# Partial Identification in Econometrics

Elie Tamer

Department of Economics, Northwestern University, Evanston, Illinois 60208;
email: tamer@northwestern.edu

## Key Words

## 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. 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 of which are not. Of course, more assumptions beget more information, 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 to 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. Therefore, 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.

"*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 in line with Coombs' (1965) principle of buying information with assumptions.

Data alone can inform us only so much, 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—and some of which are esoteric and are needed only to complete a model. These latter assumptions 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 microeconometrics models. Economic theory is largely silent regarding the choice of the distribution of unobserved heterogeneity, and in many cases, the choice of this distribution is based on folklore, familiarity, and computational grounds.[1] This is especially important in nonlinear models in which mean independence assumptions are not sufficient. In these models, it is important to examine the role played by assumptions made on the heterogeneity distribution.

In the past 30 years, there have been reactions within the empirical literature against the fragility of inferences based on suspect assumptions. And so there is a movement, 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. Conversely, the less stylized

---

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

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 especially 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 past 20 years, researchers have begun 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 Koopmans & Reiersol's (1950) dictum whereby the specification of a model ought to be based on the underlying economics, prior knowledge (such as the linearity of variable cost that people have established for this industry), or other assumptions with universal or almost universal acceptance, but should not be geared primarily toward 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 & Reiersol 1950, pp. 169–70). Once the structure is specified, the model can either have no information about the parameter of interest, restrict the parameter of interest to a nontrivial 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. Conversely, 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, 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).

The 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, whereas 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 in which one is interested in inference on the (nontrivial) identified set, statistical methods geared toward estimating sets are required, and more importantly, methods to build confidence regions for sets are needed. I 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 beginning in the early 1930s. 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 the 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 in models of demand and supply in the 1920s and 1930s. The classification of variables as exogenous and endogenous and the recognition of 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 in which identification, or the uniqueness of the optimum of some objective function, is sometimes directly assumed (especially in nonlinear models).

It is not clear why the failure of point identification and the impact of partial identification were largely ignored in both econometrics and statistics before the 1990s. This is especially surprising as 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 oftentimes 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 "[i]t 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."

Below I review the literature on partial identification in econometrics. I begin 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 1990s. I focus on works in econometrics, but there have been similar ideas in other literatures.[2] I then describe some of the recent developments in the literature. We start here with Frisch's approach to confluence analysis.

### 2.1. Partial Identification and Frisch's True Regressions

Frisch (1934) 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

---

[2]For example, in their review of structural equation models in sociology, Bielby & Hauser (1977, see p. 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 & Andrews (1944), which we review below.

if one or more of these variates are almost perfectly correlated. To disentangle true rela-
tionships between variates that ultimately are used in regressions analysis, Frisch consid-
ered the case in which 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 studying correlation among random variables and allowing for
the possibility that this true correlation is among variables that are unobserved. These
unobserved random variables constitute the "true regression." Therefore, this is a classic
identification problem in which the observed data consist of the vector $x$, the object of
interest is the second moment matrix of the vector $x'$, and $x = x' + x''$. 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 dimension-
ality of the problem. As an example, 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.
Therefore, it is easy to see that $cov(x_1, x_2) = cov(x_1', x_2')$, for example. Assuming that the
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 $\beta_1$ 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" $\beta \in \left[ \frac{cov(x_1, x_2)}{var(x_2)}, \frac{var(x_1)}{cov(x_1, x_2)} \right]$. The end points 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." (Frisch 1934, p. 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 because 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 end points, 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 assump-
tions 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 Frisch's 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 a 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 the estimation of reliability ratios (e.g., see Fuller 1987). This typically yields point identification, but these data are not easy to obtain in typical economic surveys.

## 2.2. The Partial Identification Approach of Marschak & Andrews

In an important paper, Marschak & 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), 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 uses economic restrictions to derive bounds in a parametric model of supply. We review the approach using an example employed by Nerlove (1965, see chapter II) in his review of MA's paper.[3]

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

$$Y_{0f} - \alpha_1 X_{1f} - \alpha_2 X_{2f} = A + u_{0f},$$
$$\beta_0 X_{0f} - \beta_1 X_{1f} = B_1 + u_{1f}, \qquad (1)$$
$$\beta_0 X_{0f} - \beta_2 X_{2f} = B_2 + u_{2f},$$

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

The first line of Equation 1 is the production function (in logs), whereas the second and third lines represent the first-order conditions from profit maximizations, taking into account the demand function in inputs 1 and 2 (i.e., they deal with the more general case of imperfect competition), where a specific functional form for the demand, for example, is such that

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

where the $\eta_i$'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 influence the production of one factor, and not others.

MA take a different approach. With the assumption 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, $(\alpha_1, \alpha_2)$. Therefore, when there is perfect competition ($\beta$'s equal to 1), $(\alpha_1, \alpha_2)$ must lie in the

---

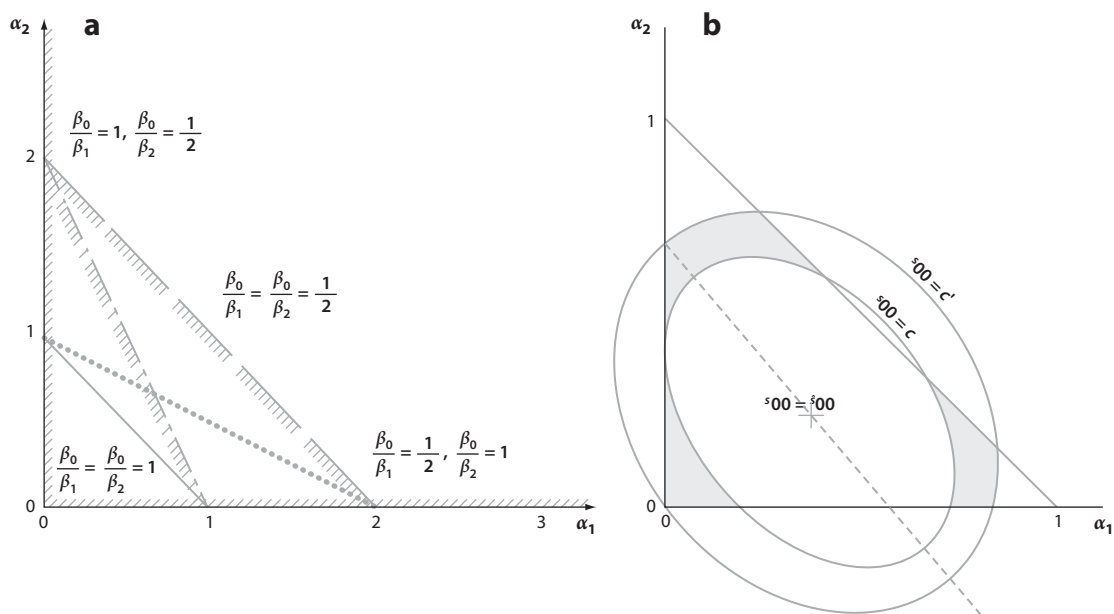[3]Nerlove's chapter reviewing the MA approach is titled "Partial-Identification: The Marshack-Andrews Approach."

