Partial Identification in Econometrics*

Elie Tamer
Northwestern University

October 2, 2009

Abstract

Identification in econometric models maps prior assumptions and the data to information about a parameter of interest. The partial identification approach to inference recognizes that this process should not result in a binary answer that consists of whether the parameter is point identified or not. Rather, given the data, the partial identification approach characterizes the informational content of various assumptions by providing a menu of estimates, each based on different sets of assumptions, some of which are plausible and some aren't. Of course, more assumptions begets more information, and so stronger conclusions can be made at the expense of more assumptions. The partial identification approach advocates a more fluid view of identification and hence provides the empirical researcher with methods that help study the spectrum of information that we can harness about a parameter of interest using a menu of assumptions. This approach links conclusions drawn from various empirical models to sets of assumptions made in a transparent way. It allows researchers to examine the informational content of their assumptions and their impacts on the inferences made. Naturally, with finite sample sizes, this approach leads to statistical complications, as one needs to deal with characterizing sampling uncertainty in models that do not point identify a parameter. So, new methods for inference are developed. These methods construct confidence sets for partially identified parameters, and confidence regions for sets of parameters, or identifiable sets.

*I thank Brendan Kline and especially Charles Manski for comments and the National Science Foundation for Research Support.

"*The law of Decreasing Credibility:* The credibility of inference decreases with the strength of the assumptions maintained," Manski (2003)

"A fragile inference is not worth taking seriously," Leamer (1985)

1 Introduction

Partial identification in econometrics is an approach to conducting inference on parameters in econometric models that recognizes that identification is not an all or nothing concept and that models that do not point identify parameters of interest can and typically do contain valuable information about these parameters. This partial identification approach favors the principle that inference -and conclusions and actions-based on empirical models with fewer suspect assumptions is more robust, hence more sensible and believable. Stronger assumptions will lead to more information about a parameter, but less credible inferences can be conducted. This is inline with Coombs (1965)'s principle of buying information with assumptions.

There is only so much that data alone can inform us, and generally, it is not possible to do inference without any assumptions, i.e., without a model. The partial identification approach to econometrics views economic models as sets of assumptions, some of which are plausible -e.g., based on economic principles that respect constraints and optimizing behavior- but others that are esoteric and only needed to complete a model. These are usually termed functional forms or distributional assumptions. Partial identification calls for analyzing the sensitivity of our inferences on the parameter of interest to these esoteric assumptions. This approach to inference in econometric models does not advocate that the only way to learn about parameters is via nonparametric models with minimal assumptions. On the one hand, it tries to determine as a first step the limit of what we can learn with only the empirical evidence (the data -in a nonparametric setup). On the other hand, in a fully parametric model, this partial identification approach examines the effect assumptions have on the information the model contains about a parameter of interest.

For example, it is accepted that (unobserved) heterogeneity plays a key role in empirical micro econometrics models. Economic theory is largely silent regarding the choice of the distribution of unobserved heterogeneity, and in many cases, the choice of this distribution has been made based on folklore, familiarity, and computational grounds[1]. This is especially important in nonlinear models where mean independence assumptions are not sufficient. In these models, it is important to examine the role played by assumptions made on the heterogeneity distribution.

[1] Heckman and Singer (1984) examine the role distributional assumptions play in duration models.

In the last 30 years, there have been reactions within the empirical literature against the fragility of inferences based on suspect assumptions. And so, there is a move especially in the labor economics literature to look for the smallest set that delivers point identification. Some of these approaches, championed by the semiparametric econometricians, provided models that rely less and less on ad-hoc assumptions while maintaining point identification. These semiparametric models use strong support conditions on the observed data in addition to the commonly used exclusion restrictions. On the other hand, the less stylized the empirical model is, the harder it is to obtain a semiparametric assumption that point identifies the parameters. This tension between guaranteeing point identification while maintaining the weakest possible set of assumptions has partially limited the use of semiparametric approaches in more complicated (nonlinear) models.

Parametric models are useful. Adding plausible assumptions that are widely accepted and based on economic principles can be a vehicle to communicate insights and conclusions. It can help advance the scientific exercise and enrich it. But, in some situations, it can also provide misleading answers: What good are sharp results that ignore model uncertainty or model misspecification? In the last 20 years researchers have began to embrace the idea that identification is NOT an all or nothing matter, and that a set of plausible assumptions that does not deliver point identification, can still contain useful information about parameters of interest.

This partial identification view has been motivated by the fact that point identification is not the objective by itself, and in essence takes us back to the Koopmans and Reiersol's dictum whereby the specification of a model ought to be based on the underlying economics, prior knowledge (such as linearity of variable cost that people have established for this industry) or other assumptions with universal or almost universal acceptance, but not geared primarily towards point identifying the parameters. "Scientific honesty demands that the specification of a model be based on prior knowledge of the phenomenon studied and possibly on criteria of simplicity, but not on the desire for identifiability of characteristics that the researcher happens to be interested in" (Koopmans and Reiersol (1950) Pages 169-170). Once the structure is specified, the model can either have no information about the parameter of interest, restrict the parameter of interest to a non-trivial set, or point identify the parameter of interest. This is exactly the domain of identification analysis, with partial identification taking the view that identification is not only about verifying whether the third case holds, but also determining the extent of information contained in the second and linking this to the type of assumptions that the researcher proposes in the model.

Partial identification analysis can be conducted from the bottom up, whereby a researcher first considers whether the data alone provide any information about the parameter of interest. Then, the researcher combines the empirical evidence with a list of assumptions and studies the effects these assumptions have on what and how one learns. On the other hand, in some examples, it is easier to start with a top-down approach in which a fully parametric model that point identifies the parameter of interest using a set of assumptions is first considered. Each of the unsettled assumptions yields to a different model that point identifies a value for the parameter of interest, and so this sensitivity analysis approach collects in a set, different values of the parameter of interest that correspond to the different models. For a similar specification analysis approach, see Leamer (1985).

Identification of a parameter of interest basically posits the existence of an infinitely large sample size and asks the question of what one can learn about this parameter. Point identification analysis answers the question of whether the parameter of interest can be recovered uniquely given this infinite data set, while partial identification considers the question of what can be learned about this parameter in the presence of an infinite data set and when considering various sets of assumptions. The requirement of studying identification with an infinitely large sample separates the question of identification from the distinct but also important question of statistical inference from a finite sample size. The two questions, identification and statistical inference, are linked and partial identification has created an important set of new statistical inference problems that require new methods and new approaches. For example, in cases where one is interested in inference on the (nontrivial) identified *set*, statistical methods geared towards estimating sets are required, and more importantly, methods to build "confidence regions for sets" are needed. I will summarize the main issues that econometricians face when handling statistical issues related to partial identification below.

This article starts with a literature review that highlights the ideas of partial identification from the early 1930's on. The next section discusses two important examples in which the partial identification approach is described and applied. Section 4 discusses statistical inference, and Section 5 concludes.

2 Literature Review

The literature on identification of economic models has been a cornerstone of the empirical research program in econometrics dating back to the early work on estimating simultaneous equations model of demand and supply in the 1920's and 1930's. The classification of variables into exogenous and endogenous, and recognizing the identification problems that this endogeneity creates have been considered a distinguishing feature of econometrics in relation to statistics, which is typically concerned with the statistical properties of estimators and where identification, or uniqueness of the optimum of some objective function, is sometimes directly assumed (especially in nonlinear models).

It is not clear why the issue of failure of point identification and the impact of partial identification have been largely ignored in both econometrics and statistics before the 1990's. This is especially surprising since even a slight breakdown in point identification leads to changes in the asymptotic theory of estimators, which in turns requires a modification of the standard procedures derived under point identification. Given that point identification is often times assumed, it is surprising that not enough work has been devoted to studying properties of models that fail to point identify the parameters and the effect of this failure on the statistical properties of estimators. Phillips (1989), for example, states that " It seems important that we should understand the implications of identification failure for statistical inference. Yet, this is a subject that seems to be virtually untouched in the literature."

I will review the literature on partial identification in econometrics. I start with the early works, which have largely been glanced over by most econometricians, and have had (almost) no influence on the empirical literature before the 1990's. I focus on works in econometrics, but there have been similar ideas in other literatures[2]. I will then describe some of the main recent developments in the literature. We start here with Frisch's approach to confluence analysis.

2.1 Partial Identification and Frisch's "True Regressions"

In his 1934 manuscript (Frisch 1934), R. Frisch was concerned with the problem of "confluency" in linear regression whereby the results of a linear regression of one variable on a set of variates are suspicious if one or more of these variates are almost perfectly correlated. To disentangle true relationships between "variates" that ultimately are used in regressions analysis, Frisch considered the case where the object of interest is the matrix of correlation among a set of unobserved variates "x'" when we observe a vector x that is a convolution of x' with a set of "accidental disturbances" that are of "no interest."

The main motivation is the study of correlation among random variables and allowing for the possibility that the this true correlation is among variables that are unobserved. These unobserved random variables constitute the "True Regression." So, this is a classic identification problem where the observed data consists of the vector x

[2]For example, Bielby and Hauser (1977) in their review of structural equation models in sociology (see page 150 and references therein), noted that identification is not an "all-or-nothing proposition" and cited partial identification works in sociology based on insights from Marschak and Andrews (1944), which we review below.

while the object of interest is the second moment matrix of the vector x' and where $x = x' + x''$. Frisch writes the observed second moment matrix as a function of the true moment matrix and nuisance parameters that represent the effect of the disturbances, or measurement error. He uses a set of assumptions, uncorrelation between the cross equation disturbances, and uncorrelation between the disturbances and the systematic parts, to reduce the dimensionality of the problem. To illustrate, Frisch considered the two variate model

$$x_1 = x_1' + x_1''$$
$$x_2 = x_2' + x_2''$$

where we observe the second moment matrix $E[(x_1, x_2)'(x_1, x_2)]$ and thus we are able to relate it to the second moment matrix of (x_1', x_2') under the uncorrelation assumptions. So, it is easy to see that for example, $cov(x_1, x_2) = cov(x_1', x_2')$. Assuming that this latter is positive, we have

$$\frac{cov(x_1, x_2)}{var(x_2)} = \beta_1 \frac{var(x_2')}{var(x_2') + var(x_2'')} \leq \beta_1$$

where β_1 is the slope in the true regression of x_2' on x_1'. It is also easy to see that

$$\frac{var(x_1)}{cov(x_1, x_2)} = \beta_1 + \frac{var(x_1'')}{\beta_1 var(x_2')} \geq \beta_1$$

Frisch concludes that the slope coefficient of the true regression of x_1' on x_2' must lie in the **"possibility set"** $\beta \in [\frac{cov(x_1,x_2)}{var(x_2)}, \frac{var(x_1)}{cov(x_2,x_2)}]$. The endpoints of this interval

> "form limits between which the true slope must lie whenever the assumptions specified hold good. But there is nothing in the observed correlation matrix which permits to choose between the above two limits, or to fix any number intermediate between them. Thus it is when, and only when, there is a good agreement" between the endpoints do we get to draw "definite conclusions about the true regression slopes" (page 86).

