# RECENT DEVELOPMENTS IN PARTIAL IDENTIFICATION

BRENDAN KLINE AND ELIE TAMER

ABSTRACT. Identification strategies concern what can be learned about the value of a parameter based on the data and the model assumptions. The literature on partial identification is motivated by the fact that it is not possible to learn the exact value of the parameter for many empirically relevant cases. A typical result in the literature on partial identification is a statement about characterizing the identified set, which summarizes what can be learned about the parameter of interest given the data and model assumptions. For instance, this may mean that the value of the parameter can be learned to necessarily be within some set of values. First, the review surveys the general frameworks that have been developed for conducting a partial identification analysis. Second, the review surveys some of the more recent results on partial identification.

## 1. INTRODUCTION

An important part of any empirical analysis in economics is the identification strategy: how are the assumptions and the data used to answer the empirical question? Often, this involves the empirical researcher telling a story that involves motivating key assumptions and describing specific features of the data that are used or "exploited" to recover information about the value of the relevant parameter of a statistical model. The parameter could be, for instance, a coefficient in a regression model or an average treatment effect in a potential outcomes model. This process of establishing the identification strategy can involve reference to a formal theorem that characterizes the relationship between the parameter and the data,

UNIVERSITY OF TEXAS AT AUSTIN
HARVARD UNIVERSITY
*E-mail addresses*: brendan.kline@austin.utexas.edu, elietamer@fas.harvard.edu.
Date: August 2022. Prepared for the Annual Review of Economics.

given a particular set of assumptions. Alternatively, it can involve a more informal discussion surrounding considerations like "sources of exogenous variation" that identify the parameter.

The general presumption tends to be that an identification strategy should provide a way to learn the exact value of the parameter. If so, then the parameter is said to be point identified. Identification strategies concern what can be learned based on the population data. This abstracts away from finite sample considerations that can be a different reason that the exact value of the parameter is not learned in an empirical analysis. Results in the literature on point identification that are motivated by this presumption often can be characterized as an attempt to find a set of assumptions and features of the data that are sufficient for point identification. In other words, the literature on point identification often takes the question "under what assumptions and with what kind of data is it possible to learn the exact value of the parameter?" as its motivation. This is an all-or-nothing approach, a binary of view of identification: the parameter value is either (point) identified or it is not.

The literature on partial identification provides an alternative perspective on empirical analysis that avoids this binary view. The literature on partial identification asks "what can be learned about a parameter under a given set of assumptions and with a given kind of data?" Unlike the literature on point identification, the literature on partial identification does not prioritize searching for assumptions and data with the goal of a result of learning the exact value of the parameter. Indeed, often the literature on partial identification does not prioritize considerations relating to how much is learned about the parameter. Rather, the literature on partial identification prioritizes working with a plausible set of assumptions and data, and studying what can be learned about the value of the parameter. In other words, with point identification, the starting point is the "goal" of learning the exact value; with partial identification, the starting point is a set of model assumptions and data.

A typical result in the literature on partial identification is a statement that the value of the parameter can be learned to necessarily be within some set of values that depends on the data and the assumptions. This set is known as an identified set. Sometimes this is referred to as "bounds" on the value of the parameter. From the point of view of identification analysis, learning an identified set is equivalent to learning about the value of the parameter.[1] The partial identification framework generalizes the point identification framework. By definition, point identification is a special case of partial identification, and so an emphasis on point identification is an unnecessary limiting restriction on empirical analyses.

This review complements the first review article published in this journal on the same topic. See Tamer (2010). The review surveys some of the common identification strategies (i.e., combinations of assumptions and data) that have led to partial identification results for specific statistical models, including ideas relating to shape constraints and ideas related to the observable implications of optimizing behavior. The review also discusses topics relating to sensitivity analysis and data limitations, among other issues. This part of the review explores how common identification strategies have been used in a variety of specific models, including treatment effects models and choice models involving a single decision maker or multiple decision makers.

This part of the review shows the sense in which point identification is an exceptional special case, in that the core ideas of many standard identification strategies result in partial identification rather than point identification. Point identification often requires additional assumptions that are not directly motivated by the underlying core idea of the identification strategy. As such, this part of the review emphasizes that partial identification can be viewed as the rule rather than the exception for many standard identification strategies. Therefore, partial identification methods provide the scope for an empirical analysis that more directly follows the core idea of the identification strategy. In such cases, this would not be possible

[1]However, from the point of view of (finite sample) statistical inference, an issue this review does not emphasize, there is a subtle difference between inference for the value of the parameter and inference for the identified set. Essentially, this is because the identified set actually is a point identified object (given it can be learned exactly) whereas by definition the value of the parameter is a partially identified object. However, since this review focuses on identification analysis, this distinction does not arise.

based on a presumption that an identification strategy needs to result in point identification of the parameter.

Based on a presumption that the goal of an identifications strategy is point identification, the empirical researcher can have any of multiple potential responses to finding that it is not possible to learn the exact value of the parameter relevant for answering an empirical question. All of them come with certain costs that can be avoided by taking a partial identification approach instead. One possibility is to conclude that an empirical analysis is not possible on the basis that the parameter is not point identified under a desired set of assumptions and based on the available data. However, this may exaggerate the inability to answer the empirical question. Partial identification allows the empirical researcher to learn something about the parameter, which is more informative than concluding that an empirical analysis is not possible. Another possibility is to conclude that stronger assumptions are needed to answer the empirical question, and consequently base the empirical analysis in part on those additional assumptions even if they were not initially part of the preferred analysis. However, this can diminish the credibility of the resulting empirical analysis, to the extent that the additional assumptions are difficult to motivate. Partial identification can avoid the use of such assumptions, thereby resulting in a more credible analysis. Another possibility is to find or generate alternative or additional data, for instance conducting a randomized experiment to overcome endogeneity in observational data. This can be a particularly convincing response, although it does not apply in all cases, and it can substantially raise the cost of the research, and in many cases there is no plausible data that would be sufficient for point identification. Partial identification can be based on the available data.

More generally, the review defines some of the general principles of an identification analysis, and surveys the frameworks that have been developed for conducting a partial identification analysis, including the moment inequality, criterion function, random sets, and generalized instrumental variables approaches. Each provides a particular “language” in which to conduct an identification analysis. These frameworks are parallels to frameworks like generalized method of moments or Z-estimation/M-estimation.