**Figure 1**

Bounds on $(\alpha_1, \alpha_2)$. (*a*) The identification region under various conduct parameter assumptions. The functional form assumptions on the production function, the demand functions, and the optimization restrictions were used to derive these bounds. The smallest triangle [the line connecting (0,1) to (1,0)] is obtained under perfect competition. (*b*) Added restrictions (Marschak & Andrews 1944) on the size of the disturbance in the production function, shrinking the regions substantially. Figure adapted from Nerlove (1965).

triangle bounded by the line connecting (0,1) to (1,0). The different bounds on $(\alpha_1, \alpha_2)$ are graphed in **Figure 1**, where we see that the size of the identified set increases as we move away from perfect competition. The importance of the MA approach is recognizing that, although 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 Equation 1. Arguing that $u_0$ represents the technical efficiency of the firm (given as a deviation from the mean of efficiency of all firms), they assume that the so-called best firm can be no more than five times as good as the worst firm, for example, but also can be no less than four times as good as the worst firm, which leads to bounds on $var(u_0)$. These proposed bounds on this variance, along with the first line in Equation 1, can be used to narrow the set. For example, we see in **Figure 1***b* that there are substantial identification gains using these variance restrictions.

In MA's paper, we find a first example of the partial identification approach to inference in an economic model using assumptions that are motivated by the underlying economic problem. The results appear in diagrams in the paper, but it would be interesting to map them into estimates of the production function and to compare them with 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 revisits MA and Leontief and examines the problem of learning about elasticities of demand and supply in a 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), whereas a reverse regression provides an upper bound.

## 2.3. Other Influential Works

In addition to the above, the work of Fréchet (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. Fréchet 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

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

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 (for more details, and other results regarding the Fréchet bounds, see Nelson 1999). This is an important result that has been extended and applied in many areas.[4]

Other early work that contains partial identification ideas is Duncan and Davis's work (see Duncan & 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 that contains data on $(X, Z)$ but not $Y$. Duncan and Davis conducted partial identification analyses that were later studied more formally by Cross & Manski (2002).

## 2.4. Early Reaction to Partial Identification

As shown above, the ideas about partial identification have been around in econometrics since the 1930s, and the rationale for such an exercise could not have been more relevant or pertinent. However, these ideas were largely ignored in the empirical and theoretical fields in econometrics until the late 1980s. Researchers have been tentative at best in considering partial identification. The semiparametric literature in microeconometrics has certainly looked for a model that makes fewer assumptions, but one that (almost) always guarantees that these assumptions deliver point identification, regardless of whether these sufficient point identification conditions are suspect, or whether they hold in their data.

I speculate and offer two reasons about this incertitude regarding partial identification. One main motivation for empirical work in economics is to evaluate policies, with an important purpose of decision making. Therefore, 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's (1965) comment on MA's partial identification approach described above. Nerlove finds the "fundamental defect" of 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 by themselves, we are told, are not informative; the bounds are usually wide, so partial identification analysis by itself is not useful. However, one rationale for partial identification is exactly that: Although data alone are sometimes not useful, data and theory together are. Therefore,

---

[4]Given knowledge of the marginals, bounds on joint distribution can be used in competing risks models that are useful in economics, for example.

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 conclusions do not; this is the essence of partial identification.[5] The skepticism regarding multiple answers does not rest on solid, coherent arguments, so it is likely to 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 based on assumptions,[6] so it is not clear where the line should be drawn. Which assumptions are plausible and which are not? Is the choice of regressors and what regressors to condition on a more plausible assumption than conditional independence? These are valid questions and should be examined only in relation to a particular empirical problem. For example, in Frisch's 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, however, there is a large set of assumptions or models that may be considered plausible, and guarding against all does not seem feasible, both conceptually and computationally. Therefore, some recent work on partial identification has considered fully parametric models, and researchers 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 that the community of researchers considers 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, in which 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 (for similar ideas, see Heckman & Hotz 1989 and the rejoinder).[7]

## 2.5. Recent Literature

Given the above reservations, the econometrics literature, both theoretical and empirical, did not pay much attention to partial identification largely until the work of Manski and his collaborators, starting in the late 1980s. This work, and other work inspired by it, has revived and buttressed the partial identification view within empirical economics as a valid, coherent, and sensible approach to inference. Starting with work that analyzed the problem of self-selection into treatment from a partial identification perspective (Manski 1989, 1990), 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 about the usefulness of parametric models of selection (e.g., see

---

[5]Moreover, with the 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, as one should account for statistical uncertainty when reporting results, which makes the preference of point-like to set-like argument questionable.

[6]Here, one can argue that even raw data are driven by theory and that routine assumptions, such as being independent and identically distributed, are conditioned on some assumptions (see Coombs 1965 for more details).

[7]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.

LaLonde 1986). Manski 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 since made contributions to partial identification in many settings (for a clear synthesis of these results, see Manski 2003, 2007). A Manski-style approach to partial identification advocates the bottom-up approach in which first worst-case bounds are derived, gradually stronger assumptions are added, and their effects are analyzed. For example, Manski (1997) examines 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. In another example, Manski & Pepper (2000) revisit an important issue in labor economics: estimating wage regression as a function of schooling. The conceptual contribution of this 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, with higher values of one leading to higher values of the other. The paper shows how this type of assumption contains identifying power. Similar ideas have been studied recently by Nevo & Rosen (2009).

With regard to other partial identification works, Bollinger (1996) extends the Frisch bounds to cases with misclassification, and Hotz et al. (1997) study contaminated instruments as applied to the effect of teenage pregnancy on later outcomes, using the partial identification approach to contaminated models in Horowitz & Manski (1995). Tamer (2003) studies inference in entry models with multiple equilibria, which is generalized further to cover cases without assuming Nash equilibrium in Aradillas-Lopez & Tamer (2008). Manski & Tamer (2002) focus on inference on parameters in linear regressions with interval data on outcomes or a regressor using a partial identification approach, and Blundell et al. (2007) study bounds on wage distribution in the United Kingdom using worst-case and other more informative bounds. This is a just sample of recent works on partial identification on various problems.