So, here Frisch covers the essential principles in a partial identification analysis: he derives the identified set, or as he calls it the "possibility set," which can be estimated since it is a function of the observed variables, and also argues that this set is sharp, i.e., any value in the set including the endpoints cannot be rejected as the true value of the slope. He provides a simple and clear analysis of the relationship between the data and the underlying structure. Frisch posits a classical measurement error model in which the disturbances are uncorrelated with the systematic variables, and under these assumptions he derives the information about the true slope.

The current literature on measurement error in linear models uses instrumental variables to obtain point identification of the true slopes. This direction of the literature is not driven by the belief that exclusion restrictions are more robust than the Frisch uncorrelation restrictions, but rather it seems to be driven by the need to obtain "definite conclusions," in terms of point identification. Sometimes, relying on exclusion to obtain point identification in measurement error model is reasonable, but empirical analysts can easily compute Frisch like bounds as an approach to learning about the parameters when exclusion restrictions are suspect, or are not available. The statistics literature relies on one having knowledge of the measurement error process via validation data or estimation of reliability ratios (see Fuller (1987) for example). This typically yields point identification, but these data are not easy to obtain in typical economics surveys.

2.2 The Partial Identification Approach of Marschak and Andrews

In an important paper, Marschak and Andrews (1944) (MA) studied the problem of inference on production functions and showed that by exploiting the economic theory of production, such as conditions for profit maximization under constraints for example, that the parameters of this production function "can be confined to relatively localized regions of the parameter space on the basis of available observations." This "Partial Identification" approach[3] uses economic restrictions to derive bounds in a parametric model of supply. We review the approach using an example that was used by Nerlove in his review of MA's paper (See chapter II in Nerlove (1965)).

Suppose we are interested in estimating the Cobb-Douglas production function

$$\begin{aligned} Y_{0f} - \alpha_1 X_{1f} - \alpha_2 X_{2f} &= A + u_{0f} \\ \beta_0 X_{0f} - \beta_1 X_{1f} &= B_1 + u_{1f} \\ \beta_0 X_{0f} - \beta_2 X_{2f} &= B_2 + u_{2f} \end{aligned} \tag{2.1}$$

where Y_{0f}, X_{1f}, X_{2f} denote the output and inputs 1 and 2 respectively for firm f, and the unobservable u's are interpreted as the distance between its production and the average production. These can measure efficiency of the firm, but also contain other unobserved qualities that are all aggregated into this unobservables.

The first equation above is the production function (in logs), while equations 2 and 3 represent the first order conditions from profit maximizations taking into account the demand function in inputs 1 and 2 (They deal with the more general case of imperfect competition), where for example a specific functional form for the demand is such that

$$\beta_i = 1 + \frac{1}{\eta_i},\ i = 0, 1, 2$$

[3]The title of Nerlove's chapter reviewing the MA approach is "Partial-Identification: The Marshack-Andrews Approach."

where η_i's are the elasticities which one can show obey the following inequalities: $0 < \beta_0 \leq 1$ and $\beta_i \geq 1$ for $i = 1, 2$. The standard current approach for inference on the above model uses exclusion restrictions, or variables that influences the production of one factor, and not others.

MA take a different approach. Assuming that firms maximize profits taking into account their production function and market demand, MA use the second order optimization conditions to place bounds on the parameters of the production function, (α_1, α_2). So for example, when there is perfect competition (β's equal to one), (α_1, α_2) must lie in the triangle bounded by the line connecting $(0, 1)$ to $(1, 0)$. The different bounds on (α_1, α_2) are graphed in Figure 1 below (taken from Nerlove) where we see that as we move away from perfect competition, the size of the identified set increases. The importance of the MA approach is recognizing that though the bounds without any assumptions can be wide, one can use more assumptions motivated by the economics of the problem to shrink the allowable regions.

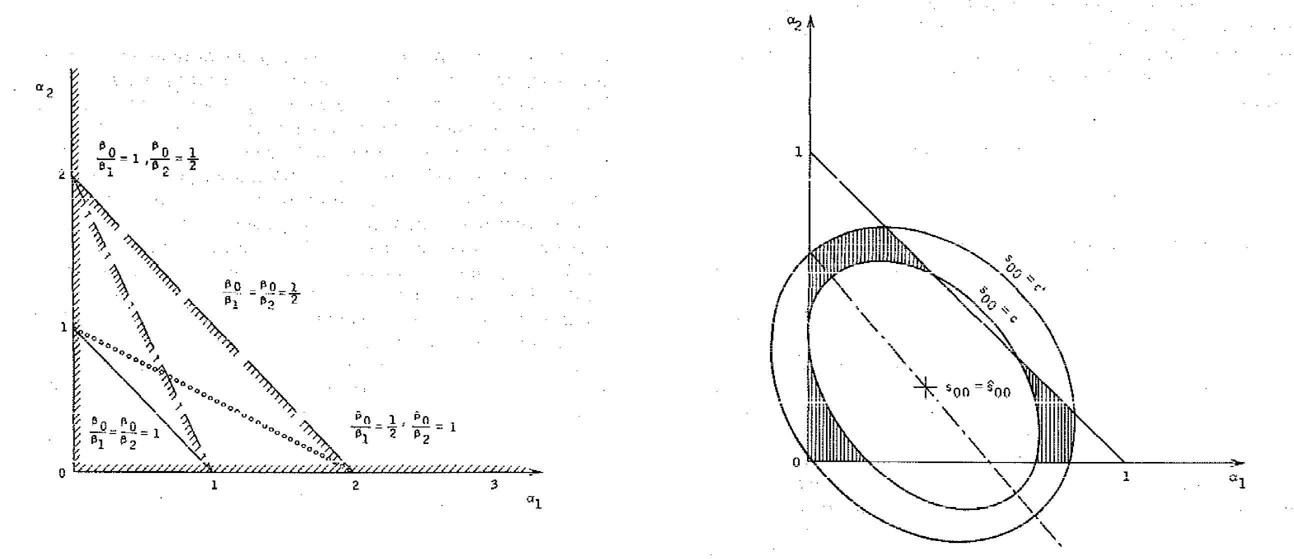
MA use restrictions on the variance of u_0 in (2.1) above. Arguing that u_0 represents technical efficiency of the firm (given as a deviation from the mean of efficiency of all firms), they assume that the "best" firm cannot be more than say 5 times as good as the "worst firm" but no less than 4, which leads to bounds on $var(u_0)$. These proposed bounds on this variance can be used along with the first equation in (2.1) to narrow the set. For example, we see in the RHS panel of Figure 1 that there are substantial identification gains using these variance restrictions.

In the MA paper, we find a first example where the partial identification approach to inference was used in an economic model using assumptions that are motivated by the underlying economic problem. The results are left as diagrams in the paper, but it would be interesting to map them into estimates of the production function, and to compare them to ones obtained using instrumental variables. This approach to inference in parametric structural models has largely been skimmed over and ignored. A noted exception, is the work of Leamer (1981), who revisited MA and Leontief and examined the problem of learning about elasticities of demand and supply in linear simultaneous equations system with uncorrelated errors. He shows that sets of parameters that lie on a hyperbola are identified. This means that the forward regression of price on quantity understates the elasticity of supply (in case the latter is positive), while a reverse regression provides an upper bound.

2.3 Other influential works

In addition to the above, the work of Frechet (1951) in the statistics literature on whether knowing the marginal distributions of continuous random variables X and Y tells us anything about their *joint* distribution, is important and influential. Frechet

Figure 1: Bounds on (α_1, α_2)



LHS: This panel provides the identification region under various conduct parameter assumptions. It uses the functional forms assumptions on the production function and the demand functions and the optimization restrictions to derive these bounds. The smallest triangle (the line connecting $(0,1)$ to $(1,0)$) is obtained under perfect competition. RHS: Here MA add restrictions on the size of the disturbance in the production function which shrinks the regions substantially.

showed that given knowledge of the distribution functions $F(.)$ and $G(.)$ of X and Y respectively, their joint distribution $K(.;.)$ is such that for all $(a,b) \in R^2$

$$\max(F(a) + F(b) - 1, 0) \leq K(a,b) \leq \min(F(a), F(b))$$

These bounds are attainable, i.e., $\min(F(a), F(b))$ and $\max(F(a) + F(b) - 1, 0)$ are proper distribution functions with extreme forms of correlation structures[4]. This is a very important result that has been extended and applied in many areas[5].

Another early work that contains partial identification ideas is Duncan and Davis's work (See Duncan and Davis (1953)) on the ecological problem. This problem is one of learning the conditional distribution of a random variable Y given (X, Z) when data are available from two random samples. One that contains data on (Y, Z) but not X, and the other contains (X, Z) and not Y. Duncan and Davis conducted partial identification analysis that were later studied more formally in Cross and Manski (2002).

[4]See Nelson (1999) for more about this and other results regarding the Fréchet bounds.

[5]Bounds on joint distribution given knowledge of the marginals can be used for example in competing risks models that are useful in economics.

2.4 Early Reaction to Partial Identification

As we see from above, the ideas about partial identification have been around in econometrics since the 1930's, and the rationale for such exercise could not have been more relevant or pertinent. But, it is safe to say that the empirical and theoretical fields in econometrics have largely ignored or given little attention to these ideas at least until the late 1980's. Researchers have been tentative at best in considering partial identification. The semiparametric literature in microeconometrics has certainly looked for the model that makes fewer assumptions, but (almost) always guaranteeing that these assumptions deliver point identification, regardless of whether these sufficient point identification conditions are suspect, or whether they hold in their data.

