WORKING PAPERS

N° 16-752

January 2017

"Set Identification, Moment Restrictions and Inference"

Christian Bontemps and Thierry Magnac

Toulouse School of Economics

Set Identification, Moment Restrictions and Inference

Christian Bontemps* Thierry Magnac†

First Version: September 2016

This Version: January 12, 2017

Abstract

For the last ten 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 in characterizing the identified set but also for inference. In this perspective, we review inference methods using convex analysis or inversion of tests and detail how geometric characterizations can be useful.

Keywords: set identification, moment inequality, convex set, support function.

*Toulouse School of Economics, ENAC, Université de Toulouse, 21 allée de Brienne, 31000 Toulouse, France, email: christian.bontemps@tse-fr.eu

†Toulouse School of Economics, Université de Toulouse Capitole, 21 allée de Brienne, 31000 Toulouse, France, email: thierry.magnac@tse-fr.eu

1 Introduction[1]

The importance of the standard notion of point identification, that appears in standard econometric textbooks (for instance, Wooldridge, 2010), has been questioned for the last thirty years especially by Manski and his coauthors (1989 and seq) who reintroduced and developed 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 was developed by Gini (1921) and Frisch (1934) for the simple regression model with measurement errors, Reiersol (1941), Marschak and Andrews (1944) for simultaneous equation models, Hoeffding (1943) and Fréchet (1955) for bounds on the joint distributions of variables when only marginal distributions are observed (in two different surveys, for example) and Klepper and 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 that he himself summarized in a book (Manski, 2003). Many of Manski's students helped to develop this literature (see in particular the surveys by Tamer, 2010, and Molchanov and Molinari, 2014).

The general reasoning that leads to partial identification is the notion of 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 since individuals can never be observed simultaneously in treatment and off-treatment. 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

[1]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 N° 295298 and from the Agence Nationale de Recherche ANR-05-BLAN-0024-01. The usual disclaimer applies.

equilibrium selection mechanism in a game). However, choosing a single completion of the model is arbitrary, and point identification becomes implausible.

However, this approach provides the first insight of 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 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, concavity). The identifying power of different assumptions may be compared in terms of the size of the set which is identified.

As presented, 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 by point estimates of the boundary of the set. In inference, the presentation using confidence regions does not change since 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 below 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 identification, estimation and inference. The point 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 since the space of convex sets, by being homeomorphic to their support functions (as we shall define below), have 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 to construct standard test statistics. This is one of the threads that we follow in this survey by devoting Section 3 to what we call convex set identification. This prepares the ground for discussing its implications for inference.

In Section 5 we briefly review the empirical literature that, while 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 recipes. 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 lower case letters for elements of these sets, e.g. θ. We use lower-case 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 (which can be found in Molchanov and Molinari, 2015) or the theory of random sets (see Molchanov, 2005).

This section first defines concepts of observational equivalence and point identification and second the notions of complete and incomplete models as well as of sharp identification. Then, we make the link 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 Set-up and Definitions

We adopt a setting in which random variables, say $\mathbf{w}$, whose interrelationships 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 since the population probability distribution depends on the finite dimensional parameter of interest $\theta \in \Theta \subset \mathbb{R}^d$, whose true value is θ_0, and on other nuisance parameters η. These nuisance parameters, whose true values are η_0, can be as general as one wants (for instance, they can be distribution functions). They will be 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. 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 will see 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 and 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 $P_{\theta,\eta}$ which are all compatible with the population distribution function. In contrast, a model is said to be *complete* when its parameters are point identified.

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. Alternatively, 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 $\mathbf{w}$.[2]

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.

Set identification 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:

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

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

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 because 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. They are all written as restrictions on the parameter $\eta\in\mathcal{E}$, which can be analyzed according to their degree of credibility:

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

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 which completes the partially identified model.

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

2.2 Example 1: Interval censoring and best single-dimensional linear prediction

Manski and Tamer (2002) analyze 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.[3] 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, ask households or individuals the exact level of their income or assets and, if households do not want to answer for privacy reasons, ask the same question but in the form of intervals. Is your income between 0 and 500 dollars? Or between 500 and 1000 dollars? etc. 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 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 bounds (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:

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

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 3 intervals.[4] 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. It includes point identified parameters derived from considering all possible distribution functions η of the variable $\mathbf{t}$.

The analysis with nonparametric completion proceeds as in the general definition. First, identify

[3] 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 since the right end is of measure 0. This is the purpose of the assumption stating the non-atomicity of probability distribution functions introduced later on.

[4] 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).

parameter β_1 in every completed model. Second, consider the union of all point identified values. As

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

we can derive the value of parameter β_1 as:[5]

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

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

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

in which $\mathbf{1}\{\cdot\}$ 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,

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

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 non-negative 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 of 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, and Magnac and Maurin, 2008).