## 2.6. Work on Inference

Most of the above literature has been concerned with identification as a question separate from statistical inference. One of the first papers in econometrics to tackle some aspects of inference in models with partially and nonidentified parameters was Phillips (1989), which proposes novel statistical inference methods in models in which some of the parameters are partially identified or nonidentified. The methods rely on rotations of the parameter space into one where all parameters are identified. Throughout the 1990s, most partially identified models were such that the boundary of the identified set could be derived explicitly as a functional of the observed data distribution. The typical bound on a parameter $\theta$, for example, would be of the form $F_1 \leq \theta \leq F_2$, and $(F_1, F_2)$ could be consistently estimated from the data. Moreover, a confidence region for $[F_1, F_2]$ was usually constructed by jointly bootstrapping the end points. Imbens & Manski (2004) then noticed that narrower confidence regions can be reported if one considered covering not the set, but the actual nonidentified parameter (these bounds were later refined in Stoye 2009). Conversely, and in the context of linear models with interval data, Manski & Tamer (2002) construct sets of parameters that are consistent in the Hausdorff metric to the argmin of a particular objective function under some conditions. This result generalizes the classical case of the consistency of m-estimators to cover the case in which the argmin of the population

objective function is not a singleton. Manski & 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 et al. (2007) provide a more general set of consistency results with rates of convergence for argmins of general objective functions. They also provide 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 by Romano & Shaikh (2008, 2010) using subsampling. Beresteanu & Molinari (2008) introduce tools from random set theory to study inference about the identified set. As many of the partially identified models can be represented by moment inequalities, there has also been a flurry of recent papers that develop inferential methods designed to cover moment inequalities (and equalities) (e.g., Bugni 2007, Canay 2007, Andrews & Jia 2008, Rosen 2008, Andrews & Soares 2009, Stoye 2009).

## 3. TWO EXAMPLES

This section provides 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, employing a more nonparametric approach as compared with the second example. The second example considers a fully parametric model in which a set of parametric assumptions is relaxed, and hence the identified features of a set of models are studied. In this section, we also assume away sampling issues and hence are concerned about the identification problem as distinct from statistical inference. Inference from finite sample sizes is 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 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 illustrate these ideas in two examples below.

### 3.1. Example 1: Missing Data

We examine a series of examples in which the partial identification approach is highlighted. We begin 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).

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 contain no information about $Y|Z = 0$, so, without this, the above problem consists of a class of models, each of which corresponds to a

complete model with an assumed distribution for $Y|Z = 0$. One common approach 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 above model's information about the parameter of interest without further assumptions. We start with this latter approach and discuss the parametric approach in a second step.

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

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

$\Theta_I$ can also be characterized as

$$\Theta_I = \{p \in [P(Y = 1|Z = 1)P(Z = 1), P(Y = 1|Z = 1)P(Z = 1) + P(Z = 0)]\} \text{ or}$$
$$P(Y = 1|Z = 1)P(Z = 1) \le p \le P(Y = 1|Z = 1)P(Z = 1) + P(Z = 0).$$

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

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

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 $\Theta_I$.

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 because it implies that $P(Y = 1) = P(Y = 1|Z = 1)$. This assumption is motivated in situations in which 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 $\Theta_I$.

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

Without further assumptions, the identified set for this distribution is

$$\Theta_I = \{F(t) \in \mathrm{H} : F(t) = P(Y \le t|Z = 1)P(Z = 1) + K(t)P(Z = 0), \text{ for some } K(t) \in \mathrm{H}\}, \quad (2)$$

where H is the set of strictly increasing distribution functions on the real line. Note that this set $\Theta_I$ lies within the set of strictly increasing distributions bounded above and below by $P(Y \le t|Z = 1)P(Z = 1) + P(Z = 0)$ and $P(Y \le t|Z = 1)P(Z = 1)$, respectively. One can also characterize the identified set as a solution to an optimization problem. The second approach is again to make prior restrictions similar to ones made above. An overall parametric approach starts 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 may or may not lie in $\Theta_I$. In the latter scenario, the parametric model would be rejected.

### 3.1.3. 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 $(Y_1, Y_2)$,[8] where we observe one of two outcomes: We observe $Y_1$ when an observed random variable $Z = 1$ and $Y_2$ when $Z = 0$. Average treatment effects can be derived easily from the bounds above as these bounds can be extended easily to bounds on the mean. Here, we posit $F(t_1, t_2)$, the joint distribution of $(Y_1, Y_2)$, as the parameter of interest.

Here, the problem is slightly more complicated because 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.

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

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

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:

$$Q(F) = \underset{C(.,.),K_1, K_2}{\text{argmin}} \; Q(F, C, K_1, K_2), \tag{3}$$

where

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

where $P_1 = P(Z = 1)$ 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 Fréchet bounds on joint distributions. For example,

$$F(t_1, t_2) = F(t_1, t_2|Z=1)P(Z=1) + F(t_1, t_2|Z=0)P(Z=0).$$

Now using Fréchet's bounds we have,

$$F(t_1, t_2|Z=1) \leq \min\left(F_1(t_1|Z=1), F_2(t_2|Z=1)\right) \leq F_1(t_1|Z=1)$$

and similarly,

$$F(t_1, t_2|Z=0) \leq F_2(t_2|Z=0),$$

and so,

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

and these bounds are sharp. A similar bound can be derived for the lower bound.

---

[8]See the recent work of Fan & Park (2009) that derives bounds on functionals of $Y_1 - Y_2$.

As above, to identify marginal treatment effects, a common approach is to invoke (conditional) independence[9] 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 on the marginals. To identify the joint distribution, researchers use two approaches in econometrics. The first approach is based on a fully parametric model that is guided by an underlying economic model (e.g., see 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 apply its results to a broader context (for more on sensitivity analysis in parametric models, see Section 3.2 below).

The second, less parametric, approach in economics exploits another set of assumptions based on exclusion restrictions and support conditions (e.g., see Heckman & Honoré 1990, Ahn & Powell 1993). 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 the above paragraph. Comparing the various sets of estimates obtained from these various models would be useful. Even though identifying joint distributions, as in Equation 3, is complicated, economists should not trade convenience and simplicity for conviction and sensitivity analysis.[10]

> **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, mathematically, inference in a partially identified model is similar to inference on the argmin of a well-defined objective function.

**3.1.4. Inference in a linear model with interval data.** The class of models 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., $E[Y|X] = X'\beta$. Moreover, 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 $\beta$. Therefore, the main maintained assumption is that the latent conditional mean of $Y$ given $X$ is linear.