Because of the large size of the literature on partial identification, the review must be selective in the topics it covers. In particular, this review does not discuss much about the statistical inference issues associated with partial identification. Partial identification raises unique statistical inference considerations, particularly relating to uniform validity of inference, inference on components of the parameter of a statistical model, and the question of the appropriate object of inference (e.g., the identified set or the partially identified parameter). Other reviews with alternatives focuses or perspectives, some of which do cover statistical inference, include Tamer (2010), Molchanov and Molinari (2014), Bontemps and Magnac (2017), Canay and Shaikh (2017), Ho and Rosen (2017), Chesher and Rosen (2020), Molinari (2020), and Kline, Pakes, and Tamer (2021). For reviews of the overall literature on identification, often with a focus on point identification, see the comprehensive papers of Matzkin (2007, 2013) and Lewbel (2019).

## 2. OVERVIEW OF IDENTIFICATION

It is necessary to set up some of the general principles of identification.

### 2.1. **General Principles.**

A statistical model provides a connection between a parameter $\theta$ and observable data $w$. The true value of the parameter is $\theta_0$, and identification of $\theta$ concerns the question of what can be logically concluded about the value of $\theta_0$ on the basis of the combination of a maintained set of assumptions and the observed data. The identification analysis must be specialized for the particular statistical model, particularly relating to what features of the distribution of the observable data are determined by the statistical model.

For instance, the statistical model can determine a complete specification of a data generating process for $w$, resulting in a distribution $F_\theta(w)$ for each value of $\theta$. A familiar example is a parametric model of the sort that is estimated by maximum likelihood. The observed distribution of the data $F_0$ is equal to the distribution determined by the statistical model at the true value of the parameter $\theta_0$, and thus $F_0 = F_{\theta_0}$. In such a case, an identification analysis can concern the determination of the values of $\theta$ such that $F_\theta = F_0$. The set of values $\{\theta : F_\theta = F_0\}$ is the set of values of $\theta$ that generate the observed distribution of the data. The true value $\theta_0$ is in the set, by definition. Sometimes this set is referred to as the set of "observationally equivalent" values of the parameter, in the sense that they all result in the same distribution of the observed data. Consequently, the data cannot be used to distinguish between them.

Typically, the statistical model concerns a relationship among the components of $w$. In many cases, this involves an "outcome" $y$ and "explanatory variables" $x$, so that $w = (y, x)$. In such cases, often the parameter $\theta$ characterizes the relationship between $y$ and $x$. In such a case, the statistical model can result in a conditional distribution $F_\theta(y|x)$. A familiar example is a parametric model of the sort that is estimated by conditional maximum likelihood. Many discrete outcome models are of this kind. The set of values $\{\theta : F_\theta(\cdot|x) = F_0(\cdot|x) \text{ w.p.1. } x\}$ is the set of values of $\theta$ that generate the observed distribution of the data. The notation "w.p.1. $x$" means with probability 1 with respect to $x$, and is practically the same as "for all $x$". Now, "observational equivalence" requires that the conditional distributions resulting from $\theta$ result in the conditional distributions in the observed data.

Alternatively, the statistical model can determine specific features of the data generating process for $w$. A common example is the expected value $E_\theta(w)$ for each value of $\theta$. This includes the familiar example of "non-parametric" estimation of moments of the data, where $\theta$ can be simply defined to be the expected value (i.e., $E_\theta(w) = \theta$). In the case of expected values, an identification analysis can concern the determination of the values of $\theta$ such that $E_\theta(w) = E_0(w)$ where $E_0$ is the expected value in the observed data. The set of values $\{\theta : E_\theta(w) = E_0(w)\}$ is the set of values of $\theta$ that generate the observed expected value of the data. Sometimes this is also referred to as "observational equivalence" but it is important to recognize this is relative only to the expected value of the data, and not the overall distribution of the data. More generally, when specific features of the data are used to identify the parameter, it is important to recognize that "observational equivalence" is relative to those features of the data, and not the overall distribution of the data. Sometimes, the statistical model can determine features of the conditional distribution of $y$ given $x$. A familiar example is the linear model that results in a specification for $E_\theta(y|x)$ for each value of $\theta$. The set of values $\{\theta : E_\theta(y|x) = E_0(y|x) \text{ w.p.1. } x\}$ is the set of values of $\theta$ that generate the observed conditional expected values of the data.

Statistical models may determine features of the distribution of $w$ other than the features that are used in an identification analysis. In other words, in many cases, an identification analysis is based only on a restricted set of implications of the model. For instance, a particular identification analysis may be based only on selected expected values even though the model also determines other expected values, or other features of the data.

Thus, identification analysis involves the question of which features of the data are used to identify the parameter. In some cases, including the linear model under the standard assumptions used by ordinary least squares, the statistical model only determines a specific set of features of the data, like expected values. If so, then those features of the data (e.g., expected values) are the only feature of the data relevant for learning about $\theta$, since $\theta$ has no further implications. However, in other cases, the model actually does determine other features of the data generating process for $w$. For instance, the linear model with distributional assumptions on the unobservable term (e.g., normality assumptions used for finite sample inference) determines the full distribution of the data, and not just expected values. With such a model, the identification analysis could be based only on the expected values or it could be based on the full distribution of the data. The logical conclusions about $\theta$ that can be drawn from expected values could be different from the logical conclusions that can be drawn from the full distribution of the data. More generally, the logical conclusions about $\theta$ can depend on the features of the data that are used in the identification analysis. Therefore, when ambiguity is possible, it is important to be clear about which features of the data are used by an identification analysis. It is common, for example, to base an identification analysis only on specific expected values even though the model determines additional expected values, or even a full distribution; it is important to be clear about this.