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

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

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

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

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

2.3 Discussion

The resulting set-up of moment restrictions obtained in the previous example extends to many partially identified economic models. They express the identifying restrictions on parameters as inequality constraints on expectations of linear or non-linear functions of variables and parameters and therefore lead to a finite or infinite number of moment inequalities. A more difficult question is sharp identification in which case 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 since the number of restrictions is smaller (see Ciliberto and Tamer, 2009, for such an empirical strategy).

These extensions require more sophisticated tools than those we used in the previous very simple example. In structural models, and in particular those derived from game theory, Galichon and 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 to find a solution to these issues. We will 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 and Manski (1995), Manski and Pepper (2000) or more generally all works as reviewed by Manski (2003). Many topics are connected with the framework of partial identification. Ridder and Moffitt (2007) offers 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 (2016) 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 and Smolinski (2012) and Chesher and Rosen (2015a) generalize instrumental variable models. Binary models with a

”very exogenous” regressor, whose observations are interval censored, are analyzed by Magnac and Maurin (2008). Davezies and 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 and Manski (2005), Fan and Park (2010), Fan and Wu (2010) and Gomez and Pacini (2012). Nevo and Rosen (2012) and Conley et al. (2011) 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 endogeneous one they are supposed to instrument.

3 Direct and indirect uses of convexity arguments

Some reminder of random set theory is useful to make our survey article self-contained and we start by borrowing notation from the survey of Molchanov and 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 $\mathbf{t}$ whose support 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 $\mathbf{T}$.[6] We will 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, $\mathbf{w}$.

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

Second, it relates the identified set Θ_I with the so-called Aumann expectation of specific random sets, either directly or indirectly as we will see below. As explained by Beresteanu et al. (2011), the choice between these two methods depend 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 Aumann expectations. In contrast, games of complete or incomplete information, or independence restrictions, are easier to deal with capacity functionals (see also Chesher and Rosen, 2015a).

3.2 Aumann expectations and support functions

As we focus our survey on partial identification derived from moment restrictions, we will concentrate on the use of Aumann expectations, although Section 4, which deals with inference, will generally encompass both frameworks. Defining the concept of Aumann expectation comes first. From 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). While 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.[7]

The *support function* of a convex set Θ is defined as:

$$\delta^*(q; \Theta) = \sup_{\theta \in \Theta}(q^\top \theta) \text{ for all directions, } q \in \mathbb{R}^d,$$

which uniquely characterizes the convex set Θ (e.g. Rockafellar, 1970):

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

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.

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

[**Include Figure 1**]

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) that says that the support function of an Aumann expectation is equal to the expectation of the support function of the underlying random set:[8]

$$E\left[\delta(q; \mathbf{T})\right] = \delta(q; \mathbb{E}(\mathbf{T})). \tag{6}$$

There are various uses of these results in the literature. First a direct approach using Aumann expectations is developed by Beresteanu and 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. Another direct approach is used by Bontemps et al. (2012) in the same case of best linear prediction with censored-by-interval outcomes although the number of moment conditions is now 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 $M(\theta)$ which is itself an Aumann expectation of a random set and the identified set can be characterized as:

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

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

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

3.3 Convex identified sets: A direct approach

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

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

in which $E(\mathbf{x}^\top \mathbf{u}) = 0$. Completing 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 $\eta = F(. \mid \mathbf{x}, \mathbf{d}, \mathbf{y}_L)$ 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 since the support $[0,1]$ of $\mathbf{t}$ is convex.

The random set of interest 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 and 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 $\left(E(\mathbf{x}^\top \mathbf{x})\right)^{-1}$).

Second, 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 in Stoye (2007) yield the support function as a function of population moments:

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

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

The geometry of the identified set and binding moment inequalities The geometry of the set Θ_I, has consequences for inference as developed in Bontemps et al. (2012). Specifically, two characteristics of frontiers of convex sets are *exposed faces* which are non-trivial (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 for 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 when the density function of covariates is not positive everywhere on its support (Bontemps et al., 2012).

The existence of non-trivial exposed faces has an impact on the asymptotic distribution of estimates that we review in the next section. The existence of kink points affects the number of moment inequalities which 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 which 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 non-singular cone, $C(\beta_0)$, are binding.

3.4 Convex identified sets: A two-step approach

This is 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 $\mathbf{x}$.[9]

First 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:

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

Note that $\pi = E(\mathbf{z}\mathbf{y}_L) + E(\mathbf{z}\mathbf{d}\mathbf{1}\{\mathbf{x} > \mathbf{0}\})$ is observable and that if $\pi \neq 0$ the orthogonality condition $E(\mathbf{z}\mathbf{u}^U) = 0$ 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. First augment the regression by adding $\mathbf{z}$ as an additional explanatory variable:

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

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: $E(\mathbf{x}\mathbf{u}) = 0, E(\mathbf{z}\mathbf{u}) = 0$. Therefore, the unrestricted identified set, say Θ_I^U, is obtained as in Subsection 3.3 by deriving its support function.[10] Reconsidering 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:

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

where q_β is associated to β_1 and q_γ to γ.