I shall speculate and offer two reasons about this incertitude regarding partial identification. One of the main motivations for empirical work in economics is to evaluate policies. One important purpose of this is making decisions, and so with partially identified models, an empirical strategy that provides "multiple answers" is generally viewed as a drawback. This feeling among empirical economists and econometricians can be summarized using Nerlove (1965)'s comment on Marschak and Andrews' partial identification approach above. Nerlove enumerates as the the most important "fundamental defect" of their (MA's) approach is that "it is impossible to obtain unique estimates of the parameters of the production function. All that can be done is to restrict their possible values to a more-or-less narrow range." Data alone we are told are not informative, the bounds are usually wide, and so partial identification analysis in and by itself is not useful. However, one rational for partial identification is exactly that data alone are sometimes not useful, but data and theory are. And so, the purpose of empirical work is to harness and link results and conclusions to theory and models, clarifying what conclusions follow from theory, and what don't; this is the essence of partial identification[6]. This skepticism based on "multiple answers" does not rest on solid arguments that are coherent, and so it is likely that it will subside.

Another important reason for the skepticism about partial identification, is the idea that on some level, all empirical work in the social sciences is assumption based[7] and so it is not clear where we decide to draw the line. Which assumptions are "plausible" and which are not? Is the choice of regressors and what regressors to condition on a

[6]Moreover, with statistical uncertainty from using finite sample sizes, point identified models provide confidence regions for these uniquely identified parameters, where, heuristically, each parameter in these regions is statistically as likely to be the true parameter. So, conceptually, this argument of preferring a point to a set is not as solid as it might seem from a practical perspective, since when reporting results one ought to account for statistical uncertainty which makes the preference of point-like to set-like argument questionable.

[7]Here, one can argue that even raw data are theory driven, and that routine assumptions such as iid are conditioned on some assumptions. See Coombs (1965) for more on this.

more plausible assumption than conditional independence? These are valid questions and ought to be examined only in relation to a particular empirical problem. For example, in the Frisch model above, we are interested in the best linear predictor of one random variable given another under square loss. This object is valid under mild moment existence assumptions which are typically considered plausible. Overall though, there is a large set of what might be considered "plausible" assumptions or models, and guarding against all does not seem feasible both conceptually and computationally. So, some recent work on partial identification have considered fully parametric models, and have studied the identification power of *some assumptions* by examining the informational content of these models absent these assumptions. The particular assumptions that were relaxed are ones in which the community of researchers consider the most controversial, are untestable (or least likely to be testable even with more data collection) and thus constitute a leading cause for unease. See section 3 for examples. These (partial) sensitivity exercises, where one uses a partial identification approach to study the sensitivity to some assumptions, can be helpful to communicate ideas and discriminate among various modeling approaches. The answer to this second concern is to acknowledge that although full fledged sensitivity analysis against all "plausible" models is not always possible, accounting for model uncertainty even for some parts of that model is always worthwhile[8] (See Heckman and Hotz (1989) and the rejoinder for more on this).

2.5 Recent Literature

Given the above reservations, the econometrics literature, both theoretical and empirical, did not pay as much attention to partial identification largely until the work of Manski and his collaborators starting in the late 1980's. This and other work inspired by it have revived and buttressed the partial identification view within empirical economics as a valid, coherent, and sensible approach to inference. Starting with Manski (1989) and Manski (1990) that analyzed the problem of self selection into treatment from a partial identification perspective, Manski urged empirical economists to be cautious of assumption-driven conclusions especially in the aftermath of the general skepticism in the labor economics literature at the time[9] about the usefulness of parametric models of selection. Manski's work then provided a "worst case" bounding strategy that is simple to understand and easy to compute. These bounds summarize what the data, and only the data say about the parameter of interest. He has gone on since then to make contributions to partial identification in many settings. A clear synthesis of

[8] This second concern should not be used as an excuse for researchers to not worry about the sensitivity of their results to the assumptions made.

[9] See LaLonde (1986) for example.

these results can be found in Manski (2003) and Manski (2007). A Manski-style approach to partial identification advocates the bottom-up approach in which first worst case bounds are derived, and gradually stronger assumptions are added and their effects are analyzed. For example, in Manski (1997), he examined the identifying power of monotonicity (and equilibrium) when estimating demand functions using an independent set of data on quantity, and price from a cross section of markets. Another example is Manski and Pepper (2000) where the authors revisit an important issue in labor economics and that is of estimating wage regression as a function of schooling. The conceptual contribution of the paper is the approach whereby one need not assume full statistical independence between the counterfactual outcomes and an instrument, but one can explain a form of monotonicity, akin higher values of one leads to higher values of the other. The paper shows how this type of assumption contains identifying power.

Some other partial identification works include: Bollinger (1996)'s work extending the Frisch bounds to cases with misclassification; Hotz, Mullin, and Sanders (1997)'s study of contaminated instruments as applied to the effect of teenage pregnancy on later outcomes using the partial identification approach to contaminated models in Horowitz and Manski (1995); Tamer (2003)'s study on inference in entry models with multiple equilibria, generalized further to cover cases without assuming Nash equilibrium in Aradillas-Lopez and Tamer (2008); inference on parameters in linear regressions with *interval data* on outcomes or a regressor using a partial identification approach in Manski and Tamer (2002); Blundell, Gosling, Ichimura, and Meghir (2007) bounds on wage distribution in the UK using worst case and other more informative bounds. This is a sample of recent works on partial identification on various problems.

2.5.1 Work on Inference

Most of the above literatures have been concerned with identification, as a separate question from statistical inference. The first paper in econometrics to tackle some aspects of inference in models with partially and nonidentified parameters was the paper by Phillips (1989) where novel statistical inference methods were proposed in models where some of the parameters are partially identified or "nonidentified." The methods rely on rotations of the parameter space into one where all parameter are identified. Throughout the 1990's, most partially identified models were such that the boundary of the identified set can be derived explicitly as a functional of the observed data distribution. The form of the typical bound on a parameter θ, say, would be of the form $F_1 \leq \theta \leq F_2$, and (F_1, F_2) can be consistently estimated from the data. Moreover, a confidence region for $[F_1, F_2]$ was usually constructed by jointly bootstrapping the endpoints. It was then noticed in a paper by Imbens and Manski

(2004) that narrower confidence regions can be reported if one considered covering not the set, but the actual identified parameter (these bounds were later refined in Stoye (2009)). On the other hand, and in the context of *linear models* with interval data, Manski and Tamer (2002) constructed sets of parameters that are *consistent* in the Hausdorff metric to the argmin of a particular objective function. This result generalizes the classical case of consistency of m-estimators to cover the case where the argmin of the population objective function is not a singleton. Manski and Tamer's work did not contain inference or confidence region procedures. There are a number of current papers that provide approaches to inference in partially identified models. Chernozhukov, Hong, and Tamer (2007) (CHT) provided a more general set consistency results with rates of convergence for argmins of general objective functions. They also provided methods to construct confidence regions that cover the identified set or the true parameter with a prespecified probability. These methods are based on subsampling. Similar methods are also provided in Romano and Shaikh (2007), (2008) using subsampling. Beresteanu and Molinari (2008) introduced tools from random set theory to study inference about the identified set. Since many of the partially identified models can be represented by moment inequalities, there has also been a flurry of papers recently that develop inferential methods designed to cover moment inequalities (and equalities). These include Rosen (2008), Bugni (2007), Andrews and Soares (2009), Canay (2007), Stoye (2009) and Andrews and Jia (2008) among others.

3 Two Examples

Here, I provide two examples that showcase the partial identification approach to inference. The first example builds on the canonical missing data problem and shows how a partial identification approach links the assumptions used to information provided using a more nonparametric approach as compared to the second example. There, I consider a fully parametric model in which a set of parametric assumptions are relaxed, and hence the identified features of a set of models are studied. In this section also, we will assume away sampling issues and hence will be concerned about the identification problem as distinct from statistical inference. Inference from finite sample sizes will be examined in section 4 below.

Operationally, identification questions can be written as the problem of analyzing and characterizing the argmin set of a properly defined objective function. This objective function can be a likelihood, or a moment based objective function, and the important property that this function must satisfy is that it is constructed in such a way that its set of minima exhausts all the information in the model given the maintained assumptions. We will illustrate these ideas also in the next two examples.

3.1 Example 1: Missing data

We examine a series of examples in which we highlight the partial identification approach. We start with the missing outcome problem when this outcome is binary. We also characterize the mechanics of the information the model contains about the parameter of interest and express it as the argmin of an objective function.

3.1.1 Missing Binary Outcomes

Let Y be a binary 0/1 random variable that is observed only when another binary 0/1 random variable Z is equal to 1. So, we observe $(Y|Z = 1, Z)$. We are interested in $P(Y = 1)$. These and similar problems are put together and worked out in Manski (2003).

Partial Identification Approach: The parameter of interest here is $P(y = 1)$, and we require the characterization of information about this parameter contained in the observables (and the assumptions). It is easy to see that, without additional assumptions, this frequency is not point identified in general. The issue is that the data alone contains no information about $Y|Z = 0$, and so, without this, the above problem consists of a class of models, each of which corresponds to a complete model with an assumed value for $Y|Z = 0$. One approach that is common in the literature is to *add* a model for the unobservable $Y|Z = 0$ that is based on some prior scientific or economic convictions. These assumptions are not testable. Another complementary approach is to examine the model above's information about the parameter of interest without further assumptions. We will start with this latter question, and discuss the parametric approach in a second step.

1. The identified features of the problem consist of the set of values for $P(y = 1)$ that are consistent with the assumptions (and the sampling process). Given the maintained assumptions, the **sharp identified set** or the **identified set** is the set of parameters that exhaust all the information. The sharp identified set is also referred to as the **sharp set, or the identification region.** Here, this set would be

   $$\Theta_I = \{p \in [0,1] : p = P(Y = 1|Z = 1)P(Z = 1) + qP(Z = 0), \text{ for some } q \in [0,1]\}$$

   Θ_I can also be characterized as $\Theta_I = \big\{p \in \big[P(Y = 1|Z = 1)P(Z = 1), P(Y = 1|Z = 1)P(Z = 1) + P(Z = 0)\big]\big\}$ or

   $$P(Y = 1|Z = 1)P(Z = 1) \leq p \leq P(Y = 1|Z = 1)P(Z = 1) + P(Z = 0)$$

   The sharp set can be obtained by parametrizing the likelihood of the observed data as a function of the parameter of interest and other nuisance parameters.