An important component of an identification analysis is the set of maintained assumptions. Often, the assumptions are determinants of the functional forms of $F_\theta$ or $E_\theta$. For instance, consider a linear model $y = x\beta + \epsilon$. In the typical analysis, the parameter $\theta$ consists only of the parameter $\beta$, so that $\theta = \beta$. In other analyses, including some discussed below, $\theta$ may contain additional components relating to the distribution of $\epsilon$. It follows that $E_\theta(y|x) = x\beta + E_\theta(\epsilon|x)$. This depends on the unknown $E_\theta(\epsilon|x)$. The typical exogeneity assumption used in an ordinary least squares analysis that $E_\theta(\epsilon|x) = 0$ results in $E_\theta(y|x) = x\beta$. Then, based on the conditional expected value $E_0(y|x)$ from the observed data, the identification analysis concerns the set of $\beta$ such that $x\beta = E_\theta(y|x) = E_0(y|x)$. This illustrates the common circumstance that the assumptions determine the functional form for $E_\theta(y|x)$.

The parameter $\theta$ is point identified when the exact value of $\theta$ can be logically inferred based on the assumptions and the data. The parameter $\theta$ is partially identified when the value of $\theta$ can be logically inferred to be in some set of values, based on the assumptions and the data.

An identified set in a partial identification analysis is the conclusion of a logical argument about what can be learned about the parameter on the basis of the assumptions and the data. In short, an identified set answers the question: "Based on the assumptions and the data, what is something that can be logically concluded about the value of the parameter?" Note this definition does not require that an identified set answer the more difficult question: "Based on the assumptions and the data, what is the strongest possible logical conclusion about the value of the parameter?"

Therefore, by definition, any value of the parameter that is not in an identified set can be ruled out as a candidate value of the parameter. In general, it is possible that there are values of the parameter in an identified set that are not compatible with the assumptions and the data. This can happen when it is difficult to derive all of the logical implications of the model, assumptions, and data. Even though some values of the parameter in an identified set may be incompatible with the assumptions and data, it may be difficult to actually prove this. The definition of an identified set does not require that it corresponds to the strongest possible logical conclusion that is implied by the setup. Rather, the definition of an identified set implicitly accommodates the difficulty of actually deriving logical conclusions about the parameter. In particular, there is not a unique identified set in a given setup.

However, if an identified set does have the property that all values of the parameter in the identified set are indeed compatible with the assumptions and the data, then the identified set is the sharp identified set. The sharp identified set answers the question: "Based on the assumptions and the data, what is the strongest possible logical conclusion about the value of the parameter?" Hence, there is a unique sharp identified set. Often, the literature on partial identification uses the phrase the identified set (rather than an identified set) to refer to the sharp identified set. Obviously, the derivation of the sharp identified set is preferred, but in many cases the most that can be proved is the derivation of an identified set.

The fact that it is possible to work with an identified set rather than the sharp identified set represents a subtle but important feature of the partial identification approach to empirical analysis. It is difficult to derive the logical implications of a model, set of assumptions, and data. In the partial identification framework, it is possible to work with a result that draws some logical conclusions about the parameter even if the result does not exhaust all of the possible logical conclusions that could be drawn. In contrast, in the point identification framework, if it is not possible to establish conditions under which it is possible to prove that the parameter is point identified then it is not possible to conduct an empirical analysis.

This is about what is possible to prove, which is distinct from whether or not the parameter actually is point identified under some reasonable set of conditions. Given the existence of a literature on identification, it is obviously non-trivial to establish point identification results. In the partial identification framework, as long as it is possible to come to some logical conclusions about the parameter, it is possible to conduct an empirical analysis, even if more conclusions could in principle be drawn about the parameter. This means that the partial identification framework accommodates the difficult in proving identification results in a way that the point identification framework does not.

Sometimes, an identified set that is known to not be the sharp identified set is called an outer identified set. Establishing "sharpness" of an identified set should not be over-prioritized relative to establishing an identified set under credible assumptions. Of course, for a given set of assumptions, it is preferred to have the sharp identified set. But, the pursuit of "sharpness" should not influence the choice of assumptions used in the identification analysis. This directly parallels the perspective that the pursuit of "point identification" should not influence the choice of assumptions used in the identification analysis.

Further, sharpness is necessarily relative to a specific set of formal assumptions. In general, there is an important difference between the set of formal assumptions and the set of assumptions actually "believed" by the empirical researcher. Typically, the empirical researcher actually "believes" more than is represented by the set of formal assumptions. In those cases, the sharp identified set actually answers the question "What is the strongest possible logical conclusion about the value of the parameter on the basis of a subset of the assumptions actually 'believed' by the empirical researcher." It may be preferable to have a potentially non-sharp identified set under a set of assumptions that more closely matches those actually "believed," compared to a sharp identified set that is sharp only under a different set of assumptions.

2.2. **Example: The Linear Model.** It is possible to illustrate these ideas in the familiar context of a linear model. Identification in the "semi-parametric" linear model $y = x\beta + \epsilon$ with the exogeneity assumption $E_\theta(\epsilon|x) = 0$ is based on the implication that $E_\theta(y|x) = x\beta$. There is point identification when there is a unique value of $\beta$ that satisfies $x\beta = E_0(y|x)$ w.p.1. $x$, where $E_0(y|x)$ is the conditional expected value that is observed in the data. Because the true value necessarily satisfies the condition that $x\beta_0 = E_0(y|x)$ w.p.1. $x$, given the model is correctly specified, it equivalently follows that point identification is the condition that $\beta_0$ is the unique solution to $x\beta = E_0(y|x)$ w.p.1. $x$. More generally, the set $\{\beta : x\beta = E_0(y|x) \text{ w.p.1. } x\}$ is the identified set for $\beta$.

In this "semi-parametric" specification of the linear model, there are no further implications of the model, so this actually is the sharp identified set. This formulation is semi-parametric because there is an infinite-dimensional unknown: the distribution of $\epsilon$, an infinite-dimensional object, is left unspecified and indeed is often viewed as not being a parameter of the model.

However, as discussed above, other specifications of "linear models" may involve further assumptions. With further assumptions, the model may determine additional features of the data generating process beyond expected values. Furthermore, other specifications may involve the introduction of additional parameters. For instance, the specification may involve distributional assumptions on $\epsilon$, like the assumption that $\epsilon$ has a normal distribution with mean zero and variance $\sigma^2(x)$, where $\sigma^2(\cdot)$ is the variance as a function of $x$. Various additional assumptions, like the homoskedasticity assumption that $\sigma^2(x) = \sigma^2$, could be further imposed. If $\sigma^2(\cdot)$ is unknown, it is an additional parameter of the model; hence, the parameter becomes $\theta = (\beta, \sigma^2(\cdot))$. This would imply that $y|x$ has a normal distribution with mean $x\beta$ and variance $\sigma^2(x)$. In such a specification of the linear model, $\{\beta : x\beta = E_0(y|x) \text{ w.p.1. } x\}$ is still an identified set for $\beta$ because logically the true value $\beta_0$ must be in that set. Note that this identified set is for a component of $\theta$, since it concerns $\beta$ but not $\sigma^2(\cdot)$. This reflects a more general point that often identified sets are constructed for specific parameters of interest rather than the full parameter.