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

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

Moreover, the usual Sargan condition of the validity of moment restrictions, here $E(\mathbf{zu}) = 0$, 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 writes:

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

If it is not verified, the model is misspecified and moment restrictions are incompatible with the data.

Figure 2 presents the geometry of the problem for different cases in which the straight line $\gamma = 0$, (a) can cross the interior of the unrestricted identified set, Θ_I^U, and this results in interval identification of β_1, (b) can be tangent to this set and this restores point identification of β_1, (c) can have no intersection with Θ_I^U, a case of misspecification.

[Include Figure 2]

3.5 Non-convex 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). Return to the single dimensional best linear prediction

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

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

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

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

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

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:

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

Its Aumann expectation, $\mathbb{E}(\mathbf{M}(\theta))$, is convex although 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:

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

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

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

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 subsections.

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 in estimating its upper and lower bounds (e.g. Imbens and 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 in estimating the identified set first, next in 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 non-standard.

In a seminal paper, Chernozhukov et al. (2007) first estimate the identified set as the collection of points defined by values close to zero of a non-negative criterion function. For a given level of confidence, they similarly define confidence regions as the set of points whose criterion value is smaller than a critical value that is adjusted by subsampling techniques. Alternatively, Beresteanu and Molinari (2008) and Bontemps et al. (2012), estimate the support function of the convex identified set Θ_I as defined in Subsection 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 in inverting a test statistic. This method has been widely used in models that are characterized by moment inequalities (see, in particular, Romano and Shaikh, 2008, Andrews and Soares, 2010, Andrews and Shi, 2012). If Θ_I denotes the identified set and 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 non-rejected 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 here. 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). The second question is 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 set-up. In the test-inversion approach, we pay attention to the issue of selecting relevant moment inequalities (Andrews and Soares, 2011, Andrews and Barwick, 2012). We also present techniques adapted to the convexity arguments developed in Section 3. Another interesting case that we review, called *intersections of bounds* (Chernozhukov et al., 2013), is when parameters are bounded by an infinity of moments.

We end up this section by turning to the recently investigated issue that concerns 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

These two issues can be presented in a single framework.

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:

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

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, $I_n=[\hat{\beta}^L_{1,n}-\hat{c}^L_n,\hat{\beta}^U_{1,n}+\hat{c}^U_n]$, in which $\hat{\beta}^L_{1,n}$ and $\hat{\beta}^U_{1,n}$ are the estimators of the lower and upper bounds of the quantities defined by Equation (3), and in which $\hat{c}^L_n$ 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 for instance in Imbens and 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 which satisfy the asymptotic level of confidence of at least $1-\alpha$,

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

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 and Shaikh (2008, 2010) studied both in two different articles. Note however that the second condition is more restrictive than the first because singletons are degenerate regions (e.g., Henry and Onatski, 2012). Confidence regions covering sets are therefore generally larger than those covering points, $\theta \in \Theta_I(P)$.

Secondly, understanding the issue of uniformity can be approached first 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, 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 of this problem was proposed by Imbens and Manski (2004) and extended by Stoye (2009). The authors construct confidence intervals whose statistical properties are robust to the true diameter of the identified set and is particularly attractive when the model is point identified or close to be point identified.

Returning to the general case of covering a point uniformly, suppose that the true data generating process belongs to a family $\mathcal{P}$. In Example 1 of dependent variable censored by intervals 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 $1 - \alpha$ which satisfies:

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

Here too, the condition is more stringent than in the non-uniform case and uniform confidence regions are larger than those which have been previously defined. Yet this case seems the most interesting since 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

We describe inference techniques proposed in the recent literature that mainly deals with microeconometric models. This is why we assume in this Section that observations are independently 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 being 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:

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

For example, in Example 1 discussed above, arguments in Equations (4) are the two functions:

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

whose expectations are non-positive.

Chernozhukov et al. (2007) were the first to consider inference for a set defined by a non-negative 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 and Shaikh (2008), Andrews and Soares (2010) and many others.

They propose the following 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

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

An estimate of this criterion function for a sample of size n and observations $(y_i, x_i)_{i=1,.,n}$ 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 analogue

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

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 which satisfies the limit conditions:[11]

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

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 and this is done by subsampling (Chernozhukov et al., 2007, Romano and Shaikh, 2010) or by bootstrap (Bugni, 2010). Canay (2010) adopts an Empirical Likelihood approach and also proposes an adapted bootstrap method. These authors show that the confidence region thus constructed respects the asymptotic coverage condition given by Equation (9) or 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 and Sen, 2005), Andrews and Soares (2010) propose a method – which they call generalized moment selection (GMS) – for calculating

[11]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.

the critical value in an effective manner. It 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. It controls for size but is too conservative when only a few moments are binding because it increases the critical value.