The log likelihood is

$$Q(p,q) \equiv l(p,q) = E\big[y \log \frac{p-(1-P(Z=1))q}{P(Z=1)} + (1-y)\log \frac{P(Z=1)-p+(1-P(Z=1))q}{P(Z=1)}\big]$$

The argmax of the above likelihood is the set of parameters $(p,q) \in [0,1]^2$ that satisfy

$$P(Y=1|Z=1) = \frac{p-q(1-P(Z=1))}{P(Z=1)}$$

This is the identified set Θ_I.

2. Another common approach to the problem above, is to add assumptions. One assumption is $Y \perp Z$, which is called the independence assumption (this can also be conditional on some set of covariates). This assumption leads to point identification since it implies that $P(Y=1) = P(Y=1|Z=1)$. This assumption is motivated in situations where the scientist believes that the missingness of Y is not related to the value of Y. The important point is that the value for $P(Y=1)$ implied by this assumption certainly lies in Θ_I.

3.2 Missing Continuous Outcomes

These results above can be generalized to the case where Y is a continuous random variable with support on R and a strictly increasing distribution function $F(t) \equiv P(Y \leq t)$, which is the parameter of interest.

Partial Identification Approach: This is similar to the question above, and we analyze it similarly.

1. Without further assumptions, the identified set for this distribution is

$$\Theta_I = \{F(t) \in \mathrm{H}: \quad F(t) = P(Y \leq t|Z=1)P(Z=1) + K(t)P(Z=0), \text{ for some } K(t) \in \mathrm{H}\} \tag{3.2}$$

where H is the set of strictly increasing CDF's on the real line. Note that this set Θ_I lies within the set of strictly increasing distributions bounded above and below by $P(Y \leq t|Z=1)P(Z=1) + P(Z=0)$ and $P(Y \leq t|Z=1)P(Z=1)$ respectively. One can also characterize the identified set as a solution to an optimization problem.

2. The second approach is again to make prior restrictions similar to ones made above. An overall parametric approach would start with a parametric distribution for Y and tries to study the identification problem of its finite dimensional parameters using the truncated data. The ensuing distribution might or might not lie in Θ_I. In the latter scenario, the parametric model would be rejected.

3.2.1 Missing Outcomes and Treatment Effects

The above approach to identifying the distribution of an outcome is important, and is key in the literature on program evaluation which is typically interested in functionals of the joint distribution of two outcomes (Y_1, Y_2), where[10] *we observe one of the two outcomes*: we observe Y_1 when an observed random variable $Z = 1$, and Y_2 when $Z = 0$. Average treatment effects can be easily derived from the bounds above since these bounds above can be extended easily to bounds on the mean. So, here, we posit $F(t_1, t_2)$ the joint distribution of (Y_1, Y_2) as the parameter of interest.

Partial Identification Approach: Here, the problem is slightly more complicated since the sampling process is such that we do not observe both Y_1 and Y_2 for any unit. In addition, the marginal distribution of each is not point identified. As above, we study the information content of the model by first examining what can be learned without further assumptions.

1. The joint distribution of (Y_1, Y_2) can be written as

$$P(Y_1 \leq t_1, Y_2 \leq t_2) = F(t_1, t_2) = C(F_1(t_1), F_2(t_2))$$

where F_1 and F_2 are the marginals, and $C(.,.)$ is a copula, a bivariate distribution with uniform marginals. Hence, the sharp identified set on the *joint* distribution of (Y_1, Y_2) can be written as the argmin of the objective function:

$$Q(F) = \underset{C(.,.),K_1,K_2}{argmin}\ Q(F, C, K_1, K_2) \tag{3.3}$$

where

$$Q(F, C, K_1, K_2) = \mathbf{d}\left[F(t_1, t_2), C\big(F_1(t_1|Z = 1)P_1 + K_1(t_1)P_0, F_2(t_2|Z = 0)P_0 + K_2(t_2)P_1\big)\right]$$

and $P_1 = P(Z = 1)$, and $P_0 = P(Z = 0)$. The above optimization is complicated and can be hard to estimate with finite sample sizes. A slightly less cumbersome description of the identified set can be obtained by exploiting the Frechet bounds on joint distributions. There, we know that

$$\max(F_1(t_1) + F_2(t_2) - 1, 0) \leq F(t_1, t_2) \leq \min(F_1(t_1), F_2(t_2))$$

The LHS and RHS above are then bounded as follows

$$\max(F_1'(t_1|Z = 1)P(Z = 1) + F_2'(t_2|Z = 0)P(Z = 0) - 1, 0) \leq \max(F_1(t_1) + F_2(t_2) - 1, 0)$$

$$\min(F_1(t_1), F_2(t_2)) \leq \min\big(F_1(t_1|Z = 1)P(Z = 1) + P(Z = 0), F_2(t_2|Z = 0)P(Z = 0) + P(Z = 1)\big)$$

where F_1' and F_2' are the distribution function that are equal to F_1 and F_2 except for a jump at "infinity" of size $P(Z = 1)$ and $P(Z = 0)$ respectively.

[10] See the recent work of Fan and Park (2009) that derived bounds on functionals of $Y_1 - Y_2$.

2. As above, to identify *marginal* treatment effects, a common approach is to invoke (conditional) independence[11] restrictions as in $(Y_1, Y_2 \perp Z)$, which point identifies the average treatment effects for example, or any parameters that require only information of the marginals. To identify the joint, two approaches in econometrics are used. The first approach is based on a fully parametric model that is guided by an underlying economic model. See for example the classic works of Gronau (1974), Heckman (1974), and Rosen (1974). This approach is useful in providing a tight link between empirical work and theory. It allows researchers to conduct policy analysis and extrapolate off the support of the data. Results from these exercises are interpreted within this model-world. However, it is understood that this approach suffers from a potential lack of robustness when one tries to applies its results to a broader context (More on sensitivity analysis in parametric models in section 3.3 below).

   The second less parametric approach in economics has been one that exploits another set of assumptions based on exclusion restrictions and support conditions. See for example Heckman and Honoré (1990), Ahn and Powell (1993) among others. These semiparametric approaches rely on exclusion restrictions and/or support conditions on a set of regressors to *point identify* the parameter of interest. These approaches are interesting and useful and should be complementary to the one in 1) above, especially. Comparing the various sets of estimates obtained from these various models would be useful. Even though identifying joint distributions as in (3.3) is complicated, economists should not trade convenience and simplicity at the expense of conviction and sensitivity analysis[12].

Remark 3.1 *Inference in partially identified models is tied to inference in m-estimation problems with nuisance parameters. Those latter parameters are generally not of intrinsic or essential value to the analysis, but do create great problems for inference (especially for constructing confidence regions with finite sample sizes). Hence, inference in partially identified models, mathematically, is one of characterizing the argmin of an objective function.*

[11] This is the assumption that basically legitimizes the "correlation implies causation" interpretation of mean regression.

[12] To identify the joint, Peterson (1976) showed that in a competing risks model any joint distribution of outcomes can be rationalized by one in which the outcomes are independent. He further derived bounds on the joint distribution based on the competing risks model. The statistics and biostatistics literatures seem to have largely taken the Peterson results as a justification for analyzing competing risks models under independent risks. This is unwarranted, and should be reconsidered, as independence is not a reasonable assumption in most settings.

3.2.2 Inference in a linear model with interval data

The class of models we considered above is largely nonparametric. Now, we maintain the assumption that the conditional mean of Y is linear in X, a vector of regressors, i.e., $E[Y|X] = X'\beta$. Also, the outcome Y is censored in a special way. It is interval measured, i.e., we do not observe Y, but rather, we observe $[Y_1, Y_2]$ such that $P(Y \in [Y_1, Y_2]) = 1$ and the parameter of interest here is the finite dimensional parameter β. So the main maintained assumption is that the latent conditional mean of Y given X is linear.

Partial Identification Approach: There are a number of ways to go about the identification analysis of β. I group them under two main avenues.

1. First, we approach the problem without any added assumptions about where Y lies within $[Y_1, Y_2]$, and so the amount of information about β can be analyzed as follows. Here, the mechanics of characterizing the identified set can vary. For example, let

$$Y = Y_1\lambda + Y_2(1 - \lambda)$$

where λ is a random variable with distribution on $[0, 1]$. Then, the identified set for β can be written as the argmin of the following objective function:

$$\begin{aligned} \Theta_I &= \underset{b}{argmin}\, Q(b) \\ &= \underset{b}{argmin} \left[\underset{F_\lambda}{argmin}\ E_x(E[y(\lambda)|x] - x'b)^2 \right] \end{aligned} \tag{3.4}$$

where

$$E[y(\lambda)|x] = E[Y_1\lambda + Y_2(1-\lambda)|x] = E[E[\lambda|Y_1, Y_2, x](Y_1 - Y_2)|x] + E[Y_2|x]$$

and F_λ is the expectation of λ conditional on (Y_1, Y_2, x). There are other ways to characterize Θ_I by exploiting monotonicity in the problem. For example, it is easy to show that the identified set can also be written as a **moment inequality** model:

$$\Theta_I = \{b \in \mathrm{R}^k : E[Y_1|x] \le x'b \le E[Y_2|x]\}$$

which in turns can be written as the argmin of the objective function

$$Q(b) = E[w(x)(E[Y_1|x - x'b)_+^2 + (E[Y_2|x] - x'b)_-^2])]$$

where $w(.)$ is a nonnegative weight function, $(a)_+ = a1[a \ge 0]$, and $(a)_- = a1[a \le 0]$. This is the modified minimum distance approach introduced in Manski and Tamer (2002). All the above approaches will deliver Θ_I.

2. Another identification strategy is modelling the relationship between Y and (Y_1, Y_2) parametrically, using a link function. For example, assume that $Y = g(Y_1, Y_2, \theta)$ where $g()$ is known and θ is an unknown (nuisance) parameter and where $g(Y_1, Y_2, \theta)$ lies between $[Y_1, Y_2]$. It is possible now that both θ and β are point identified, and so, this would provide a simple and convenient approach to inference in regressions with interval data. As usual, results and conclusions in this approach should be compared to ones obtained in 1). A sensitivity analysis is possible and practical in this model.