However, $\{\beta : x\beta = E_0(y|x) \text{ w.p.1. } x\}$ may not be the sharp identified set, depending on whether each value of $\beta$ in that set also is compatible with the additional assumptions. If the parameter is point identified based on the "typical" specification, then obviously additional assumptions would not influence the identification result, since it is not possible to be "more" than point identified in an identification analysis. But if the parameter is partially identified based on the "typical" specification, additional assumptions may influence the identification result. With the normality assumption, the sharp identified set for $\beta$ would be $\{\beta : F_{\mathcal{N}(x\beta,\sigma^2(x))}(\cdot) = F_0(\cdot|x) \text{ w.p.1. } x \text{ for some } \sigma^2(\cdot)\}$. The notation $F_{\mathcal{N}(x\beta,\sigma^2(x))}(\cdot)$ is the distribution of $\mathcal{N}(x\beta, \sigma^2(x))$. This identified set again reflects the idea that often the identified set for a component of the parameter is of interest.

Returning to the semi-parametric specification, the typical analysis of the linear model provides conditions for point identification. For any value of the parameter such that $x\beta = E_0(y|x)$ w.p.1. $x$, it follows that $x'x\beta = x'E_0(y|x)$ w.p.1. $x$ and therefore $E_0(x'x)\beta = E_0(x'y)$. Therefore, under the "full rank" condition on the data that requires that $E_0(x'x)$ has a matrix inverse, $\hat{\beta} = (E_0(x'x))^{-1} E_0(x'y)$ is the unique solution to $x\beta = E_0(y|x)$ w.p.1. $x$, which implies that $\beta$ is point identified. This full rank condition is a sufficient condition for point identification, alongside the rest of the setup.

Certain deviations from this analysis of the linear model result in partial identification of the parameter. It is possible to see this through two simple examples. The rest of the review provides a range of other examples of partial identification.

2.2.1. *Perfect Collinearity.* Suppose that the "full rank" condition fails and instead that there is a non-zero value $\tilde{\beta}$ of the parameter such that $x\tilde{\beta} = 0$ w.p.1. $x$. This can be viewed as a particular version of a "perfect collinearity" problem. Although the use of partial identification methods would probably not be the response to a "perfect collinearity" problem in practice, this can be used to provide an example of partial identification that is simple and based on a familiar problem. By matrix algebra, the set of $\beta$ that solve $x\beta = E_0(y|x)$ w.p.1. $x$ can be expressed as $\{\beta : \beta = \tilde{\beta} + \beta_0$ for some $\tilde{\beta}$ s.t. $x\tilde{\beta} = 0$ w.p.1. $x\}$.[2] Any value of the parameter in this set results in the same conditional expected values. Obviously, the true value $\beta_0$ is in this set (taking $\tilde{\beta} = 0$) but other values of the parameter can also be in this set (for any non-zero values of $\tilde{\beta}$). Specifically, this means that the parameter is partially identified when there is "perfect collinearity" and hence a non-zero $\tilde{\beta}$. This set shows that the parameter is partially identified, but note that this set depends on the unknown $\beta_0$, and therefore does not immediately provide a way to learn about the value of the parameter in empirical practice. However, it is trivial that an identified set for $\beta$ is $\{\beta : x\beta = E_0(y|x)$ w.p.1. $x\}$. Another representation of the identified set for $\beta$ is $\{\beta : \beta = \tilde{\beta} + (E_0(x'x))^{+}E_0(x'y)$ for some $\tilde{\beta}$ s.t. $x\tilde{\beta} = 0$ w.p.1. $x\}$ where $A^{+}$ is the Moore-Penrose generalized inverse of $A$.[3]

It is possible to use this example to discuss the fact that there can be different identification results for different specific features of a statistical model. If $\beta^{(1)}$ and $\beta^{(2)}$ are two values of $\beta$ in the identified set, then $x\beta^{(1)} = x\beta^{(2)} = x\beta_0$ w.p.1. $x$. Consequently, even though $\beta$ is partially identified, the "predictions" $x\beta_0$ are actually point identified on the observed support of $x$. If the goal of the empirical analysis were to construct such predictions, then the corresponding object of interest actually turns out to be point identified even with perfect collinearity. This reflects a subtle but important point about identification: even if "the

[2]For any value $\tilde{\beta}$ such that $x\tilde{\beta} = 0$ w.p.1. $x$, it follows that $x(\tilde{\beta} + \beta) = E_0(y|x)$ whenever $x\beta = E_0(y|x)$. Consequently, any parameter value of the form $\beta^* = \tilde{\beta} + \beta_0$ is such that $x\beta^* = E_0(y|x)$ w.p.1. $x$. Furthermore, for any value $\beta^*$ such that $x\beta^* = E_0(y|x)$ w.p.1. $x$, which can be always written as $\beta^* = \tilde{\beta} + \beta_0$ with $\tilde{\beta} = \beta^* - \beta_0$, it follows that $x\tilde{\beta} = x(\beta^* - \beta_0) = E_0(y|x) - E_0(y|x) = 0$ w.p.1. $x$. Consequently, any parameter value $\beta^*$ such that $x\beta^* = E_0(y|x)$ w.p.1. $x$ is of the form $\beta^* = \tilde{\beta} + \beta_0$ with $x\tilde{\beta} = 0$ w.p.1. $x$. The same argument and result is true with $\tilde{\beta}_0$ in place of $\beta_0$ where $\tilde{\beta}_0$ is any value of the parameter such that $x\tilde{\beta}_0 = E(y|x)$ w.p.1. $x$.