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

Another way to get better finite distance behavior is to redefine the criterion $Q(\theta)$. Andrews and Barwick (2012) compare different ones. First, criterion $Q(\theta)$ in Equation (12) is of a Cramer 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 and Barwick (2012) illustrate the better performance of the KS type statistic by 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 and Barwick (2012) refine the selection procedure of Andrews and 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. A simplification of this very computationally intensive method and in particular when the number of moments is large is offered by Romano et al. (2014) 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 and Shi (2013) are specifically interested in the transformation of a finite number of conditional moment inequalities into unconditional moment inequalities whose number grows with the sample size. This is also the case of Lee et al. (2014) who propose a test of functional inequalities or conditional moments.

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

Other authors 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 examples of Section 3, convex analysis enables inference from a different perspective as in Beresteanu and 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 Subsection 3.3, can be expressed as an OLS estimator.

Namely, the expression of a point on the boundary of the identified set whose supporting hyperplane 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 in Beresteanu and 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 and Santos, 2014).

As discussed in Subsection 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 non positive 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 since it measures the distance between θ_0 and Θ_I. Note that this test statistic can be studentized since 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 p.d.f. 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 selecting 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 in Bontemps et al. (2012) based on OLS residuals.[12]

When the p.d.f. 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 non-trivial 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 LR test statistic is no longer a χ^2 distribution. Liu and Shao (2003) derive the asymptotic distribution

[12] Beresteanu and Molinari (2008) also propose estimates of the covariance operator of the support function estimator but they are slightly more complicated to compute than with this method.

of the test in this case. Conceptually, this is akin to a moment inequality set-up when we do not know the identity of binding moments although we deal here with a connected continuum of such inequalities.

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 whose support 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 since, in this case, the identified set has exposed faces not reduced to singletons and the empirical process is no longer asymptotically normal.[13]

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

4.4 Intersection of bounds

Subsection 3.4 above 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 (i.e. Manski and Pepper, 2000) and satisfy conditions as:

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

[13]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 and Santos (2014) assume that if the convex set has exposed faces, these directions are known.

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

$$\inf_z (h_n(z)) \tag{13}$$

is severely biased downwards in small samples since 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 is 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, whose variance 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 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 Subsection 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 consists in 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). A second approach is proposed by Romano and Shaikh (2008) and by Bugni et al. (2016) where the statistic of interest is concentrated out in the dimensions that are of no interest. Again, the resulting test statistics are not standard. The first paper proposes computation of the critical values by subsampling techniques while the second proposes 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 to do 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

A few authors have developed Bayesian methods for set identification.

Liao and Jiang (2010) work in a standard setting of moment inequalities. An interesting aspect 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 and Shorfheide (2012) work in a setting in which some reduced form parameters are point identified and in which this is the relationship between this parameter and the structural form parameters that generates partial identification. An entry game provides such an example since probabilities of simultaneous actions by agents are point identified. The posterior distribution function of structural parameters are derived from posteriors of the reduced form parameters. Their 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 issue by introducing a more general class of priors and the use of an inferior envelope of the posteriors to reconcile the Bayesian and frequentist approaches, at least asymptotically. The partial prior knowledge is modelled as a class and distinguished by whether 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 while 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 issue is developed by Kline and Tamer (2016). They show that there is an asymptotic equivalence between Bayesian and frequentist analyses when inference concerns the identified set rather than the partially identified parameter.

Liao and Simoni (2016) consider the estimation of closed and convex sets in a similar setting to that of Moon and Shorfheide. 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 analogue to the Bernstein and 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 papers focus on the posterior distribution of the set and not the partially identified parameter.

5 A sample of empirical applications and related topics

Even if the number of empirical applications is steadily increasing, most papers do not use these recent inference methods that we have just reviewed. A notable exception is 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 empirics of set identified structural models has not yet reached a mature level. These models are often estimated by using moment inequalities that exploit players' rational behaviour, combined with equilibrium constraints. The issue of sharpness is set apart in order to alleviate estimation costs.

Finally, convexity, though promising, is not fully exploited. We now briefly present these streams of empirical research while for a more complete and elaborate review of empirical applications, we refer the reader to Ho and Rosen (2015).

One of the first papers using partial identification concepts in an empirical framework is Hotz, Mullin and Sanders (1997). It is a reduced form model which 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 behaviour and outcomes. The instrumental variable in question is the occurrence of a miscarriage during pregnancy. Miscarriages do indeed provide a valid instrument, but for a sub-sample of the population, and is therefore contaminated in the sense of Horowitz and Manski (1995). The literature on treatment and selection also includes Manski and 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 and Pepper, 2013) and show that different assumptions lead to dramatically different conclusions.