Remark 3.2 *The nature of the identification problem in the linear model is different from the above cases in that the analysis was done under the maintained assumption that the conditional mean of Y is linear in X, which is untestable*[13] *generally given the censoring of the outcome Y. Why were we comfortable with the linearity assumption? We need not be. In fact, if our parameter of interest is $E[Y|X]$, then it is obvious that $E[Y|X]$ lies between $E[Y_1|X]$ and $E[Y_2|X]$. Linearity though of the conditional expectation is widely used in empirical work, and least squares for example can be interpreted, absence censoring, as the best linear approximation of this conditional mean function under square loss. If the parameter of interest were the best linear approximation to the (latent) conditional expectation $E[Y|X]$ under squared loss , i.e., in the case where the linearity assumption is not necessarily true, then, Θ_I above is a subset of the identified set. The identified set in this case is derived in Ponomareva and Tamer (2009). A full analysis of identification in this model depends on what the parameter of interest is and what the purpose of the analysis is. For example, the nonparametric no assumption bounds might be uninformative for example if one is interested in the density of $Y|X$ (as opposed to the conditional mean). But, these no-assumption bounds are informative if one is interested in the conditional distribution function of $Y|X$.*

Researchers have exploited the above approaches to inference in empirical work. I will only highlight two examples. Haile and Tamer (2003) study the problem of inference in English auctions, when data reveal bids from a set of independent auctions. The object of interest in these independent private values model of auctions is usually the underlying distribution of valuations. The question becomes one of linking the observed bids from an auction to the underlying valuations. The authors make two weak assumptions to analyze this inferential problem. They first assume that the winning bid is an upper bound on all losing valuations (within the minimum bid increment), and second, that no bidder is willing to bid above their valuation. These two

[13]It is certainly true that if the identified set in (3.4 is empty, then the linearity can be rejected. But, on the other hand, if this identified set is nonempty then it does not follow that the conditional mean must be linear. Linearity is a sufficient condition for the identified set to be nonempty, but is not necessary.

assumptions along with the independence of valuation within an auction imply non-trivial upper and lower bounds on each valuation within every auction. These bounds are computed by exploiting natural and well known properties of order statistics. In this setup, the authors are able to characterize information about the bid distribution under weak *necessary conditions* for equilibria in a *set* of auction models, and hence the inferential result that is based on these assumptions is consistent with all models within this set. This is in contrast to say a particular parametric model linking bids to valuations that is based on one particular model of auctions. In addition, to conduct policy experiments, one is interested in the optimal reserve price. The authors show that it is possible to place bounds on this reserve price with only the weak assumptions on behavior above. This is all done in a nonparametric framework, and where only weak behavioral assumptions are imposed motivated from the realities of auction data. Finally, to accommodate auction heterogeneity, if we assume that the conditional distribution of valuations is linear in some vector of auctions and/or bidder characteristics, the framework now fits under the interval data example described above. The paper then applies these bounds to data from U.S. Forest Service timber auctions, focusing on reserve price policy[14].

Another interesting empirical example that implemented partial identification ideas to deal with the important selection problem into the labor force using the distribution of wages in the UK is the study by Blundell, Gosling, Ichimura, and Meghir (2007). They show that worst case bounds on the wage distribution allowing for non random selection into the labor force are informative. The authors then use a set of assumptions, motivated by economic theory to narrow these bounds further. They find for example using the partial identification approach that there is evidence of increases in the relative wages of women. We next discuss another example where the partial identification approach is fruitful.

3.3 Example 2: Identification in a 2×2 entry game

In this example, we illustrate the partial identification approach to analyzing the identified features in a simple 2×2 entry game. A discrete entry game is an economics model in which two players (firms, individuals, entities, etc) decide to "enter" or not and their decision is interdependent: one player's action impacts the other's utility or payoff and vice versa. In an entry game, if a player decides not to enter, that player earns zero profits.

What are we interested in: In empirical setting with interactions, a model of strategic behavior is used, where restrictions on player information, behavior, and equilibrium

[14]For another partial identification approach to inference with auction data, see the recent work of Tang (2008).

are crucial to determining their actions. An essential part of any identification analysis in these settings is to derive the link between the data and the underlying structure given the behavioral assumptions made. For instance, the data has information on who is in the market, and so does observing $(1,1)$ mean that a duopoly is necessarily a pure strategy equilibrium of the game? What if one allows some form of rational play, but not full equilibrium, how would that affect this link? These are important questions that need to be addressed in any identification study in these settings.

With data on entry in a cross section of markets a parameter of interest can be the probability that player 1 enters the market given that player 2 is in the market, and compare that to the probability that player 1 is in the market when player 2 is not. This is a treatment effect like counterfactual that is of interest since we do not observe say what player 1 would have done had 2 not entered in markets where 2 is in. Other parameters of interest include variable profits, and the joint distribution of fixed costs. So, if researchers are interested in answering within a model questions related to policy changes, then a more parametric approach can be used in which links between profits, fixed costs and demand are more transparent and manipulable, and prediction off the support is possible.

I will start with a parametric version of the game and show how a partial identification approach to inference allows one to study the robustness of the information provided to certain assumptions that researchers have more disagreements about, such as equilibrium selection mechanisms. Mechanically, the partial identification approach portrays the sensitivity of information about the parameter of interest when we relax the non-plausible assumptions and allow that part of the model to vary in its logical domain. I also describe a nonparametric approach to the game in which worst case bounds are derived[15].

Consider the following bivariate game where we observe a random sample of ob-

Table 1: Bivariate Discrete Game

		Player 2: 0	Player 2: 1
Player 1	0	0,0	0,$X_2'\beta_2+\epsilon_2$
	1	$X_1'\beta_1+\epsilon_1$,0	$X_1'\beta_1+\Delta_1+\epsilon_1,\ X_2'\beta_2+\Delta_2+\epsilon_2$

[15]The empirical literature on estimation in games is rich and diverse. See for example Bresnahan and Reiss (1991), Berry (1992), Tamer (2003), Ciliberto and Tamer (2009), Bajari, Hong, and Ryan (2005), Aradillas-Lopez (2005), Aradillas-Lopez and Tamer (2008), Sweeting (2004), Pakes, Porter, Ho, and Ishii (2005), and Grieco (2009).

servations $(Y_{1i}, Y_{2i}, X_{1i}, X_{2i})$ for $i = 1, \ldots, N$, and Y_{li} is the binary 0/1 outcome for firm l in market i, where $l = 1, 2$. To abstract away from statistical issues and focus on the identification question of what does knowledge of the joint distribution of (Y_1, Y_2, X_1, X_2) tell us about the parameter of interest, we assume that this distribution is known. Also, and throughout this section, we will assume that the players know (ϵ_1, ϵ_2) (and the X's) but the econometrician does not observe the ϵ's. It is possible to relax this complete information assumption[16]. We also assume throughout that the Δ's are negative since duopoly profits are lower than monopoly profits.

Assumptions, Parameters of Interest and Partial Identification: Here, we consider sets of assumptions that can are made on the model here and relate those to the parameter of interest. The list is in decreasing order starting with a fully parametric model, and ending with a nonparametric model.

1-Parametric Model One approach to the above problem is to assume that (ϵ_1, ϵ_2) have a parametric distribution F_Ω that is known up to Ω, a finite dimensional parameter, and is independent of the X's. We also assume that the players are playing Nash Equilibrium. One final issue to deal with is that the game above admits multiple equilibria, i.e., fixing values for all exogenous variables (including the unobserved ones), the model predicts multiple outcomes in some cases. See Tamer (2003) and Ciliberto and Tamer (2009) for more on this. So, even with these parametric assumptions made, the model structure is incomplete and we need to model the selection function to be able to obtain a complete likelihood. This likelihood (or the choice probabilities predicted by the model) depends on $(Y_1, Y_2, X_1, X_2, \theta, S)$ where $\theta = (\beta_1, \beta_2, \Delta_1, \Delta_2, \Omega)$. The function S, is a probability function that picks one of the equilibrias in regions of multiplicity. It is a function of both the ϵ's and the X's in its most general form. Economists have little information about S. In fact, it is difficult to think of future data collection that would contain information that would allow us to consistently estimate this function. On the other hand, linearity of the systematic part of utility is not as problematic since this represents variable profits which is a function of demand and economists have information about the shape of demand (perhaps from other data sources). The joint distribution F_Ω, along with the independence restriction are more crucial. However, in the discrete choice literature, there has been a lot of work in the single agent case relaxing these assumptions. So, as a first step, one would like to study the identified features of the model -θ- and examine whether these inferences are

[16]See the work of Grieco (2009) who assumes that the players observe part of ϵ and hence considers two errors, one observed by all the players, and another observed by only one player. Both errors remain unobserved by the econometrician.

sensitive to the specification of S. The identified set can be defined as

$$\Theta_I = \underset{\theta}{argmin}\big(\underset{S}{argmin}\ -EL(Y_1, Y_2, X_1, X_2, \theta, S)\big)$$

where $L(.)$ is the likelihood of the model (for the exact form of this likelihood, see Berry and Tamer (2006)). Generally, inference about the parameter θ is rendered difficult with the presence of the function S. It is possible that the likelihood above is not point identified, i.e., there exists $(S_1, \theta_1) \neq (S_2, \theta_2)$ such that

$$EL(Y_1, Y_2, X_1, X_2, \theta_1, S_1)\big) = EL(Y_1, Y_2, X_1, X_2, \theta_2, S_2)\big) = EL_0$$

where EL_0 is the true likelihood. One can use recently developed methods based on random sets to estimate Θ_I directly using techniques developed by Beresteanu, Molinari, and Molchanov (2008) and Galichon and Henry (2009). So, this approach to inference embeds sensitivity analysis within the specification of the likelihood, and indexes the class of models by the nuisance function S. Estimates of Θ_I obtained will take into account the effect of the nuisance parameter, and hence the identified set contains the set of parameters θ from the various models for S that are consistent with the model and the data. So, the partial identification approach here is geared towards the sensitivity of our inferences with respect to a key assumption on the selection function. Another assumption whose identifying power can be examined is the routinely made behavioral assumption that the players are playing a Nash equilibrium. Analysts can examine the identifying power of this assumption by maintaining only that players be rational. See Aradillas-Lopez and Tamer (2008) for more on this where they study the identifying power of rationalizable strategies.