[3]Let $A = E_0(x'x)$. It holds $E_0((x(\beta_0 - E_0(x'x)^{+}E_0(x'y)))^2) = E_0((\beta_0 - A^{+}E_0(x'y))'x'x(\beta_0 - A^{+}E_0(x'y))) = E_0((\beta_0 - A^{+}A\beta_0)'x'x(\beta_0 - A^{+}A\beta_0)) = \beta_0'A\beta_0 - \beta_0'AA^{+}A\beta_0 - \beta_0'AA^{+}A\beta_0 + \beta_0'AA^{+}AA^{+}A\beta_0 = 0$ since $AA^{+}A = A$ as a property of the Moore-Penrose generalized inverse. Therefore, $x(\beta_0 - E_0(x'x)^{+}E_0(x'y)) = 0$ with probability 1. Therefore, $\beta$ is of the form $\beta = \tilde{\beta} + \beta_0$ with $x\tilde{\beta} = 0$ w.p.1. $x$ if and only if $\beta$ is of the form $\beta = \tilde{\tilde{\beta}} + E_0(x'x)^{+}E_0(x'y)$ with $x\tilde{\tilde{\beta}} = 0$ w.p.1. $x$, by taking $\tilde{\tilde{\beta}} = \tilde{\beta} + \beta_0 - E_0(x'x)^{+}E_0(x'y)$.

parameter" is partially identified, it is possible that the object of interest that answers the empirical question is point identified. It is important to recognize, however, the nuances of this fact. First, in this particular example, the "predictions" are point identified specifically on the observed support of $x$. If the empirical analysis involved making predictions for novel values of $x$, then the corresponding prediction would not be point identified. This follows since $\beta^{(1)} \neq \beta^{(2)}$ implies that there is some $x$ such that $x\beta^{(1)} \neq x\beta^{(2)}$. Therefore, it is important to be very careful in defining the object to be identified, since subtle issues like "support" can make a big difference. Second, in this particular example, even though these "predictions" are point identified, the marginal effects are partially identified. This follows since $\beta$ is the marginal effect, and $\beta$ itself is partially identified. In other words, identification properties are specific to specific quantities in potentially surprising ways: in this case, there is point identification of these "predictions" about the relationship between $y$ and $x$, but only for $x$ in the observed support, and yet partial identification of marginal effects. It is common to see statements like "the model is (partially) identified," which usually refers to the identification of the main parameter of the model $\theta$, but it is important to keep in mind that identification properties of $\theta$ are not necessarily directly relevant for every empirical question.

2.2.2. *Interval Data.* Alternatively suppose that the data does not contain $y$, but instead contains $(y_L, y_U)$ with the property that $y_L \leq y \leq y_U$. This can be viewed as a particular version of "interval data" specifically concerning the outcome variable (e.g., Manski and Tamer (2002)). Among other examples,[4] interval data is a feature of survey data that asks respondents to indicate in which of a set of ranges a particular numerical quantity lies (e.g., ranges of income). Although it remains true that the true value $\beta_0$ necessarily satisfies the condition that $x\beta_0 = E_0(y|x)$ w.p.1. $x$, the expected value $E_0(y|x)$ cannot be used in the identification strategy because it is not observed in the data. However, it is true that

[4]The earliest examples in the partial identification literature in economics that can also be cast as an interval data problem is the classic results of Manski in the context of the analysis of treatment response as a missing data problem (e.g., Manski (1989, 1990)). See also Manski (2003) for a summary.

$E_0(y_L|x) \leq E_0(y|x) \leq E_0(y_U|x)$, and therefore $\beta_0$ necessarily satisfies the condition that $x\beta_0 = E_0(y|x) \leq E_0(y_U|x)$ w.p.1. $x$; similarly, $x\beta_0 \geq E_0(y_L|x)$ w.p.1. $x$. Consequently, $\beta_0$ necessarily satisfies the moment inequality conditions that $E_0(y_L|x) \leq x\beta_0 \leq E_0(y_U|x)$ w.p.1. $x$. This means that it can be learned that $\beta_0$ satisfies those moment inequality conditions, so an identified set for $\beta$ is $\{\beta : E_0(y_L|x) \leq x\beta \leq E_0(y_U|x) \text{ w.p.1. } x\}$.

Partial identification as a general principle shows that there can be cases where something can be learned about a parameter, even if the exact value of the parameter cannot be learned. In a linear model, the typical perspective is that there cannot be "perfect collinearity" and that $y$ must be observed in order for an empirical analysis to proceed; however, in the partial identification framework, there is no reason to avoid an empirical analysis just because of these issues. This is because the partial identification approach prioritizes learning about the value of the parameter under reasonable conditions even if it does not result in learning the exact value of the parameter. In particular, although empirical results based on partial identification are often criticized as not being "informative" when "not enough" can be learned about the parameter of interest, the partial identification perspective can actually result in learning more about a parameter of interest compared to a perspective that avoids empirical analysis of situations that do not result in point identification. In the above examples, it is not possible to learn the exact value of the parameter, precluding an empirical analysis based on a presumption of point identification. However, the partial identification approach shows it is possible to conduct an empirical analysis and learn something about the value of the parameter.

## 3. FRAMEWORKS

With the understanding that partial identification is an important basic principle of empirical analysis, general frameworks that encompass partial identification have been developed. These general frameworks can be viewed as attempts from the partial identification literature to have equivalents to frameworks like generalized method of moments, maximum likelihood, or Z-estimation/M-estimation that are familiar in the analysis of point identified models.

These frameworks provide general ways to conduct an identification analysis. Further, these frameworks are associated with a substantial literature on finite sample estimation and inference results. An important consideration in an identification analysis is to have an identification result that can be implemented in empirical practice. If the identification result can be put into one of the frameworks discussed below, then the existing literature on finite sample considerations can be directly applied. Consequently, even from the point of view of an identification analysis, it is useful to be aware of the estimation and inference results. This can help guide the "form" of identification result. A familiar analogy is that it is convenient to have a point identification result in terms of moment conditions, because the generalized method of moments can then be used for estimation and inference. Here, the question is what "forms" of a partial identification result have corresponding estimation and inference results.