Another example of reduced form estimation in a model with selectivity is Honoré and Lleras-Muney (2006). 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, Gosling, Ichimura and Meghir (2007) in the case of changes in the returns to education for men and for women in the United Kingdom over the last 30 years while dealing with non-participation. 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 endogeneous outcomes which are the decisions of firms to enter or not in a collection of independent markets (see Berry and Reiss, 2006, 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.[14] Ciliberto and 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 and Tamer (2009) and allows for both complete and incomplete information. He applies it to the estimation of 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 and 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 paper, Chesher and 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 in Haile and Tamer (2003). Armstrong (2013) derives bounds in the presence of unobserved heterogeneity, and Gentry and Li (2014) consider entry costs in auctions and derive bounds for the distributions of interest.

Many papers exploit rationality constraints in games/situations. An example is provided by Pakes (2010) and Pakes et al. (2015). The latter develops the estimation of structural models under general rationality constraints upon ordered choices (such as the number of bank ATMs) or in non-cooperative games between hospitals and HMOs. Inequality constraints on the parameters of

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

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 taken other decisions.

Finally, 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 the case the distribution of prices is discrete. This prevents point identification of price elasticities. This work is extended in Blundell et al. (2014). Henry and Mourifié (2013) study political competition and the spatial voting model and show how to test this model despite partial identification. The authors reject it using US data. Recent papers on networks exploit pairwise stability to estimate models of network formation. De Paula et al. (2016) adapt the solution of Ciliberto and Tamer (2009) to this problem whereas Sheng (2014), due to the curse of dimensionality because of the (exponentially) increasing number of 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 in order 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 the first example given in this article. 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 like in 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 would have recommended. Popper (2005) suggested 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, since 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 be able to compare hypotheses which are increasingly binding and that reduce the size of the identified set (e.g. the empirical strategy used by Manski and Pepper, 2000). This will not be the data that justify the credibility of research results since 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, they will have the option of conducting empirical reasoning that is rich enough to say that this assumption leads to such or such 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 for collecting new data and this would strengthen the credibility of empirical approaches in economics.

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

This is why test inversion might seem costly to applied researchers since, 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 handling 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 since 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 by 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.

REFERENCES

Andrews, D.W.K., and P. Guggenberger, 2009, "Validity of Subsampling and Plug-in Asymptotic Inference for Parameters Defined by Moment Inequalities", *Econometric Theory,* 25:669-709.

Andrews, D.W.K., and P.J. Barwick, 2012, "Inference for Parameters defined by Moment Inequalities: A recommended moment selection procedure", *Econometrica*, 80:2805-2826.

Andrews, D.W.K., and G. Soares, 2011, "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection", *Econometrica,* 78:119-157

Andrews, D.W.K., and X. Shi, 2013, "Inference Based on Conditional Moment Inequalities", *Econometrica,* 81:609-666.

Armstrong, T.B., 2013, "Bounds in auctions with unobserved heterogeneity", *Quantitative Economics*, 4:377-415.

Armstrong, T.B., 2014, "Weighted KS statistics for inference on conditional moment inequalities", *Journal of Econometrics*, 2014, 181:92-116.

Armstrong, T.B., 2015, "Asymptotically exact inference in conditional moment inequality models", *Journal of Econometrics*, 2015, 186:51-65.

Beresteanu, A., and F. Molinari, 2008, "Asymptotic Properties for a Class of Partially Identified Models", *Econometrica,* 76:763-814.

Beresteanu, A., I. Molchanov and F. Molinari, 2011, "Sharp Identification Regions in Models with Convex Moment Predictions", *Econometrica*, 79:1785-1821.

Beresteanu, A., I. Molchanov and F. Molinari, 2012, "Partial identification using random set theory", *Journal of Econometrics*, 166:17-32.

Berry, S., P. Reiss, 2007, "Empirical Models of Entry and Market Structure", In: M. Armstrong and R. Porter, Editor(s), Handbook of Industrial Organization, Elsevier, 2007, Volume 3.

Blundell, R., M. Browning and I. Crawford, 2008, "Best nonparametric bounds on demand responses", *Econometrica*, 76:1227-1262.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir, 2007, "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds", *Econometrica,* 75:323-363.

Blundell, R., D. Kristensen, and R. Matzkin, 2014, "Bounding Quantile Demand Functions using Revealed Preference Inequalities", *Journal of Econometrics*, 117:112-127.

Bontemps, C., T. Magnac and E. Maurin, 2012, "Set Identified Linear Models", *Econometrica*, 80:1129-1155.

Bugni, F., 2010, "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities", *Econometrica*, 78:735-754.

Bugni, F., I. Canay, and X. Shi, 2016, "Inference for Subvectors and Other Functions of Partially Identified Parameters in Moment Inequality Models",*Quantitative Economics*, forthcoming.

Chandrasekhar, A., V. Chernozhukov, F. Molinari, and P. Schrimpf, 2012, "Inference for Best Linear Approximations to Set Identified Functions", CEMMAP working paper 43/12.

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

Chernozhukov, V., D. Chetverikov, and K. Kato, 2016, "Testing many moment inequalities", CEMMAP working paper 42/16.