2-Nonparametric Model: Now, suppose one wants to answer the counterfactual probability contrast while maintaining minimal assumptions as in Kline and Tamer (2009). This is akin to examining the identified feature of the following entry game where the π's in Table 2 are random variables that arbitrarily distributed, and where

Table 2: Bivariate Discrete Game

		Player 2: 0	Player 2: 1
Player 1	0	0,0	0,π_0^2
	1	π_0^1,0	π_1^1, π_1^2

one observes an iid data set from a cross section of markets. The data consists of pairs of outcomes (a_{1i}, a_{2i}) from market i, where $a_{ji} \in \{0, 1\}$ for $j = 1, 2$. The object of interest is to learn $P(Y_1(1) = 1)$ and $P(Y_2(1) = 1)$ using our knowledge of the data frequencies. The function $Y_1(1)$ is player 1's best response to player 2 entering the market, and similarly for $Y_2(1)$. The function $Y_1(.)$ can be considered a treatment response function for player 1, where the treatment is whether player 2 is in the market or not. As in the parametric game above, the link between the observed outcomes and the underlying best responses is complicated because of the presence of multiple equilibria and mixed strategies, both common in these setups. So, how do we proceed? Without making any assumptions on the π's, it is easy to show[17] that under complete information, $P(Y_1(1))$ is equal to

$$\begin{aligned} P(y_1(1) = 1) &= P(\pi_1^1 \geq 0) \\ &= P(\pi_1^1 \geq 0|(1,1))_{(1)} P(1,1) + P(\pi_1^1 \geq 0|(0,0))P(0,0) \\ &\quad + P(\pi_1^1 \geq 0|(1,0)P(1,0) + P(\pi_1^1 \geq 0|(0,1))P(0,1) \end{aligned}$$

where the unconditional probabilities are identified from the data (i.e., $P(1,1) = P(a_1 = 1, a_2 = 1)$). The object of interest $P(Y_1(1))$ is not point identified, but rather under rationality (which is weaker than Nash,) we get

$$P(Y_1(1) = 1) \in [0, P(Y_1 = 1)]$$

while if we assume that the players are playing only pure strategies, then the sharp bounds become

$$P(Y_1(1) = 1) \in [P(Y_1 = 1, Y_2 = 1), P(Y_1 = 1)]$$

So, it is possible to learn about these counterfactual probabilities without making assumptions on the forms of the profit function. Which way to proceed, a top down approach or a bottom up approach depends on what the empirical researcher is interested in.

As we can see, the partial identification approach to inference in this section enriches and strengthens the exercise, allowing the researcher to really explore the source of the results and what influences his/her estimates. This approach to inference is not a substitute for economic modeling, but rather, a disciplining tool that measures the cost of information with respect to the assumptions that are made.

4 Statistical Inference

The identification analysis above supposes that we have access to an arbitrarily large sample size. This was done to focus on the problem of identification, or the question of what we can learn under ideal conditions. Practically, empirical work deals with data

[17]See Kline and Tamer (2009) for more on this.

with a finite sample size and hence one needs to account for statistical uncertainty when conducting inference about parameters. The general problem that arises in partially identified models is inference on the set of minimizers of an objective function, and more crucially allowing for cases in which this set is non-singleton. The literature on inference is involved, and so here, we highlight the important issues that come up, describe general methods for inference, and point to relevant literatures for more detailed results. We first study the general inference problem with a generic objective function. Then we analyze a case in which the identified set can be written as the set of parameters that satisfy a vector of *moment inequalities*. We will only discuss cases in which θ belongs to some finite dimensional space. Partial identification approaches in semiparametric models in which the parameter of interest is infinite dimensional is a developing area of research.

Is the identified set the parameter of interest? It is certainly interesting to conduct inference on the set of minimizers of an objective function. This set represents generally the values of the parameter that are consistent with the maintained assumptions. Each parameter within this set is related to a complete model. This is particularly useful in parametric cases in which sets of parametric models are considered, each of these models correspond to a different parameter that belongs to the identified set.
Another parameter of interest is the "true" parameter which generated the data, θ^*. This parameter is not point identified, but all we know is that $\theta^* \in \Theta_I$. Inference on the (potentially) **non-point-identified parameter** takes the view that the unique data generating process lies in the basin of the class of models under consideration. In general, both the identified set and the true parameter are objects of interest and statistical inference on either face delicate and subtle problems that are different than ones faced in models with point identified parameters.

The general framework for statistical inference is one of M-estimation in cases where the objective function $Q(.)$ admits a non-unique minimum. The identified set of interest can be expressed as

$$\Theta_I = \operatorname*{argmin}_{\theta} Q(\theta) \tag{4.5}$$

where we assume for convenience that $Q(\theta) \geq 0$ for all θ. We will first define a set estimator for Θ_I, and discuss consistency of this estimator in the Hausdorff distance. Consistency is naturally only worked out for the identified set. The formal results, and assumptions needed including regularity conditions are only referenced, and I focus on heuristic descriptions that are meant to give a flavor of statistical work in this area.

4.1 Consistency

The object of interest is Θ_I defined in (4.5) above, where the function $Q(.)$ must obey a set of conditions, such as lower semicontinuity and where there exists a well defined sample objective function $Q_n(.)$ that converges uniformly to $Q(.)$. See the exact conditions in Condition C.1 on page 1252 of CHT. The sample estimators we consider are sequences of properly defined *level sets* for the objective function $Q_n(.)$. Results on consistency of level sets in econometrics were first given in Manski and Tamer (2002) and sharpened in CHT, where rates of convergence were also provided under more general conditions. The estimator for Θ_I is the level set

$$C_n(c_n) = \{\theta : Q_n(\theta) \leq c_n\} \tag{4.6}$$

and we use the Hausdorff distance between sets to define consistency[18], i.e., a sequence of sets A_n converges to A is $d_H(A_n, A) = o_p(1)$ as $n \to \infty$ where n is the sample size. When $c_n \sim \frac{\ln n}{n}$ for example, it is shown in CHT that

$$d_H(C_n(\frac{\ln n}{n}), \Theta_I) = O_p(\sqrt{\max(\ln n, 1)/n})$$

which is close[19] to the $\sqrt{n}$ parametric rate[20].

The consistency theorem is not as useful with general objective functions since it is not clear in practice how one would choose c_n. In cases in which the boundary of the identified set is an explicit function that can be consistently estimated, then a consistent estimator of the identified set can easily be obtained by replacing the boundary by its sample analog. This is a common class of problem in which for example the identified set is an interval as in $\Theta_I = [\theta_1, \theta_2]$ where θ_1 and θ_2 can both be consistently estimated from the data.

4.2 Confidence Regions

It is important for empirical work to summarize sampling uncertainty due to small sample sizes using a confidence region. Typically, econometricians use large sample approximations to do that. I will highlight this approach here from a frequentist perspective[21]. The confidence regions for Θ_I and those for θ^* are different here. We will

[18] The Hausdorff distance betweeb A and B, $d_H(A, B) = \max[\sup_{a\in A} d(a, B), \sup_{b\in B} d(b, A)]$ where $d(a, B) = \inf_{b\in B} \|a - b\|\}$.

[19] For exact conditions such as existence of polynomial minorants on $Q_n(\theta)$, see CHT.

[20] In cases where the identified set Θ_I has a nonempty interior, or more precisely when any point on the boundary of Θ_I is arbitrarily close to some point on the interior, it is possible to set $c_n = 0$ and get the sharp rate of $\sqrt{n}$.

[21] For recent work on inference in partially identified model from a Bayesian perspective, see the recent work of Moon and Schorfheide (2009), and Liao and Jiang (2009).

first highlight a subsampling based empirical approach to construct these confidence regions. Subsampling is a resampling technique that is used to approximate large sample distributions. This will be general and can be used in identified sets defined as minimizers of an objective function. I will then discuss different approaches, some that have been shown to be more powerful than subsampling, for the class of problems where the identified set is defined through moment inequalities. Inference in these models is complicated due to the fact that there exist problems with nuisance parameters that are manifested through a subtle property, mainly that of uniformity of coverage with respect to all possible DGP's. This uniformity issue is described in more details below. Overall, the discussion is descriptive, but sufficiently detailed to allow one to obtain a snapshot of the kinds of work in this literature.

4.2.1 Confidence Regions for Identified Set using Subsampling

Consider the objective function $Q_n(.)$, where we are interested in constructing a confidence region for the set Θ_I. One approach, that we describe here, follows CHT and is based on subsampling. Another similar approach is detailed in Romano and Shaikh (2008). The set that we construct, similar to the consistent set, is a level set $C_n(c_n)$ where the c_n is chosen in a particular way. For instance, we consider the level set in (4.6). Consider B_n subsets[22] of size $b << n$. Then, 1) Let $c_{0n} \geq inf_\theta Q_n(\theta)$. Usually, one can set C_{0n} to be 10% higher than the infimum of Q_n. 2) Compute c_{1n} as the α-quantile of the sample $\hat{C}_{j,b} = \{ \underset{C_n(c_{0n})}{sup} Q_{j,b}(.) : \; j = 1, \ldots, B_n\}$ where $Q_{j,b}$ is the objective function evaluated at the jth subsample. 3) Repeat the step 2 above 3-4 times to get $\hat{c}$. Report $C_n(\hat{c})$ which is a valid confidence region, i.e., $P\{\Theta_I \subseteq C_n(\hat{c})\} = \alpha$, and a consistent set estimator. For the precise statement of the theorem, see CHT Theorem 3.3. See also Romano and Shaikh (2007).

The result above is powerful and applies to general objective functions, but the approach using subsampling relies on conditions on the objective function that need to be satisfied. For specific problems such as objective functions based on moment inequalities for example, it is possible to use a modified bootstrap procedure, or simulation methods to do inference on the set Θ_I. See for example Remarks 4.5 and 4.6 in CHT. Finally, for another approach to do inference on sets based on random set theory, see Beresteanu and Molinari (2008).

[22] In larger samples, confidence regions should be computed for a series of subsample sizes starting say with $b = n/5$. For more on subsample sizes, see Politis, Romano, and Wolf (1999).