3.1. **Moment Inequalities.** The moment inequality framework encompasses the partial identification results that can be represented as statements concerning (conditional) moments. This framework allows for the possibility that the underlying statistical model implies moment inequality conditions, such that the true value of the parameter $\theta_0$ satisfies the moment inequality condition $E_0(m(w, \theta_0)) \geq 0$ where $m(w, \theta)$ is a known function of the data $w$ and parameter $\theta$. In general, $m$ is vector-valued. Correspondingly, the identified set is $\{\theta : E_0(m(w, \theta)) \geq 0\}$. Typically there are multiple solutions to moment inequality conditions, and hence typically there is partial identification of the parameter based on moment inequality conditions. However, it is possible that the parameter is point identified even with moment inequality conditions.

This compares to the generalized method of moments framework. In the generalized method of moments framework, the main idea is that the true value of the parameter $\theta_0$

satisfies the moment equality condition $E_0(m(w, \theta_0)) = 0$. If $\theta_0$ is the unique value of the parameter to satisfy $E_0(m(w, \theta)) = 0$, then the parameter is point identified.

Often, but not always, results in the moment inequality condition framework also allow for moment equality conditions. In some cases, this is allowed by accommodating a moment condition with $m_c(w, \theta) = -m_{c+1}(w, \theta)$ for some components $c$ and $c+1$ of the vector-valued $m$, in which case the moment inequality condition imposes that $E_0(m_c(w, \theta_0)) = 0$. Thus, as a special case, the moment inequality approach allows for the case that the underlying statistical model implies moment equality conditions with multiple solutions.

Some statistical models imply conditional moment inequality conditions. Although other approaches are also possible, these can be converted into unconditional moment inequalities using the result that $E_0(m(w, \theta)|z) \geq 0$ implies that $E_0(\alpha(z)m(w, \theta)) \geq 0$ for any non-negative function $\alpha(\cdot)$.

There is a large literature of results that are focused on the finite sample issues of estimation and inference. These results are typically intended to apply to a general setting. This literature has been reviewed in particular for instance in Canay and Shaikh (2017), among the other reviews of partial identification mentioned in the introduction.

A common issue in estimation and inference of partially identified models, and the moment inequality approach in particular, is the failure of continuity conditions that are standard in the analysis of point identified models. For instance, suppose the identified set can be represented as $\{\theta : \theta \in [\mu_L, \mu_U]\}$ where $\mu = (\mu_L, \mu_U)$ is the vector-valued moment of observable data. It might seem reasonable to estimate the identified set with the sample analogue $\{\theta : \theta \in [\mu_{L,N}, \mu_{U,N}]\}$ where $\mu_N = (\mu_{L,N}, \mu_{U,N})$ is an estimator of $\mu$. However, even if $\mu_N$ is a consistent estimator of $\mu$, it is not necessarily the case that the estimate of the identified set is a consistent estimate of the identified set. This is because the set $\{\theta : \theta \in [\mu_L, \mu_U]\}$ is actually not a continuous mapping from $\mu$. In particular, continuity can fail as $\mu_N \to \mu_0$ with $\mu_{L,0} = \mu_{U,0}$, since it can happen that $\mu_{L,N} > \mu_{U,N}$ so the estimate of the identified set can be the empty set (for each sample size along the sequence) even though the true identified set is not the empty set. The behavior of the estimated identified set depends on the joint distribution of $\mu_N$: it could be that $\mu_{L,N} \leq \mu_{U,N}$ w.p.1. or it could be that there is a non-vanishing probability that $\mu_{L,N} > \mu_{U,N}$. As this review is focused on identification, this review does not explore this issue further.

3.2. **Criterion Functions.** The criterion function framework encompasses the partial identification results that can be represented as statements concerning a criterion function. This framework allows for the possibility that the underlying statistical model implies that the true value of the parameter $\theta_0$ has some relation to a criterion function $Q_0(\theta)$. The leading case is that $\theta_0$ is such that $Q_0(\theta_0) = 0$, in which case the identified set is $\{\theta : Q_0(\theta) = 0\}$.

This compares to the Z-estimation and M-estimation frameworks. Similarly, there are are multiple important special cases of the criterion function approach. For instance, if the parameter satisfies the moment inequality conditions $E_0(m(w, \theta_0)) \geq 0$, then the parameter satisfies the criterion function condition that $Q_0(\theta) = ||[E_0(m(w, \theta_0))]_-|| = 0$, where $a_- = \min(a, 0)$ component-wise, and for any norm $|| \cdot ||$.[5] With moment inequality conditions, other criterion functions can be specified with the same property that $Q_0(\theta) = 0$ exactly when $E_0(m(w, \theta)) \geq 0$. Although moment inequalities are typically the motivation for the criterion function approach, this framework allows for a more general identification strategy that does not involve moment inequality conditions.

Related finite sample estimation and inference results include Chernozhukov, Hong, and Tamer (2007) and Romano and Shaikh (2008, 2010). As with the results on Z-estimation and M-estimation, the most general results in this literature are not based on the specific structure of particular settings, like moment inequality models. Thus, the most general results tend to be based on higher-level conditions which are more general but potentially more difficult

[5]Note that $a_-$ is 0 in a component $c$ with $a_c \geq 0$ and is $a_c < 0$ in components with $a_c < 0$. Therefore, $[E_0(m(w, \theta))]_-$ is the zero vector exactly when $E_0(m(w, \theta)) \geq 0$ and otherwise $[E_0(m(w, \theta))]_-$ is a non-zero vector. Consequently, $Q_0(\theta) = 0$ exactly when $E_0(m(w, \theta)) \geq 0$.

to establish in particular settings. In practice, often the papers do work out the details for moment inequality models. Similar to above, a common issue in the literature on estimation and inference in the criterion function framework is a failure of continuity conditions, and therefore a failure of direct application of the analogy principle for estimation. In particular, even if there is an estimator $Q_n(\cdot)$ that converges to $Q_0(\cdot)$, the set $\{\theta : Q_n(\theta) = 0\}$ need not converge to $\{\theta : Q_0(\theta) = 0\}$.