Chernozhukov, V., H. Hong, and E. Tamer, 2007, "Inference on Parameter Sets in Econometric Models", *Econometrica,* 75:1243-1284.

Chernozhukov, V., E. Kocatulum and K. Menzel, 2015, "Inferences on sets in finance", *Quantitative Economics*, 6:309-358.

Chernozhukov, V., S. Lee, and A.M. Rosen, 2013, "Intersection Bounds: Estimation and Inference", *Econometrica,* 81: 667-737.

Chesher, A., 2005, "Non Parametric Identification under Discrete Variation", *Econometrica,* 73:1525-1550.

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

Chesher, A., and A. Rosen, 2015a, "Characterizations of Identified Sets delivered by Structural Econometric Models", CEMMAP working paper 63/15.

Chesher, A., and A. Rosen, 2015b, "Identification of the Distribution of Valuations in an

Incomplete Model of English Auctions", CEMMAP working paper 30/15.

Chesher, A., and K. Smolinski, 2012, "IV models of ordered choice", *Journal of Econometrics*, 166(1), 33-48.

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

Conley, T., C. Hansen., and P. Rossi, 2012, "Plausibly Exogenous", *The Review of Economic and Statistics*, 94:260-272

Davezies, L., and X., d'Haultfoeuille, 2012, "Partial identification with Missing Outcomes", preprint.

Erickson, T., 1993, "Restricting Regression Slopes in the Errors in Variables Model by Bounding the Error Correlation", *Econometrica,* 61:959-69.

Fan, Y., and S. Park, 2010, "Sharp Bounds on the Distribution of the Treatment Effects and Their Statistical Inference", *Econometric Theory*, 26:931-951.

Fan, Y., and J. Wu, 2010, "Partial Identification of the Distribution of Treatment Effects in Switching Regimes Models and its Confidence Sets", *Review of Economic Studies*, 77:1002-1041

Fréchet, M., 1951, "Sur les tableaux de corrélation dont les marges sont données", *Annales de l'Université de Lyon*, III° Série Sci. A, 53-77.

Frisch, R., 1934, *Statistical Confluence Analysis by Means of Complete Regression Systems,* Oslo, Norway: University Institute of Economics.

Galichon, A., and M. Henry, 2009, "A Test of Non-Identifying Restrictions and Confidence Regions for Partially Identified Parameters", *Journal of Econometrics,* 152:186-196.

Galichon, A., and M. Henry, 2011, "Set Identification in Models with Multiple Equilibria", *Review of Economic Studies,* 78:1264-1298.

Galichon, A., and M. Henry, 2013, "Dilatation Bootstrap: a methodology for constructing confidence regions with partially identified models." *Journal of Econometrics,* forthcoming.

Gentry, M., and T. Li, 2014, "Identification in Auctions With Selective Entry", *Econometrica*, 82:315–344.

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., and D. Pacini 2013, "Measuring Inequality from Contaminated Data", preprint Bristol U.

Grieco, P.L.E. , 2014, "Discrete games with flexible information structures: an application to local grocery markets", *RAND Journal of Economics*, 45:303-340.

Guggenberger, P., F. Kleibergen, S. Mavroeidis and L. Chen, 2012, "On the Asymptotic Sizes of Subset Anderson Rubin and Lagrange Multiplier Tests in Linear Instrumental Variables Regression", *Econometrica*, 80, 2649-2666.

Haile, P., and E.Tamer, 2003, "Inference with an Incomplete Model of English Auctions", *Journal of Political Economy*, 2003, 111:1-51.

Heckman, **J.J,** 1978, "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica,* 46:931-960.

Henry, M., R., Meango and M., Queyranne, 2015, "Combinatorial approach to inference in partially identified incomplete structural models", *Quantitative Economics*, 6:499–529.

Henry, M., and I., Mourifié, 2013, "Euclidean Revealed Preferences: Testing the Spatial Voting Model", *Journal of Applied Econometrics*, 28(4), 650-666.

Henry, M. and A. Onatski, 2012, "Set Coverage and Robust Policy", *Economics Letters,* 115.2: 256-257.

Ho, K,. and A. Rosen, 2015, "Partial Identification in applied Research: Benefits and Challenges", CEMMAP Working paper 64/15.

Hoeffding, W., 1940, "Masstabinvariante Korrelationstheorie", *Shriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin,* 5(3): 179-233.

Honoré, B., and A. Lleras-Muney, 2004, "Bounds in Competing Risks Models and the War on Cancer", *Econometrica* 74: 1675-1698.

Horowitz, J.L., and C.F. Manski, 1995, "Identification and Robustness with Contaminated and Corrupted Data", *Econometrica*, 63:281-302.

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

Hotz, V. J., C. Mullin, and S.G. Sanders, 1997, "Bounding causal effects using data from a contaminated natural experiment: analysing the effects of teenage childbearing", *The Review of Economic Studies*, 64(4), 575-603.

Imbens, G., and C.F. Manski, 2004, "Confidence Intervals for Partially Identified Parameters", *Econometrica,* 72:1845-1859.