There are a number of ways to go about the identification analysis of $\beta$. I group them under two main avenues below. First, we approach the problem without any added assumptions about where $Y$ lies within $[Y_1, Y_2]$, so the amount of information about $\beta$ can be analyzed as follows. Here, the mechanics of characterizing the identified set can vary. For example, let

---

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

[10]Peterson (1976) shows that, in a competing risks model, any joint distribution of outcomes can be rationalized by one in which the outcomes are independent. He further derives bounds on the joint distribution based on the competing risks model. The statistics and biostatistics literatures seem to have largely taken Peterson's 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.

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

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

$$\Theta_I = \underset{b}{\text{argmin }} Q(b),$$
$$= \underset{b}{\text{argmin}}\left[\underset{F_\lambda}{\text{argmin }} E_x(E[y(\lambda)|x] - x'b)^2\right], \tag{4}$$

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 $F_\lambda$ is the expectation of $\lambda$ conditional on $(Y_1, Y_2, x)$. There are other ways to characterize $\Theta_I$ 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 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 $(a)_- = a\,1[a \le 0]$. This is the modified minimum distance approach introduced in Manski & Tamer (2002). All the above approaches will deliver $\Theta_I$.

Another identification strategy is to model 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, $\theta$ is an unknown (nuisance) parameter, and $g(Y_1, Y_2, \theta)$ lies between $[Y_1, Y_2]$. It is possible now that both $\theta$ and $\beta$ are point identified, 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 with ones obtained above. 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 nonparametric cases in that the analysis was done under the maintained assumption that the conditional mean of $Y$ is linear in $X$, which is untestable[11] 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 $E[Y_2|X]$. However, linearity of the conditional expectation is widely used in empirical work, and linear 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 in which the linearity assumption is not necessarily true), then $\Theta_I$ above is a subset of the identified set (see Ponomareva &

---

[11]It is certainly true that if the identified set in Equation 4 is empty, then the linearity can be rejected. However, 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 it is not necessary.

Tamer 2009). A full analysis of identification in this model depends on the parameter of interest and the purpose of the analysis. For example, the non-parametric no-assumption bounds might be uninformative if one is interested in the density of $Y|X$ (as opposed to the conditional mean). However, 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. Here I highlight only two examples. Haile & 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 models 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 then they assume that no bidder is willing to bid above her valuation. These two assumptions, along with the independence of valuation within an auction, imply nontrivial 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 based on these assumptions is consistent with all models within this set. This contrasts with a particular parametric model linking bids to valuations based on one particular model of auctions, for example. 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, in which only weak behavioral assumptions are imposed, motivated by the realities of auction data. Finally, if we assume that the conditional distribution of valuations is linear in some vector of auctions and/or bidder characteristics to accommodate auction heterogeneity, the framework now fits under the interval data example described above. Haile & Tamer (2003) then apply these bounds to data from U.S. Forest Service timber auctions, focusing on reserve price policy (for another partial identification approach to inference with auction data, see Tang 2008).

Blundell et al. (2007) is another interesting empirical example that implements partial identification ideas to deal with the important selection problem into the labor force using the distribution of wages in the United Kingdom. The authors show that worst-case bounds on the wage distribution allowing for nonrandom selection into the labor force are informative. They then use a set of assumptions, motivated by economic theory, to narrow these bounds further. Using the partial identification approach, they find evidence of increases in the relative wages of women. We next discuss another example in which the partial identification approach is fruitful.

## 3.2. Example 2: Identification in a 2 × 2 Entry Game

This example illustrates 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 (e.g., firms, individuals, entities) decide to enter or not enter, 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.

**3.2.1. What are we interested in?** In empirical settings with interactions, a model of strategic behavior is used in which restrictions on player information, behavior, and equilibrium 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 provide information on who is in the market; therefore, 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, or if one allows for mixed strategies? How would that affect this link? These are important questions to address 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 we can compare that with 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 because we do not observe what player 1 would have done had player 2 not entered the market that player 2 is in. Other parameters of interest include variable profits and the joint distribution of fixed costs. Therefore, if researchers are interested in answering questions related to policy changes within a model, 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 begin 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 about which researchers have disagreed, such as equilibrium selection mechanisms. Mechanically, the partial identification approach portrays the sensitivity of information about the parameter of interest when we relax the nonplausible 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.[12]

Consider the following bivariate game in which we observe a random sample of observations $(Y_{1i}, Y_{2i}, X_{1i}, X_{2i})$ for $i = 1, \ldots, N$, and $Y_{li}$ is the binary 0/1 outcome for firm $l$ in market $i$, where $l = 1, 2$ (**Table 1**). To abstract away from statistical issues and focus on the identification question of what the knowledge of the joint distribution of $(Y_1, Y_2, X_1, X_2)$ tells us about the parameter of interest, we assume that this distribution is known. Moreover, throughout this section we assume that the players know $(\varepsilon_1, \varepsilon_2)$ (and the $X$'s) but that the econometrician does not observe the $\varepsilon$'s. It is possible to relax this complete information assumption.[13] We also assume throughout that the $\Delta$'s are negative, as duopoly profits are lower than monopoly profits.

Here we consider sets of assumptions that are made on the model and relate those to the parameter of interest. These models are listed in decreasing order, starting with a fully parametric model and ending with a nonparametric model.

**3.2.2. Parametric Model.** One approach to the above problem is to assume that $(\varepsilon_1, \varepsilon_2)$ have a parametric distribution $F_\Omega$ that is known up to $\Omega$, a finite-dimensional parameter,

---

[12]The empirical literature on estimation in games is rich and diverse (e.g., see Bresnahan & Reiss 1991, Berry 1992, Tamer 2003, Sweeting 2004, Aradillas-Lopez 2005, Bajari et al. 2005, Pakes et al. 2005, Aradillas-Lopez & Tamer 2008, Ciliberto & Tamer 2009, Grieco 2009).

[13]See the work of Grieco (2009), who assumes that the players observe part of $\varepsilon$ 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.