Somewhat similarly, in some models it is possible to express an identified set for a partially identified parameter $\theta$ as a mapping from a (finite-dimensional) point identified parameter $\mu$. For example, in a choice model, the statistical model often takes the form $P_0(\cdot|x) = P_\theta(\cdot|x)$, where $P_0(\cdot|x)$ is the observed choice probabilities and $P_\theta(\cdot|x)$ are the choice probabilities according to the model. The identified set for $\theta$ is $\{\theta : P_0(\cdot|x) = P_\theta(\cdot|x) \text{ w.p.1. } x\}$, which can be viewed as a mapping from a point identified "parameter" $\mu$, which in this case is the stacked values of $P_0(\cdot|x)$ across $x$. More generally, it can be possible to express the identified set as $\Theta_I(\mu)$, where $\Theta_I(\mu)$ is known based on the statistical model and assumptions. This can be viewed as related to the criterion function representation, with the functional form $Q_0(\theta) = 1 - 1[\theta \in \Theta_I(\mu_0)]$. In this setting, Kline and Tamer (2016) provide results for Bayesian inference. It is also possible to do Bayesian inference in a setting with moment inequalities, see Kline (2011), Liao and Jiang (2010), Moon and Schorfheide (2012), and Liao and Simoni (2019). See also Poirier (1998), Giacomini and Kitagawa (2021), and Giacomini, Kitagawa, and Read (2022).

Finally, Chen, Christensen, and Tamer (2018) consider the question of inference with optimal criterion functions, such as likelihood or optimally-weighted GMM under partial identification. The paper proposes to use quasi-Bayes methods by building a confidence set for the identified set using simulations (such as MCMC) drawn from an appropriately constructed (pseudo) posterior. These methods are computationally simple to use and are amenable to inference on components of $\theta$, especially for a scalar parameter of interest.

3.2.1. *Identified Sets as Solutions to Linear Programs.* In some cases, the identified set can be characterized as the set of values of $\theta$ that solve a linear program (LP). A generic example of this case would be the interval outcome regression above where the support of $x$ is discrete, i.e., the case where the identified set is characterized as $\{\beta : E_0(y_L|x) \leq x\beta \leq E_0(y_U|x) \text{ w.p.1. } x\}$ where the support of $x$ is discrete and finite. This identified set has the feature that it can be viewed as the feasibility constraints in a linear program. One early example of posing the computational problem as an LP is Honoré and Tamer (2006) in the context of a nonlinear panel data model. More recently, Syrgkanis, Tamer, and Ziani (2021) used an LP as the main approach for inference on auction fundamentals and counterfactuals. For inference, a defining feature of the model is again this mapping between a finite-dimensional parameter (i.e., the stacked values of $E(y|x)$) that is point identified (and can be consistently estimated), to the identified set which is the set of parameters for which the linear program is feasible. Hence, one can use the computationally attractive Bayesian approaches in Kline and Tamer (2016). Leveraging the LP structure, Syrgkanis, Tamer, and Ziani (2021) provided (frequentist) finite sample methods to construct confidence sets for identified sets. These sets control size. For recent other approaches for this problem, see also Fang, Santos, Shaikh, and Torgovitsky (2021).

3.3. **Random Sets and Generalized Instrumental Variables.** The random sets framework for partial identification is based on the observation that "sets" are central to partial identification. This recognition goes beyond the idea that the identification result concerns a set of parameter values, to the recognition that the identification analysis itself can be usefully conducted as involving sets of random variables. The review of partial identification in Molinari (2020) emphasizes the random sets framework.

As an example, consider a "non-parametric" version of interval data. That is, suppose that the data is $w = (y_L, y_U, x)$ with the property that $y_L \leq y \leq y_U$ where $y$ is the "true" outcome variable. This setting implies that $y$ is a selection from the random set $[y_L, y_U]$. The formal definitions of terms like "selection" and "random set" can be found in sources like Molchanov (2005), Beresteanu, Molchanov, and Molinari (2012), Molchanov and Molinari (2014, 2018), and Molinari (2020). Generally, the definitions formalize intuitive ideas, an important aspect of the random sets approach. For instance, $[y_L, y_U]$ is a random set because it is a set (obviously) and random (slightly less obviously, to formalize the measurability condition). Similarly, $y$ is a selection from $[y_L, y_U]$ because each realization of $y$ is necessarily an element of (and hence a selection from) the corresponding realization of $[y_L, y_U]$.

Suppose the identification analysis concerns learning about the distribution of $y$ given $x$. The observed data reveals the distribution of the random set $[y_L, y_U]$, and so the identification analysis concerns what the distribution of $[y_L, y_U]$ implies about the distribution of $y$. A central result in the theory of random sets, the Artstein (1983) inequality, directly gives the result. The result is that the distribution of $y$ given $x$ can be any distribution that satisfies the inequality that $P(y \in K|x) \leq P_0([y_L, y_U] \cap K \neq \emptyset|x)$ for all compact sets $K$. Therefore, using the random sets framework provides the identification result almost immediately, an advantage of the random sets framework.

Intuitively, this says that the probability that $y$ is in some set $K$ can be no larger than the probability that $[y_L, y_U]$ is observed to "intersect" with that set $K$. The fact that $P(y|x)$ must satisfy this inequality is intuitive: $y \in K$ implies that $y \in [y_L, y_U] \cap K$ since always $y \in [y_L, y_U]$, so $y \in K$ implies in particular that $[y_L, y_U] \cap K \neq \emptyset$, implying the probability inequality holds.

Molinari (2020, Section 2.3) provides additional details on this specific identification analysis, and other examples of partial identification analysis in the random sets framework. Molchanov (2005) and Molchanov and Molinari (2014, 2018) provide comprehensive treatments of random sets. Beresteanu, Molchanov, and Molinari (2012) particularly focuses on the application to partial identification. Beresteanu, Molchanov, and Molinari (2011) use random set theory to derive the sharp identified set in a large class of models with set-valued outcomes.

Galichon and Henry (2011) use optimal transport theory to derive identified sets, covered comprehensively in Galichon (2018). Beresteanu and Molinari (2008) provide estimation and inference results for identified sets in the random sets framework. Galichon and Henry (2009) provide estimation and inference results for identified sets in the optimal transport framework.

The generalized instrumental variables framework uses random set theory as applied to statistical models that imply that the observed data $w = (y, z)$ can be partitioned into endogenous $y$ (quantities determined by the model) and exogenous $z$, with an unobservable $\epsilon$, that satisfies an equation $f(y, z, \epsilon) = 0$. The unknowns of the model are $f$ and the distribution $P_{\epsilon|z}$ of $\epsilon|z$. A main point of the generalized instrumental variables framework is to work with statistical models for which there is a set of values of the unobservable that are compatible with the observed data, rather than a single value as in traditional instrumental variables analysis. As such, the assumptions on how the unobservable relates to the observable (e.g., "exclusion restrictions") that are used in the identification analysis must deal with these sets. Chesher and Rosen (2020) provide a review of the results that have been derived in the generalized instrumental variables framework.