4.2.2 Confidence Regions for Identified Parameter using Subsampling

Imbens and Manski (2004) considered inference on the identified parameter in simple setups. They argued that there is usually a unique parameter θ^* that is the "true" parameter of interest, even though this parameter is not point identified. Most importantly, confidence regions for θ^* are no larger that the confidence regions for the identified set. Here, we highlight an approach to obtaining a confidence regions C_n for θ^*, i.e., $P(\theta^* \in C_n) \geq \alpha$ as $n \to \infty$. Imbens and Manski (2004) also pointed out that there is a subtle but important property that comes up while constructing these confidence regions, mainly that one needs to ensure that these confidence regions are **uniform** in the DGP, i.e., that in cases where the underlying model is point identified, or close to point identified, the size of these confidence region is maintained. This is a problem similar to inference in models with parameters on the boundary where values of the true parameter near or on the boundary causes problems for standard large sample approximations. Uniformity problems arise typically in nonstandard problems in which the asymptotic distribution is a non differentiable function of the true parameter. The validity of subsampling methods in non-point identified models has been considered by CHT, Romano and Shaikh (2008) and especially Andrews and Guggenberger (2009b) and Andrews and Guggenberger (2009a).

The general idea of constructing a confidence interval for the "true" parameter θ^* is to exploit the duality between testability and confidence region. In essence, a confidence region is the set of parameters that cannot be rejected. So, this pointwise approach to constructing a confidence region for the "true" parameter considers every parameter in the parameter space and uses a criterion based test to determine whether one fails to reject the hypothesis of whether this parameter is the truth. The collection of all the parameters that cannot be rejected constitutes the confidence region. There are many approaches to do this in econometrics especially for models defined by *moment inequalities*: See for example CHT, Romano and Shaikh (2008), Andrews and Soares (2009), Bugni (2007), Canay (2007), Rosen (2008) among others. Andrews and Jia (2008) provide a way to construct an objective function based on moment inequalities that allows for confidence regions that are size valid but also optimal from some power criterion. See also Chiburis (2009). Here, we follow the approach in CHT and Romano and Shaikh (2008) in constructing a confidence region. These methods apply not just for moment inequalities, but in other general models that are based on minimizing an objective functions.

Start by choosing a parameter θ in the parameter space, and then we use the value of $nQ_n(\theta)$ as the test statistic under the null that $\theta^* = \theta$, where we add θ to the confidence region if $nQ_n(\theta) \leq \hat{c}(\theta)$, where $\hat{c}(\theta)$ is a critical value that we construct using subsampling[23]. We outline an approach for general objective functions, and then consider in details a simple example below.

We assume here that $Q(\theta) \geq 0$, for all θ, and $Q(\theta_0) = 0$. So, to test the hypothesis that $\theta_0 = \theta$, we compute the critical value, $c_n(\alpha, \theta)$ of the test statistic $nQ_n(\theta)$ as follows. Let

$$c_n(\alpha, \theta) = inf\{t : \frac{1}{B_n} \sum_i \{b\left(Q_{i,b}(\theta) - Q_n(\theta)\right) \leq t\} \geq \alpha\}$$

where we recenter here since recentering can lead to better finite sample values. The confidence region is then $\mathcal{C}_n = \{\theta \in \Theta : nQ_n(\theta) \leq c_n(\theta, \alpha)\}$. In empirical examples, it is preferable to redefine $Q_n(\theta)$ and use instead $Q_n'(\theta) = Q_n(\theta) - \inf_t Q_n(t)$ to ensure that the confidence region is non-empty.

4.3 Confidence regions in interval bounds: A Simple Example

Here, I highlight the inference approaches in the canonical example of a scalar parameter θ^* and where the identified set is

$$\Theta_I = [\theta_l, \theta_h] \tag{4.7}$$

and where θ_l and θ_h can be consistently estimated. This is a simple and important example covering cases in which the parameter of interest is scalar and where one is able to solve for the upper and lower bounds as functionals of the observed data distribution. We first start with a confidence interval that covers Θ_I with a prespecified probability α. Then, we highlight various approaches to construct intervals that cover the identified parameter θ^* is probability α.

4.3.1 Confidence Region for $\Theta_I = [\theta_l, \theta_h]$

Here, we are covering the interval (4.7) above, and so heuristically a confidence region would be a set of intervals. One way to do that is to use a joint confidence regions on the endpoints and map that into a confidence region for Θ_I, imposing the fact that the joint confidence region on the endpoints is under the constraint that it is an interval where the right endpoint is higher than the left one. One easy approach is to generate via the bootstrap a set of intervals and take the smallest interval that fits α% of the generated intervals within it. This was used for example in Horowitz and Manski

[23]For example conditions needed and statements of the theorems, see CHT Section 5, and Romano and Shaikh (2008) Section 3.2. CHT which also provided critical values constructed using bootstrap and simulations (See Theorem 5.2).

(2000) and other papers. Here, we follow our analysis above and derive an objective function that is minimized on Θ_I. One such objective function is

$$Q(\theta) = (\theta_l - \theta)_+^2 + (\theta_h - \theta)^2 - \tag{4.8}$$

where $(a)_+^2 = a^2 1[a > 0]$ and similarly for $(a)_-^2$. Notice that $Q(\theta) \geq 0$ and $Q(\theta) = 0$ if and only if $\theta \in \Theta_I$. Assuming that we have $\hat{\theta}_l$ and $\hat{\theta}_h$ such that

$$\begin{bmatrix} \hat{\theta}_l - \theta_l \\ \hat{\theta}_h - \theta_h \end{bmatrix} \rightarrow_d \mathcal{Z} = \begin{bmatrix} \mathcal{Z}_1 \\ \mathcal{Z}_2 \end{bmatrix}$$

where $\mathcal{Z}$ is a bivariate normal distribution with a strictly positive variance. As above, our confidence region will be a level set as in $C_n(c) = \{\theta : nQ_n(\theta) \leq c\}$ where

$$Q_n(\theta) = (\hat{\theta}_l - \theta)_+^2 + (\hat{\theta}_h - \theta)^2 - \tag{4.9}$$

We know that the event $\{\Theta_I \subseteq C_n(c)\}$ is equivalent to the event $sup_{\theta \in \Theta_I} nQ_n(\theta) \leq c$ and so the asymptotic behavior of $\sup_{\theta \in \Theta_I} nQ_n(\theta)$ is used to determine the coverage probability. In particular, it is easy to show that

$$\sup_{\theta \in \Theta_I} nQ_n(\theta) \rightarrow_d \max((\mathcal{Z}_1)_+^2, (\mathcal{Z}_2)_-^2)$$

So, one can obtain, c_α, the $\alpha-$quantile of the asymptotic distribution above via simulation to obtain the confidence region $C_n(c_\alpha)$.

4.3.2 Confidence region for $\theta^* \in [\theta_l, \theta_h]$:

Imbens and Manski (2004) argued that that one can report confidence region on the parameter $\theta^* \in [\theta_l, \theta_h]$. To build this confidence interval one can collect all the parameters θ that cannot be rejected under some appropriate test that they belong to Θ_I. There are many approaches to building such an interval. Here, as in above, we build these based on the simple objective function (4.8). The choice of the objective is relevant in moment inequality models and can impact the power of the test. For more on this, see Andrews and Jia (2008). An important issue that comes up in these settings is that of uniform consistency of the testing procedure which impacts the asymptotic behavior of the test statistic.

Uniformity: I will provide here a simple and heuristic discussion of the issue[24]. Some but not all procedures are uniformly consistent, which is a property stronger than consistency. Recall that a procedure is consistent if for any true null hypothesis the

[24] For a discussion of uniformity in the interval bounds setting, see Imbens and Manski (2004). For a thorough discussion of uniformity in these contexts and other, see Andrews and Guggenberger (2009a) and Andrews and Guggenberger (2009b).

rejection rate in repeated sampling is not much more than the nominal rejection rate, as long as the samples are at least of some minimal size. The exact definition of uniform consistency is technical and varies somewhat between papers, but at a minimum uniform consistency strengthens consistency to require that the same minimal sample size controls the rejection rate for all true null hypotheses. In addition, uniform consistency often requires also that the rejection rate be controlled across not only different true null hypotheses but also different data generating processes.

Uniform consistency is best understood in the context of a simple example. Consider again the moment inequality model with identified set $\Theta_I = \{\theta : \theta_l \leq \theta \leq \theta_h\}$. Suppose that we are interested in constructing a 95% confidence set for $\theta^* \in \Theta_I$. Consistency requires that, any $\theta \in [\theta_l, \theta_h]$, is an element of the confidence set with probability close to at least 95% in repeated sampling, as long as the samples are at least of some minimal size. The minimal sample size is allowed to depend on the exact value of the parameter θ, and therefore consistency does not rule out that for any sample size there are many elements of the identified set that are rejected by the procedure with very high probability. Uniform consistency, on the other hand, guarantees that there is one fixed minimal sample size such that any $\theta \in [\theta_l, \theta_h]$ is an element of the confidence set with probability close to at least 95%, as long as the samples are of at least that one fixed minimal size. In other words, the minimal sample size is no longer allowed to depend on the exact value of θ. In particular, uniform consistency will maintain the size of the confidence regions even when the DGP is such that $\theta_l = \theta^* = \theta_h$, i.e., θ^* is point identified. This requires some modification of the standard approach of constructing the confidence interval for a fized DGP. See for example Imbens and Manski (2004) for a discussion of a procedure that is consistent but not uniformly consistent across a family of data generating processes including point identification.

There are many ways to construct a confidence region that is uniformly consistent. We use an approximation to the asymptotic distribution of $Q_n(.)$ above where we can easily show that under the null (See CHT section 5) that

$$nQ_n(\theta) \to_d \mathcal{Q} = (\mathcal{Z}_1 + \xi_1(\theta))_+^2 + (\mathcal{Z}_2 + \xi_2(\theta))_-^2 \tag{4.10}$$

$\xi_1(\theta) = -\infty$ if $\theta_l < \theta$ and equal to zero otherwise, and $\xi_2 = +\infty$ if $\theta_h > \theta$ and equal to zero otherwise. The ξ's are parameter that cannot be estimated consistently. But, we can estimate the α-quantile of $\mathcal{Q}$, by simulating the distribution of the random variable $\mathcal{Q}_n$,

$$\mathcal{Q}_n = (\mathcal{Z}_1 + \xi_1^n(\theta))_+^2 + (\mathcal{Z}_2 + \xi_2^n(\theta))_-^2$$

where we can simulate the distribution of $(\mathcal{Z}_1, \mathcal{Z}_2)$ (using the bootstrap for example) and $\xi_1^n = -\infty$ if $\hat{\theta}_l + c_1\sqrt{\log n/n} < \theta$ and equal to zero otherwise for some positive constant c_1, and similarly for ξ_2^n. Other approaches to construct a confidence interval in this case can be used such as ones based on the the pseudo likelihood approach of Rosen (2008), the empirical likelihood as in Canay (2007), the bootstrap, as in Bugni (2007), or the GMS procedures introduced in Andrews and Soares (2009) and further refined in Andrews and Jia (2008). Subsampling based intervals can also be constructed as in CHT and Romano and Shaikh (2008).