**Table 1** Bivariate discrete game

| | | Player 2 | |
|---|---|---|---|
| | | 0 | 1 |
| **Player 1** | 0 | 0,0 | $0, X_2'\beta_2 + \varepsilon_2$ |
| | 1 | $X_1'\beta_1 + \varepsilon_1, 0$ | $X_1'\beta_1 + \Delta_1 + \varepsilon_1, \; X_2'\beta_2 + \Delta_2 + \varepsilon_2$ |

and that 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 (for more details, see Tamer 2003, Ciliberto & Tamer 2009). Therefore, even with these parametric assumptions made, the model structure is incomplete, and we need to model the selection function 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 equilibria in regions of multiplicity. It is a function of both the $\varepsilon$'s and the *X*'s in its most general form. Economists have little information about *S*. In fact, it is difficult to think of a future data collection that would contain information that would allow us to consistently estimate this function. Conversely, the linearity of the systematic part of utility is not as problematic because this represents variable profits, which are a function of demand, and economists have information about the shape of demand (perhaps from other data sources). The joint distribution $F_\Omega$, along with the independence restriction, is more crucial. However, in the discrete choice literature, there has been a lot of work in the single-agent case relaxing these assumptions. Therefore, as a first step, one would like to study the identified features of the model—$\theta$—and examine whether these inferences are sensitive to the specification of *S*. The identified set can be defined as

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

where $L(.)$ is the likelihood of the model (for the exact form of this likelihood, see Berry & Tamer 2006). Generally, inference about the parameter $\theta$ 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. Therefore, 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 $\Theta_I$ can be obtained using sieve semiparametric likelihood methods as in Chen et. al. (2010) and Grieco (2009). This will take into account the effect of the nuisance parameter, and hence the identified set contains the set of parameters $\theta$ from the various models for *S* that are consistent with the model and the data. So the partial identification approach here is geared toward the sensitivity of our inferences with respect to a key assumption on the selection function. One can also use interesting recent work on random set theory to characterize the set $\Theta_I$ using techniques developed by Beresteanu et al. (2008) and Galichon & Henry (2009).

With regard to 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 [for more details, see Aradillas-Lopez & Tamer (2008), who study the identifying power of rationalizable strategies in entry games and in first price auctions].

### 3.2.3. Nonparametric Model.

Now suppose one wants to answer the counterfactual probability contrast while maintaining minimal assumptions, as in Kline & Tamer (2009). This is similar to examining the identified feature of the following entry game in which the $\pi$'s in **Table 2** are random variables that are arbitrarily distributed and one observes an independent and identically distributed data set from a cross section of markets. The data consist 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, in which the treatment is whether player 2 is in the market. 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. Then how do we proceed? Without making any assumptions on the $\pi$'s, it is easy to show 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., $P(1,1) = P(a_1 = 1, a_2 = 1)$] (for more details, see Kline & Tamer 2009). The object of interest $P(Y_1(1))$ is not point identified, but rather under rationality (which is weaker than Nash), we get

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

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

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

Therefore, 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 the interest of the empirical researcher.

As demonstrated above, 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 the influences of her estimates. This approach to inference is not a substitute for economic modeling, but rather is a disciplining tool that measures the cost of information with respect to the assumptions made.

**Table 2    Bivariate discrete game with general payoffs**

|  |  | Player 2 | |
|---|---|---|---|
|  |  | 0 | 1 |
| Player 1 | 0 | 0,0 | $0, \pi_0^2$ |
|  | 1 | $\pi_0^1, 0$ | $\pi_1^1, \pi_1^2$ |

Changes may still occur before final publication online and in print.

## 4. STATISTICAL INFERENCE

The identification analysis above supposes that we have access to an arbitrarily large sample size. This allows us to focus on the problem of identification, or the question of what we can learn under ideal conditions. Practically, empirical work deals with data 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, allows for cases in which this set is nonsingleton. The literature on inference is involved, so here we highlight some important issues, describe general methods for inference, and point to relevant papers 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 discuss only cases in which $\theta$ belongs to some finite-dimensional space. Partial identification approaches in semiparametric models in which the parameter of interest is infinite dimensional compose a developing area of research.

### 4.1. 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, with each of these models corresponding to a different parameter that belongs to the identified set.[14]

Another parameter of interest is the so-called true parameter that generated the data, $\theta^*$. 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 (DGP) lies in the basin of the class of models under consideration, which is not the case if the class of models is misspecified. In general, both the identified set and the true parameter are objects of interest, and in doing statistical inference on either, one faces delicate and subtle problems that are different than those faced in models with point-identified parameters.

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

$$\Theta_I = \operatorname*{argmin}_{\theta} \; Q(\theta), \tag{5}$$

where we assume for convenience that $Q(\theta) \geq 0$ for all $\theta$. We first define a set estimator for $\Theta_I$ and discuss the 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 sample of statistical work in this area.

---

[14]Also, in misspecified partially identified models, one can define the parameter of interest as the set of parameters that minimize a certain distance between the model set and the DGP (the data). There is not a true parameter in this set, but rather the parameters in this set are defined through the distance function, and hence inference on the whole set is more plausible.

## 4.2. Consistency

The object of interest is $\Theta_I$, defined in Equation 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 Chernozhukov et al. 2007, condition C.1, p. 1252). The sample estimators we consider are sequences of properly defined level sets for the objective function $Q_n(.)$. Results on the consistency of level sets in econometrics were first given by Manski & Tamer (2002) and Chernozhukov et al. (2007), who also provide rates of convergence under more general conditions. The estimator for $\Theta_I$ is the level set

$$C_n(c_n) = \{\theta : Q_n(\theta) \leq c_n\}, \tag{6}$$

and we use the Hausdorff distance between sets to define consistency. When $c_n \sim \dfrac{\ln n}{n}$, for example, Chernozhukov et al. (2007) show that

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

which is close[15] to the $\sqrt{n}$ parametric rate.[16]

The consistency theorem is not as useful with general objective functions as 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 estimated consistently, then a consistent estimator of the identified set can be obtained easily by replacing the boundary by its sample analog. This is a common class of problem in which the identified set is an interval, for example, as in $\Theta_I = [\theta_1, \theta_2]$, where $\theta_1$ and $\theta_2$ can both be estimated consistently from the data.

## 4.3. 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 highlight this approach here from a frequentist perspective (for recent work on inference in partially identified models from a Bayesian perspective, see Liao & Jiang 2009, Moon & Schorfheide 2009). The confidence regions for $\Theta_I$ and those for $\theta^*$ are different here. We first highlight a subsampling-based empirical approach to construct these confidence regions. Subsampling is a resampling technique used to approximate large sample distributions. This technique is general and can be used in identified sets defined as minimizers of an objective function. I then discuss different approaches, some of which have been shown to be more powerful than subsampling, for the class of problems in which the identified set is defined through moment inequalities. Inference in these models is complicated because there exist problems with nuisance parameters that are manifested through a subtle property, mainly that of uniformity of coverage with respect to all possible DGPs. This uniformity issue is described in more detail below. Overall, the discussion is descriptive but sufficiently detailed to allow one to obtain a snapshot of the kinds of work in this literature.