Kaido, H., F. Molinari, and J. Stoye, 2016, "Confidence Intervals for Projections of Partially Identified Parameters", working paper.

Kaido, H., and A. Santos, 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", Working paper.

Klepper, S., and E.E. Leamer, 1984, "Consistent Sets of Estimates for Regressions with Errors in all Variables", *Econometrica,* 52:163-183.

Kline, B., and E. Tamer, 2016, "Bayesian inference in a class of partially identified models", *Quantitative Economics*, 7, 329–366.

Komarova, T., 2013, "Partial Identification in Asymmetric Auctions in the Absence of Independence", *The Econometrics Journal*, 16(1):60-92.

Leamer, E.E., 1987, "Errors in Variables in Linear Systems", *Econometrica,* 55(4): 893-909.

Lee, D.S., 2009, "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects", *Review of Economic Studies,* 76(3):1071-1102.

Lee, S., K. Song, and Y. Whang, 2014, "Testing for a general class of functional inequalities", CEMMAP working paper 09/14.

Liao, Y., and W. Jiang., 2010, "Bayesian Analysis in Moment Inequality Models", *The Annals of Statistics*, 38(1), 275-316.

Liao, Y., and A. Simoni, 2016, "Bayesian inference for partially identified convex models: is it valid for frequentist inference?", working paper.

Liu, X., and Y. Shao, 2003, "Asymptotics for Likelihood ratio tests under loss of Identifiability", *The Annals of Statistics*, 31:807:832.

Magnac, T., and E. Maurin, 2008, "Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data", *Review of Economic Studies,* 75:835-864.

Manski, C.F., 1989, "Anatomy of the Selection Problem", *Journal of Human Resources,* 24:343-60.

Manski, C.F., 2003, *Partial Identification of Probability Distributions*, Springer-Verlag: Berlin.

Manski, C.F., 2005, "Partial Identification in Econometrics", *New Palgrave Dictionary of Economics.*

Manski, C.F., and J.V. Pepper, 2000, "Monotone instrumental variables: with an application to the returns to schooling." *Econometrica,* 68(4): 997-1010.

Manski, C.F., and J.V. Pepper, 2013, "Deterrence and the Death Penalty: Partial Identification Analysis Using Repeated Cross Sections", *Journal of Quantitative Criminology*, 29(1), 123-141.

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

Marschak, J., and W.H. Andrews, 1944, "Random Simultaneous Equations and the Theory of production", *Econometrica*, 12:143-205.

McFadden, D., 1994, "Contingent Valuation and Social Choice", *American Journal of Agricultural Economics*, 76:689-708.

Menzel, K., 2014, "Consistent estimation with many moment inequalities", *Journal of Econometrics*, 182(2), 329-350.

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

Molchanov, I., and F. Molinari, 2015, "Applications of random set theory in econometrics", *Annu. Rev. Econ.*, 6(1), 229-251.

Moon, H. R., and F. Schorfheide, 2012, "Bayesian and frequentist inference in partially identified models", *Econometrica*, 80(2), 755-782.

Nevo, A., and A. Rosen, 2012, "Identification with imperfect instruments", *Review of Economics and Statistics*, 94(3), 659-671.

Pacini, D., 2016, "Two-Sample Least Squares Projection", forthcoming *Econometric Reviews.*

Pakes, A., 2010, "Alternative Models for Moment Inequalities", *Econometrica*, 78:1783-1822.

Pakes, A., J. Porter, K. Ho and J. Ishii, 2015, "Moment Inequalities and their Applications", *Econometrica*, 83(1), 315-334.

de Paula, A., 2016, "Econometrics of Network Models", CeMMAP Working Paper 06/16.

de Paula, A., S. Richards-Shubik and E. Tamer, 2015, "Identification of Preferences in Network Formation Games", CeMMAP Working Paper 29/15.

Ponomareva, M., 2010, "Inference in models defined by conditional moment inequalities with continuous covariates", Unpublished manuscript, Department of Economics, University of Western Ontario.

Ponomareva, M. and E. Tamer, 2011, "Misspecification in moment inequality models: back to moment equalities", *Econometrics Journal,* 14:186-203.

Popper, K., 2005, *The Logic of Scientific Discovery*, London and New York: Routledge / Taylor and Francis.

Redner, R.A., 1981, "On the assumptions used to prove asymptotic normality of maximum likelihood estimates", *Ann. Math. Statistic*, 9:225-228.

Reiersol, O., 1941, "Confluence analysis by means of lag moments and other methods of confluence analysis ", *Econometrica*, 9:1-24.

Ridder, G. E., and R. Moffitt, 2007, "The econometrics of data combination", in J. J.Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, North-Holland, Amsterdam.

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

Romano, J.P., and A.M. Shaikh, 2008, "Inference for Identifiable Parameters in Partially Identified Econometric Models", *Journal of Statistical Planning and Inference*, 138:2786-2807.

