitz JL, Manski CF. 2006. Identification and estimation of statistical functions using incomplete data. *J. Econom.* 132:445–59
Horowitz JL, Manski CF, Ponomareva M, Stoye J. 2003. Computation of bounds on population parameters when the data are incomplete. *Reliab. Comput.* 9:419–40
Hotz VJ, Mullin C, Sanders SG. 1997. Bounding causal effects using data from a contaminated natural experiment: analysing the effects of teenage childbearing. *Rev. Econ. Stud.* 64(4):575–603
Imbens G, Manski CF. 2004. Confidence intervals for partially identified parameters. *Econometrica* 72:1845–59
Kaido H, Molinari F, Stoye J. 2016. Confidence intervals for projections of partially identified parameters. arXiv:1601.0094 [math.ST]
Kaido H, Santos A. 2014. Asymptotically efficient estimation of models defined by convex moment inequalities. *Econometrica* 82(1):387–413
Kitagawa T. 2012. *Estimation and inference for set-identified parameters using posterior lower probability*. Work. Pap., Cent. Microdata Methods Pract., London
Klepper S, Leamer EE. 1984. Consistent sets of estimates for regressions with errors in all variables. *Econometrica* 52:163–83
Kline B, Tamer E. 2016. Bayesian inference in a class of partially identified models. *Quant. Econ.* 7:329–66
Komarova T. 2013. Partial identification in asymmetric auctions in the absence of independence. *Econom. J.* 16(1):60–92
Leamer EE. 1987. Errors in variables in linear systems. *Econometrica* 55(4):893–909
Lee DS. 2009. Training, wages, and sample selection: estimating sharp bounds on treatment effects. *Rev. Econ. Stud.* 76(3):1071–102
Lee S, Song K, Whang Y. 2014. *Testing for a general class of functional inequalities*. Work. Pap. 09/14, Cent. Microdata Methods Pract., London
Liao Y, Jiang W. 2010. Bayesian analysis in moment inequality models. *Ann. Stat.* 38(1):275–316
Liao Y, Simoni A. 2016. *Bayesian inference for partially identified convex models: Is it valid for frequentist inference?* Work. Pap. 2016-07, Rutgers Univ., New Brunswick, NJ
Liu X, Shao Y. 2003. Asymptotics for likelihood ratio tests under loss of identifiability. *Ann. Stat.* 31:807–32
Magnac T, Maurin E. 2008. Partial identification in monotone binary models: discrete regressors and interval data. *Rev. Econ. Stud.* 75:835–64
Manski CF. 1989. Anatomy of the selection problem. *J. Hum. Resour.* 24:343–60
Manski CF. 2003. *Partial Identification of Probability Distributions*. Berlin: Springer-Verlag
Manski CF, Pepper JV. 2000. Monotone instrumental variables: with an application to the returns to schooling. *Econometrica* 68(4):997–1010
Manski CF, Pepper JV. 2013. Deterrence and the death penalty: partial identification analysis using repeated cross sections. *J. Quant. Criminol.* 29(1):123–41

Manski CF, Tamer E. 2002. Inference on regressions with interval data on a regressor or outcome. *Econometrica* 70:519–46
Marschak J, Andrews WH. 1944. Random simultaneous equations and the theory of production. *Econometrica* 12:143–205
Menzel K. 2014. Consistent estimation with many moment inequalities. *J. Econom.* 182(2):329–50
Molchanov I. 2005. *Theory of Random Sets*. Berlin: Springer-Verlag
Molchanov I, Molinari F. 2015. Applications of random set theory in econometrics. *Annu. Rev. Econ.* 6(1):229–51
Moon HR, Schorfheide F. 2012. Bayesian and frequentist inference in partially identified models. *Econometrica* 80(2):755–82
Nevo A, Rosen A. 2012. Identification with imperfect instruments. *Rev. Econ. Stat.* 94(3):659–71
Pacini D. 2017. Two-sample least squares projection. *Econom. Rev.* In press. **https://doi.org/10.1080/07474938.2016.1222068**
Pakes A. 2010. Alternative models for moment inequalities. *Econometrica* 78:1783–822
Pakes A, Porter J, Ho K, Ishii J. 2015. Moment inequalities and their applications. *Econometrica* 83(1):315–34
Ponomareva M. 2010. *Inference in models defined by conditional moment inequalities with continuous covariates*. Unpublished manuscript, Dep. Econ., Univ. West. Ont., London, Can.
Ponomareva M, Tamer E. 2011. Misspecification in moment inequality models: back to moment equalities. *Econom. J.* 14:186–203
Popper K. 2005. *The Logic of Scientific Discovery*. London/New York: Routledge/Taylor & Francis
Redner RA. 1981. On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Stat.* 9:225–28
Reiersol O. 1941. Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica* 9:1–24
Ridder GE, Moffitt R. 2007. The econometrics of data combination. In *Handbook of Econometrics*, Vol. 6, ed. JJ Heckman, EE Leamer, pp. 5469–547. Amsterdam: North-Holland
Rockafellar RT. 1970. *Convex Analysis*. Princeton, NJ: Princeton Univ. Press
Romano JP, Shaikh AM. 2008. Inference for identifiable parameters in partially identified econometric models. *J. Stat. Plan. Inference* 138:2786–807
Romano JP, Shaikh AM. 2010. Inference for the identified set in partially identified econometric models. *Econometrica* 78(1):169–211
Romano JP, Shaikh AM, Wolf M. 2014. A practical two-step method for testing moment inequalities. *Econometrica* 82(5):1979–2002
Rosen AM. 2008. Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *J. Econom.* 146:107–17
Rothenberg TJ. 1971. Identification in parametric models. *Econometrica* 39:577–91
Sheng S. 2014. *A structural econometric analysis of network formation games*. Work. Pap., Dep. Econ., Univ. Calif., Los Angeles
Silvapulle M, Sen P. 2005. *Constrained Statistical Inference*. London: Wiley
Stoye J. 2007. Bounds on generalized linear predictors with incomplete outcome data. *Reliab. Comput.* 13:293–302
Stoye J. 2009. More on confidence intervals for partially identified parameters. *Econometrica* 77:1299–315
Tamer E. 2003. Incomplete simultaneous discrete response model with multiple equilibria. *Rev. Econ. Stud.* 70(1):147–65
Tamer E. 2010. Partial identification in econometrics. *Annu. Rev. Econ.* 2:167–95
Wooldridge JM. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press