---

[15]For exact conditions, such as the existence of polynomial minorants on $Q_n(\theta)$, see Chernozhukov et al. (2007).

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

### 4.3.1. Confidence regions for an identified set using subsampling.
Consider the objective function $Q_n(.)$, where we are interested in constructing a confidence region for the set $\Theta_I$. One approach, described here, follows Chernozhukov et al. (2007) and is based on subsampling. Another similar approach is detailed in Romano & Shaikh (2008). The set that we construct, similar to the consistent set, is a level set $C_n(c_n)$, where $c_n$ is chosen in a particular way. For instance, we consider the level set in Equation 6. Consider $B_n$ subsets[17] of size $b \ll n$. First, let $c_{0n} \geq inf_\theta Q_n(\theta)$. Usually, one can set $C_{0n}$ to be 10% higher than the infimum of $Q_n$. Second, compute $c_{1n}$ as the $\alpha$-quantile of the sample $\hat{C}_{j,b} = \{\sup_{C_n(c_{0n})} Q_{j,b}(.) : j = 1, \ldots, B_n\}$, where $Q_{j,b}$ is the objective function evaluated at the $j$-th subsample. Third, repeat the second step above three to four 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 Chernozhukov et al. 2007, theorem 3.3; see also Romano & Shaikh 2010).

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, it is possible to use a modified bootstrap procedure, or simulation methods to do inference on the set $\Theta_I$ (e.g., see Chernozhukov et al. 2007, remarks 4.5 and 4.6). Finally, for another approach to do inference on sets based on random set theory, we refer the reader to Beresteanu & Molinari (2008).

### 4.3.2. Confidence regions for an identified parameter using subsampling.
Imbens & Manski (2004) consider inference on the identified parameter in simple setups. They argue that there is usually a unique parameter $\theta^*$ that is the true parameter of interest, even though this parameter is not point identified. Most importantly, confidence regions for $\theta^*$ are no larger that the confidence regions for the identified set. Here, we highlight an approach to obtaining a confidence region $C_n$ for $\theta^*$, i.e., $P(\theta^* \in C_n) \geq \alpha$ as $n \to \infty$. Imbens & Manski (2004) also point 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 in which the underlying model is point identified, or close to point identified, the size of these confidence regions is maintained. This is a problem similar to inference in models with parameters on the boundary, in which values of the true parameter near or on the boundary cause problems for standard large-sample approximations. Uniformity problems arise typically in nonstandard models in which the asymptotic distribution is a nondifferentiable function of the true parameter. The validity of subsampling methods in non-point-identified models has been considered by Chernozhukov et al. (2007), Romano & Shaikh (2008), and especially Andrews & Guggenberger (2009a,b).

The general idea of constructing a confidence interval for the true parameter $\theta^*$ is to exploit the duality between testability and the confidence region. In essence, a confidence region is the set of parameters that cannot be rejected. Therefore, 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

---

[17]In larger samples, confidence regions should be computed for a series of subsample sizes starting with $b = n/5$, for example (for more details on subsample sizes, see Politiset al. 1999).

cannot be rejected constitutes the confidence region. There are many approaches to do this in econometrics, especially for models defined by moment inequalities (e.g., see Bugni 2007, Canay 2007, Chernozhukov et al. 2007, Romano & Shaikh 2008, Rosen 2008, Andrews & Soares 2009). Andrews & 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 Chernozhukov et al. (2007) and Romano & Shaikh (2008) in constructing a confidence region. These methods apply not just to moment inequalities, but also to other general models based on minimizing objective functions.

Start by choosing a parameter $\theta$ in the parameter space, and then use the value of $nQ_n(\theta)$ as the test statistic under the null that $\theta^* = \theta$, where we add $\theta$ 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 (for example conditions needed and statements of the theorems, see Chernozhukov et al. 2007, section 5, and Romano & Shaikh 2008, section 3.2; Chernozhukov et al. 2007, see theorem 5.2, also provide critical values constructed using bootstrap and simulations). We outline an approach for general objective functions and then consider a simple example below.

We assume here that $Q(\theta) \geq 0$, for all $\theta$, and $Q(\theta_0) = 0$. Therefore, 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

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

where we recenter, as recentering can lead to better finite sample values. The confidence region is then $C_n = \{\theta \in \Theta : nQ_n(\theta) \leq c_n(\theta, \alpha)\}$. 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 nonempty.

## 4.4. Confidence Regions in Interval Bounds: A Simple Example

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

$$\Theta_I = [\theta_l, \theta_b], \tag{7}$$

where $\theta_l$ and $\theta_b$ can be consistently estimated. This is a simple and important example covering cases in which the parameter of interest is scalar and 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 $\Theta_I$ with a prespecified probability $\alpha$. Then we highlight various approaches to construct intervals that cover the identified parameter $\theta^*$ with probability $\alpha$.

**4.4.1. Confidence region for $\Theta_I = [\theta_l, \theta_b]$.** Here, we are covering the interval in Equation 7 above, so heuristically a confidence region would be a set of intervals. One way to do that is to use a joint confidence region on the end points and map that into a confidence region for $\Theta_I$, imposing the fact that the joint confidence region on the end points is under the constraint that it is an interval in which the right end point 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 $\alpha\%$ of the generated intervals within it. This approach was used, for

example, in Horowitz & Manski (2000) and other papers. Here we follow our analysis above and derive an objective function that is minimized on $\Theta_I$. One such objective function is

$$Q(\theta) = (\theta_l - \theta)_+^2 + (\theta_b - \theta)_-^2, \tag{8}$$

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 $\theta \in \Theta_I$. Assume that we have $\hat\theta_l$ and $\hat\theta_b$ such that

$$\begin{bmatrix} \hat\theta_l - \theta_l \\ \hat\theta_b - \theta_b \end{bmatrix} \to_d \mathbb{Z} = \begin{bmatrix} \mathbb{Z}_1 \\ \mathbb{Z}_2 \end{bmatrix},$$

where $\mathbb{Z}$ 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

$$Q_n(\theta) = (\hat\theta_l - \theta)_+^2 + (\hat\theta_b - \theta)_-^2. \tag{9}$$

We know that the event $\{\Theta_I \subseteq C_n(c)\}$ is equivalent to the event $\sup_{\theta \in \Theta_I} nQ_n(\theta) \leq c$, 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

$$\sup_{\theta \in \Theta_I} nQ_n(\theta) \to_d \max\left((\mathbb{Z}_1)_+^2, (\mathbb{Z}_2)_-^2\right).$$

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

### 4.4.2. Confidence region for $\theta^* \in [\theta_l, \theta_b]$.

Imbens & Manski (2004) argue that one can report the confidence region on the parameter $\theta^* \in [\theta_l, \theta_b]$. To build this confidence interval, one can collect all the parameters $\theta$ that cannot be rejected under some appropriate test that they belong to $\Theta_I$. There are many approaches to building such an interval. Here, as in above, we build these based on the simple objective function (Equation 8). The choice of the objective is relevant in moment inequality models and can impact the power of the test (for more details, see Andrews & Jia 2008). An important issue that arises in these settings is that of the uniform consistency of the testing procedure, which impacts the asymptotic behavior of the test statistic.