5 Conclusion

The partial identification approach to inference in econometric models takes as its starting point a set of assumptions that define a model and the data in order to learn about a parameter of interest, which is a finite dimensional parameter or a function in a general space. This approach to identification clarifies what and how can one learn about this parameter, by first describing what can we hope to learn about this parameter with an infinite data set, and then sets out to characterize the statistical uncertainty with a finite sample (as opposed to knowledge of the population). The key distinguishing feature of this approach to inference is the view that identification is not a one or zero event and that instead of looking for point identification assumptions, researchers should understand the map between information and assumptions characterizing what can be learned under what sets of assumptions.

Economists have long put a value on models to gain insights, polish and discipline their communication, and examine what can happen within these toy economies under what policies. Data also play an important role in shedding lights on certain postulates, disciplining theories and informing policy. However, most models contain a subset of assumptions that are made based only on convenience - analytical or computational- such as functional forms or distributional assumptions. The choice of what assumptions to investigate is problem specific. But, in most cases, those are the ones where there is no widespread consensus and accord about their validity. The partial identification approach allows one to probe and scrutinize the importance of these assumptions by examining their effects on conclusions drawn about the parameter of interest. It quantifies the (old) view that sensitivity analysis is important.

There is a lot of work ahead. For example, some inference methods referenced above are computationally intensive -to construct level sets for example- and so advances in computational methods tailored towards these problems is essential. Also, inference in models in which the parameter of interest is infinite dimensional, and where this parameter is partially identified is also a very important area of research. A good step in that direction is the work of Santos (2008). On a broader level, theoretical work such as that of Gilstein and Leamer (1983) is certainly in the spirit of partial identification, but its practical usefulness has not been exploited. I conjecture that the reason for this is that this robustness approach implements what is theoretically attractive -collecting the estimates from a large set of models- but is practically challenging in a general setup. I believe one can take this Gilstein and Leamer (1983) vision to more specific problems, and make use of the body of work so far on inference procedures to implement it.

The hallmark of microeconometric work in the last thirty years has been concern with semiparametric models with its main motivation that of robustness against one class of assumptions or another. In the last two decades, partial identification, and the analysis of econometric models that are not necessarily point identified have entered the realm of what econometricians accept, think about and allow and this approach to inference has appeared in important empirical work. So, there is no better time for empirical economists to be clear about what inferences can be made with what assumptions. This will lead to a better empirical program, one that is clear and transparent, combining both the data and valid economic assumptions, which is exactly what is required from any serious scientific program.

References

Ahn, H., and J. Powell (1993): "Semiparametric Estimation of Censored Selection Models," *Journal of Econometrics*, 58, 3–29.

Andrews, D. W. K., and P. Guggenberger (2009a): "Hybrid and size-corrected subsampling methods," forthcoming, Econometrica.

——— (2009b): "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. Jia (2008): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," Working Paper.

Andrews, D. W. K., and G. Soares (2009): "Inference for Parameters Defined by Moment Inequalities using Generalize Moment Selection," forthcoming, Econometrica.

Aradillas-Lopez, A. (2005): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information," Working Paper, Princeton University.

Aradillas-Lopez, A., and E. Tamer (2008): "The Identification Power of Equilibrium in Simple Games," *Journal of Business and Economic Statistics*, 26, 261–283.

Bajari, P., H. Hong, and S. Ryan (2005): "Identification and Estimation of Discrete Games of Complete Information," Working Paper.

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

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

Berry, S. (1992): "Estimation of a model of entry in the airline industry," *Econometrica*, 60(4), 889–917.

Berry, S., and E. Tamer (2006): "Identification in Models of Oligopoly Entry," in *Advances in Economics and Econometrics. Ninth World Congress of the Econometric Society*, ed. by R. Blundell, W. Newey, and T. Persson, vol. II, pp. 46 – 85. Cambridge University Press.

Bielby, W., and R. Hauser (1977): *Annual Review of Sociology*3, 137–161.

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

Bollinger, C. (1996): "Bounding mean regressions when a binary regressor is mismeasured," *Journal of Econometrics*, 73(2), 387 – 399.

Bresnahan, T., and P. Reiss (1991): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48, 57—81.

Bugni, F. (2007): "Bootstrap Inference in Partially Identified Models," Working Paper, Northwestern University.

Canay, I. (2007): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," Northwestern Working Paper.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Chiburis, R. (2009): "Approximately Most Powerful Tests for Moment Inequalities," Working Paper.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," forthcoming, Econometrica.

Coombs, C. (1965): *Theory of Data*. John Wiley and Sons.

Cross, P., and C. F. Manski (2002): "Regressions, Short and Long," *Econometrica*, 70(1), 357–368.

Duncan, O., and B. Davis (1953): "An Alternative to Ecological Correlation," *American Sociological Review*, 18, 665–666.

Fan, Y., and S. S. Park (2009): "Sharp Bounds on the Distribution of Treatment Effects and their Statistical Inference," *Econometrics Theory, forthcoming.*

Frechet, M. (1951): "Sur Les Tableaux de Correlation Dont les Marges sont Données," *Annals de l'Université de Lyon A*, Serie 3, 53–77.

Frisch, R. (1934): *Statistical Confluence Analysis*. Publication No. 5, Oslo: University Insitute of Economics.

Fuller, W. (1987): *Measurement Error models*. Wiley.

Galichon, A., and M. Henry (2009): "Inference in models with multiple equilibria," Working Paper.

Gilstein, C. Z., and E. E. Leamer (1983): "Robust Sets of Regression Estimates," *Econometrica*, 51(2), 321–333.

Grieco, P. (2009): "Inference in Games with Complete and Incomplete Information Errors," Northwestern University Working Paper, in progress.

Gronau, R. (1974): "Wage Comparisons–A Selectivity Bias," *The Journal of Political Economy*, 82(6), 1119–1143.

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

Heckman, J. (1974): "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42(4), 679–694.

Heckman, J., and B. Honoré (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121–1149.

Heckman, J., and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*.

Heckman, J. J., and V. J. Hotz (1989): "Choosing Among Alternative Nonemperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association*, 84(408), 862–874.

Horowitz, J., and C. Manski (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and outcome data," *Journal of the American Statistical Association*, 95, 77–84.

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

Hotz, V. J., C. H. Mullin, and S. G. Sanders (1997): "Bounding Causal Effects Using Data From a Contaminated Natural Experiment: Analysis the Effects of Teenage Chilbearing," *The Review of Economic Studies*, 64(4), 575–603.

Imbens, G., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

Kline, B., and E. Tamer (2009): "Bounds for Best Response Functions in Binary Games," Working Paper.

Koopmans, T., and O. Reiersol (1950): "The identification of Structural Characteristics," *The Annals of Mathematical Statistics*, 21(2).

LaLonde, R. J. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *The American Economic Review*, 76(4), 604–620.

Leamer, E. E. (1981): "Is it a Demand Curve, Or Is It A Supply Curve? Partial Identification through Inequality Constraints," *The Review of Economics and Statistics*, 63(3), 319–327.

——— (1985): "Sensitivity Analyses Would Help," *The American Economic Review*, 75(3), 308–313.

Liao, Y., and W. Jiang (2009): "Bayesian Analysis in Moment Inequality Models," Working Paper, forthcoming, Annals of Statistics.

Manski, C. F. (1989): "Anatomy of the Selection Problem," *The Journal of Human Resources*, 24(3), 343–360.

——— (1990): "Nonparametric Bounds on Treatment Effects," *The American Economic Review*, 80, Papers and Proceedings, 319–323.

——— (1997): "Monotone Treatment Response," *Econometrica*, 65(6), 1311–1334.

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

——— (2007): *Identification for Prediction and Decision*. Harvard University Press, forthcoming.

Manski, C. F., and J. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68(4), 997–1010.

Manski, C. F., and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–547.

Marschak, J., and W. H. Andrews (1944): "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 812, 143–203.

Moon, R., and F. Schorfheide (2009): "Bayesian and Frequentist Inference in Partially Identified Models," Penn Working Paper.

Nelson, R. (1999): *An Introduction to Copulas*. Springer, New York.

Nerlove, M. (1965): *Estimation and Identification of Cobb-Douglas Production Functions*. Rand McNally & and Company, Chicago.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2005): "Moment Inequalities and Their Applications," Working Paper.

Peterson, A. V. (1976): "Bounds for a joint distribution function with fixed subdistribution functions: applications to competing risks," *Proc. National Acad. Sci. USA*, 73, 11–13.

Phillips, P. C. (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5(2), 181–240.

Politis, D., J. Romano, and M. Wolf (1999): *Subsampling*, Springer Series in Statistics. Springer:New York.

Ponomareva, M., and E. Tamer (2009): "Misspecification in Moment Inequality Models: Back to Moment Equalities?," Working Paper, Northwestern University.

Romano, J., and A. Shaikh (2007): "Inference for the Identified Set in Partially Identified Econometric Models," Working Paper.

——— (2008): "Inference for the Identified Set in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138(9), 2786–2807.

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

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82, 34–55.

Santos, A. (2008): "Inference in Nonparametric Instrumental Variables with Partial Identification," Working Paper.

Stoye, J. (2009): "More on Confidence Regions for Partially Identified Parameters," forthcoming, Econometrica.

Sweeting, A. (2004): "Coordination Games, Multiple Equilibria, and the Timing of Radio Commercials," Working Paper.

Tamer, E. (2003): "Incomplete Bivariate Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147–167.

Tang, X. (2008): "Bounds on the Counterfactual Distribution of Revenues in Auctions with Reserve Prices," UPenn Working Paper.