Romano, J.P., and A.M. Shaikh, 2010, "Inference for the identified set in partially identified econometric models", *Econometrica*, 78(1), 169-211.

Romano, J.P., A.M. Shaikh, and M. Wolf, 2014, "A Practical Two-Step Method for Testing Moment Inequalities", *Econometrica*, 82(5), 1979-2002.

Rosen, A.M., 2008, "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146:107-117.

Rothenberg, T. J., 1971, "Identification in parametric models", *Econometrica,* 39:577-591.

Sheng, S., 2014, "A Structural Econometric Analysis of Network Formation Games", working paper.

Silvapulle, M., and P. Sen, 2005, *Constrained Statistical Inference*, Wiley:London.

Stoye, J., 2007, "Bounds on Generalized Linear Predictors with Incomplete Outcome Data", *Reliable Computing*, 13: 293–302.

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

Tamer, E., 2003, "Incomplete simultaneous discrete response model with multiple equilibria", *The Review of Economic Studies*, 70(1), 147-165.

Tamer, E., 2010, "Partial Identification in Econometrics", *Annual Review Econ.*, 2:167–95

Wooldridge, J. M., 2010, *Econometric Analysis of Cross Section and Panel Data*, MIT Press:Cambridge.

Figures

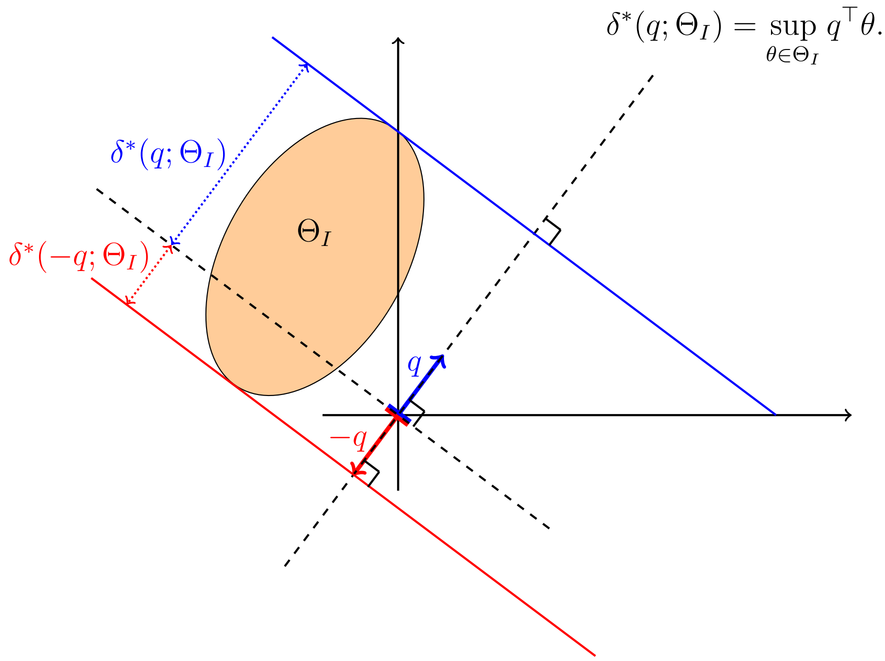
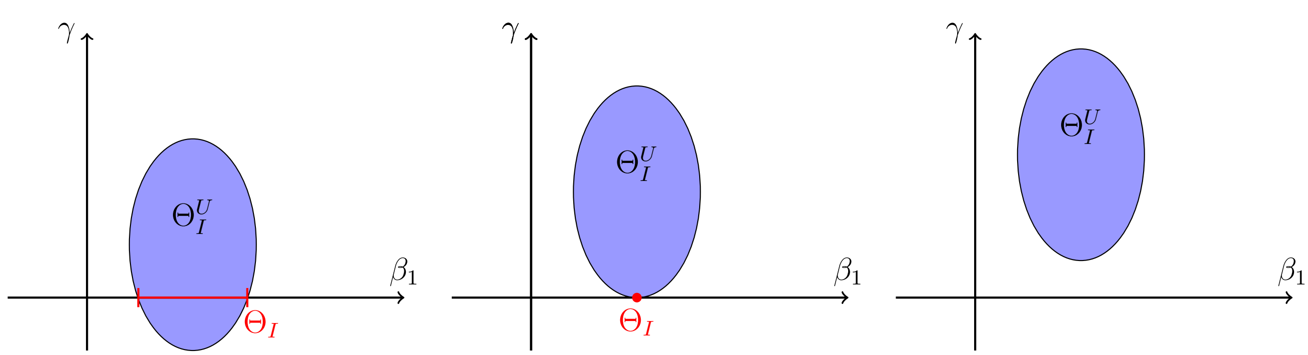


Figure 1: The support function



(a) Θ_I shrinks (b) Point Identification (c) Misspecification

Figure 2: Geometry with one supernumerary instrument