I provide here a simple and heuristic discussion of the issue of uniformity (for a discussion of uniformity in the interval bounds setting, see Imbens & Manski 2004; for a thorough discussion of uniformity in these, and other, contexts, see Andrews & Guggenberger 2009a,b). 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 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 DGPs.

Uniform consistency is best understood in the context of a simple example. Consider again the moment inequality model with the identified set $\Theta_I = \{\theta: \theta_l \leq \theta \leq \theta_b\}$. 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 a 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 $\theta$, 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, conversely, 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 a probability close to at least 95%, as long as the samples are 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 $\theta$. 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., $\theta^*$ is point identified. This requires some modification of the standard approach of constructing the confidence interval for a fixed DGP (e.g., see Imbens & Manski 2004 for a discussion of a procedure that is consistent but not uniformly consistent across a family of DGPs 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 in which we can easily show under the null (see Chernozhukov et al. 2007, section 5) that

$$nQ_n(\theta) \to_d Q = \left( \mathbb{Z}_1 + \xi_1(\theta) \right)_+^2 + \left( \mathbb{Z}_2 + \xi_2(\theta) \right)_-^2, \tag{10}$$

where $\infty_1(\theta) = -\infty$ if $\theta_l < \theta$ and is equal to zero otherwise, and $\xi_2 = +\infty$ if $\theta_h > \theta$ and is equal to zero otherwise. The $\xi$'s are parameters that cannot be estimated consistently. However, we can estimate the $\alpha$-quantile of $Q$ by simulating the distribution of the random variable $Q_n$,

$$Q_n = \left( \mathbb{Z}_1 + \xi_1^n(\theta) \right)_+^2 + \left( \mathbb{Z}_2 + \xi_2^n(\theta) \right)_-^2,$$

where we can simulate the distribution of $(\mathbb{Z}_1, \mathbb{Z}_2)$ and set, for example, $\xi_1^n = -\infty$ if $\hat{\theta}_l + c_1\sqrt{\log n/n} < \theta$ and is equal to zero otherwise for some positive constant $c_1$, and similarly for $\xi_2^n$. Other approaches to constructing a confidence interval in this case can be used, such as ones based on the pseudo-likelihood approach of Rosen (2008), empirical likelihood (as in Canay 2007), bootstrap (as in Bugni 2007), or the generalized moment selection procedures (introduced in Andrews & Soares 2009 and further refined in Andrews & Jia 2008). Subsampling-based intervals can also be constructed, as in Chernozhukov et al. (2007) and Romano & 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 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 setting 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 esoteric point identification assumptions, researchers should understand the map between information and assumptions characterizing what can be learned under different sets of assumptions.

Economists have long valued models to gain insights, polish and discipline their communication, and examine what can happen within these toy economies under different policies. Data also play an important role in shedding light on certain postulates, disciplining theories and informing policy. However, most models contain a subset of assumptions based only on convenience—analytical or computational—such as functional forms or distributional assumptions. The choice of assumptions for investigation is problem specific. However, in most cases, these suspect assumptions are 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—e.g., to construct level sets—so advances in computational methods tailored toward these problems are essential. Also, inference in models in which the parameter of interest is infinite dimensional, and in which this parameter is partially identified, is also an 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 & Leamer (1983) is certainly in the spirit of partial identification, but its practical usefulness has not been exploited. I conjecture that this may be because 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 Gilstein & Leamer's (1983) vision to more specific problems and make use of the body of work thus far on inference procedures to implement it. A step toward this goal is the recent work of Chen et. al. (2010).

The hallmark of microeconometric work in the past 30 years has been concern with semiparametric models, with its main motivation being that of robustness against one class of assumptions or another. In the past two decades, partial identification, and the analysis of econometric models that are not necessarily point identified, has entered the realm of what econometricians accept, think about, and consider, and this approach to inference has appeared in important empirical work. Therefore, 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. This is exactly what is required from any serious scientific program.

## DISCLOSURE STATEMENT

## ACKNOWLEDGMENTS

## LITERATURE CITED

Ahn H, Powell J. 1993. Semiparametric estimation of censored selection models. *J. Econom.* 58:3–29

Andrews DWK, Guggenberger P. 2009a. Hybrid and size-corrected subsampling methods. *Econometrica* 77(3):721–62

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

Andrews DWK, Jia P. 2008. *Inference for parameters defined by moment inequalities: a recommended moment selection procedure*. Work. Pap., Yale Univ.

Andrews DWK, Soares G. 2009. Inference for parameters defined by moment inequalities using generalize moment selection. *Econometrica* In press

Aradillas-Lopez A. 2005. *Semiparametric estimation of a simultaneous game with incomplete information*. Work. Pap., Princeton Univ.

Aradillas-Lopez A, Tamer E. 2008. The identification power of equilibrium in simple games. *J. Bus. Econ. Stat.* 26:261–83

Bajari P, Hong H, Ryan S. 2005. *Identification and estimation of discrete games of complete information*. Work. Pap., Stanford Univ.

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

Beresteanu A, Molinari F, Molchanov I. 2008. *Sharp identification regions in games*. Work. Pap., Cornell Univ.

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

Berry S, Tamer E. 2006. Identification in models of oligopoly entry. In *Advances in Economics and Econometrics, Ninth World Congress of the Econometric Society*, ed. R Blundell, W Newey, T Persson, 2:46–85. Cambridge, UK: Cambridge Univ. Press

Bielby W, Hauser R. 1977. Structural equation models. *Annu. Rev. Sociol.* 3:137–61

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

Bollinger C. 1996. Bounding mean regressions when a binary regressor is mismeasured. *J. Econom.* 73(2):387–99

Bresnahan T, Reiss P. 1991. Empirical models of discrete games. *J. Econom.* 48:57–81

Bugni F. 2007. *Bootstrap inference in partially identified models*. Work. Pap., Northwestern Univ.

Canay I. 2007. *EL inference for partially identified models: large deviations optimality and bootstrap validity*. Work. Pap., Northwestern Univ.

Chen X, Tamer E, Torgovitsky A. 2010. *Inference in semiparametric likelihoods with partial identification*. Work. Pap., Northwestern Univ.

Chernozhukov V, Hong H, Tamer E. 2007. Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75:1243–84

Chiburis R. 2009. *Approximately most powerful tests for moment inequalities*. Work. Pap., Univ. Texas, Austin

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

Coombs C. 1965. *Theory of Data*. New York:Wiley & Sons

Cross P, Manski CF. 2002. Regressions, short and long. *Econometrica* 70:357–68

Duncan O, Davis B. 1953. An alternative to ecological correlation. *Am. Sociol. Rev.* 18:665–66

Fan Y, Park SS. 2009. Sharp bounds on the distribution of treatment effects and their statistical inference. *Econ. Theory* In press

Fréchet M. 1951. Sur les tableaux de correlation dont les marges sont données. *Ann. Univ. Lyon A* 3:53–77

Frisch R. 1934. *Statistical Confluence Analysis*. Publ. No. 5. Oslo: Univ. Inst. Econ.

Fuller W. 1987. *Measurement Error Models*. New York: Wiley

Galichon A, Henry M. 2009. *Inference in models with multiple equilibria*. Work. Pap., Univ. Montreal

Gilstein CZ, Leamer EE. 1983. Robust sets of regression estimates. *Econometrica* 51:321–33

Grieco P. 2009. *Inference in games with complete and incomplete information errors*. Work. Pap., Northwestern Univ.

Gronau R. 1974. Wage comparisons—a selectivity bias. *J. Polit. Econ.* 82(6):1119–43

Haile P, Tamer E. 2003. Inference with an incomplete model of English auctions. *J. Polit. Econ.* 111(1):1–51

Heckman J. 1974. Shadow prices, market wages, and labor supply. *Econometrica* 42:679–94

Heckman J, Honoré B. 1990. The empirical content of the Roy model. *Econometrica* 58:1121–49

Heckman J, Singer B. 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52:271–320

Heckman JJ, Hotz VJ. 1989. Choosing among alternative nonemperimental methods for estimating the impact of social programs: the case of manpower training. *J. Am. Stat. Assoc.* 84:862–74

Horowitz J, Manski C. 2000. Nonparametric analysis of randomized experiments with missing covariate and outcome data. *J. Am. Stat. Assoc.* 95:77–84

Horowitz J, Manski CF. 1995. Identification and robustness with contaminated and corrupted data. *Econometrica* 63:281–302

Hotz VJ, Mullin CH, Sanders SG. 1997. Bounding causal effects using data from a contaminated natural experiment: analysis the effects of teenage childbearing. *Rev. Econ. Stud.* 64(4):575–603

Imbens G, Manski CF. 2004. Confidence intervals for partially identified parameters. *Econometrica* 72:1845–57

Kline B, Tamer E. 2009. *Bounds for best response functions in binary games*. Work. Pap., Northwestern Univ.

Koopmans T, Reiersol O. 1950. The identification of structural characteristics. *Ann. Math. Stat.* 21:165–81

LaLonde RJ. 1986. Evaluating the econometric evaluations of training programs with experimental data. *Am. Econ. Rev.* 76(4):604–20

Leamer EE. 1981. Is it a demand curve, or is it a supply curve? Partial identification through inequality constraints. *Rev. Econ. Stat.* 63(3):319–27

Leamer EE. 1985. Sensitivity analyses would help. *Am. Econ. Rev.* 75(3):308–13

Liao Y, Jiang W. 2009. *Bayesian analysis in moment inequality models*. *Ann. Stat.* In press

Manski CF. 1989. Anatomy of the selection problem. *J. Hum. Resour.* 24(3):343–60

Manski CF. 1990. Nonparametric bounds on treatment effects. *Am. Econ. Rev. Pap. Proceed.* 80:319–23

Manski CF. 1997. Monotone treatment response. *Econometrica* 65:1311–34

Manski CF. 2003. *Partial Identification of Probability Distributions*. Springer Ser. Stat. New York: Springerr

Manski CF. 2007. *Identification for Prediction and Decision*. Cambridge, MA: Harvard Univ. Press

Manski CF, Pepper J. 2000. Monotone instrumental variables: with an application to the returns to schooling. *Econometrica* 68:997–1010

Manski CF, Tamer E. 2002. Inference on regressions with interval data on a regressor or outcome. *Econometrica* 70:519–47

Marschak J, Andrews WH. 1944. Random simultaneous equations and the theory of production. *Econometrica* 12:143–203

Moon R, Schorfheide F. 2009. *Bayesian and frequentist inference in partially identified models*. Work. Pap., Univ. Penn.

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

Nerlove M. 1965. *Estimation and Identification of Cobb-Douglas Production Functions*. Chicago: Rand McNally

Nevo A, Rosen A. 2009. *Inference with imperfect instrumental variables*. Work. Pap., Northwestern Univ.

Pakes A, Porter J, Ho K, Ishii J. 2005. *Moment inequalities and their applications*. Work. Pap., Harvard Univ.

Peterson AV. 1976. Bounds for a joint distribution function with fixed subdistribution functions: applications to competing risks. *Proc. Natl. Acad. Sci. USA* 73:11–13

Phillips PC. 1989. Partially identified econometric models. *Econom. Theory* 5(2):181–240

Politis D, Romano J, Wolf M. 1999. *Subsampling.* Springer Ser. Stat. New York: Springer

Ponomareva M, Tamer E. 2009. *Misspecification in moment inequality models: back to moment equalities?* Work. Pap., Northwestern Univ.

Romano J, Shaikh A. 2008. Inference for identifiable parameters in partially identified econometric models *J. Stat. Plan. Inference* 138(9):2786–807

Romano J, Shaikh A. 2010. Inference for the identified set in partially identified econometric models. *Econometrica* In press

Rosen AM. 2008. Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *J. Econom.* 146(1):107–17

Rosen S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *J. Polit. Econ.* 82:34–55

Santos A. 2008. *Inference in nonparametric instrumental variables with partial identification.* Work. Pap., Univ. Calif., San Diego

Stoye J. 2009. More on confidence regions for partially identified parameters. *Econometrica* 77:1299–315

Sweeting A. 2004. *Coordination games, multiple equilibria, and the timing of radio commercials.* Work. Pap., Duke Univ.

Tamer E. 2003. Incomplete bivariate discrete response model with multiple equilibria. *Rev. Econ. Stud.* 70:147–67

Tang X. 2008. *Bounds on the counterfactual distribution of revenues in auctions with reserve prices.* Work. Pap., Univ. Penn.