As an example, consider again the idea of interval data, except now the interval data concerns a right hand side variable. In particular, suppose that $y_1 = x\beta + \epsilon$ with $\beta \geq 0$. Suppose that $x$ is endogenous, with corresponding instrumental variable $z$. Further, suppose that $x$ is not observed, but $(x_L, x_U)$ is observed with $x_L \leq x \leq x_U$. That is, the data is $w = (y, z) = (y_1, x_L, x_U, z)$ where $y = (y_1, x_L, x_U)$ are the "endogenous" quantities and $z$ is the only "exogenous" quantity. The set of values of $y$ that are consistent with the model for a given value of $\beta$ and a given realization of $\epsilon$ is $\{y : x_L\beta+\epsilon \leq y_1 \leq x_U\beta+\epsilon\}$. Overall, this setup implies that $f(y, z, \epsilon) = 0$ for instance for $f(y, z, \epsilon) = 2-1[y_1-x_L\beta-\epsilon \geq 0]-1[x_U\beta+\epsilon-y_1 \geq 0]$.

The set of values of $\epsilon$ that are consistent with the model for a given value of $\beta$ and a given realization of $y$ is $\{\epsilon : y_1 - x_U\beta \leq \epsilon \leq y_1 - x_L\beta\}$. Therefore, $\epsilon$ must be a selection from the random set $[y_1 - x_U\beta, y_1 - x_L\beta]$.

The generalized instrumental variable framework makes it possible to use a variety of assumptions about the distribution of the unobservable. For instance, suppose that $\epsilon$ is independent of $z$. Then, again applying Artstein's inequality similarly to above, the identified set for $(\beta, P_\epsilon)$ is $\{(\beta, P_\epsilon) : P_0([y_1-x_U\beta, y_1-x_L\beta] \subseteq K|z) \leq P_\epsilon(\epsilon \in K)$ for all compact sets $K$ w.p.1. $z\}$. Note that $P_0$ is the true distribution of the observed data, whereas $P_\epsilon$ is a candidate value of the distribution of $\epsilon$. The intuition is that if $[y_1-x_U\beta, y_1-x_L\beta] \subseteq K$ (and $\epsilon \in [y_1-x_U\beta, y_1-x_L\beta]$) then it must be that $\epsilon \in K$.

Chesher and Rosen (2020) provides additional details on this specific identification analysis, and other examples of partial identification analysis in the generalized instrumental variables framework. Chesher and Rosen (2017) provide many of the main ideas. Chesher, Rosen, and Smolinski (2013) focus on discrete choice models.

Often, the identified sets from the generalized instrumental variable framework can be represented in terms of moment inequality conditions, so that estimation and inference proceeds based on the existing literature for moment inequality conditions. For instance, continuing the above example of interval data, $E(\epsilon|z)$ must be an element of $[E_0(y_1 - x_U\beta|z), E_0(y_1 - x_L\beta|z)]$. The intuition is that $\epsilon$ must be a selection from $[y_1 - x_U\beta, y_1 - x_L\beta]$, and so the (conditional) expected value must be an element of the expected value of $[y_1 - x_U\beta, y_1 - x_L\beta]$. Formally, this uses results on the Aumann expectation of a random set. Suppose the assumption is that $E(\epsilon|z) = 0$; then, the statistical model implies the conditional moment inequality conditions $E_0(y_1 - x_L\beta|z) \geq 0$ and $E_0(x_U\beta - y_1|z) \geq 0$ w.p.1. $z$.

## 4. MAIN IDEAS AND EXAMPLES

4.1. **Shape Restrictions and Sign Restrictions.** Partial identification is relevant for statistical models involving shape restrictions. A shape restriction can be viewed as restricting a quantity to be an element of some set, a circumstance that is particularly amenable to a partial identification analysis. Alternatively, a shape restriction can be viewed as a "weak" assumption that does not suffice for point identification, also a circumstance that is particularly amenable to a partial identification analysis.

Shape restrictions are commonly implied by economic theory or are suggested by informal empiricism. One common shape restriction is a sign restriction, or a monotonicity restriction: the assumption that some feature of a model entails a monotone relationship. In a parametric model, this can be that some "slope" coefficient has a particular sign. Economic theory provides many monotonicity results, like results that supply/demand curves are upward/downward sloping.

Economic theory also involves many other shape restrictions. In some cases, economic theory commonly assumes shape restrictions in order to correspond to a credible assumption on model primitives or because the absence of those shape restrictions can allow for results that are viewed as unreasonable, like the assumption of concavity of a utility function to impose risk aversion or the assumption of concavity of a production function to avoid increasing returns to scale.

Informal empiricism also provides many shape restriction claims, like claims that wages are increasing in education or that wages are concave in education (decreasing returns).

In some cases, shape restriction assumptions can be used to establish a point identification result, or improve the behavior of the finite sample estimator in a case that would be point identified even without the shape restriction. These uses have been reviewed for instance in Matzkin (1994) and Chetverikov, Santos, and Shaikh (2018).

Shape assumptions are common in potential outcomes models. In these models, each observational unit is characterized by a response function $y(d)$ which gives the outcome $y(d)$ that the unit would experience in response to treatment $d$. These models are central in the analysis of treatment effects. Generally, the observed data consists of an observed outcome $y$ and an observed treatment $x$, and potentially other quantities.

In some cases, the shape restriction is placed directly on the response function. Manski (1997) uses the assumption that the response functions are non-decreasing. Okumura and Usui (2014) use the assumption that the response functions are concave. Twinam (2017) uses the assumption that the response functions (involving multiple treatments) are supermodular. Kim, Kwon, Kwon, and Lee (2018) use the assumption that the response functions satisfy smoothness conditions. Ishihara (2021) uses the assumption that the response functions are monotone and concave in a setting with instruments with limited support. Often, multiple shape restrictions assumptions are used jointly. The potential outcomes model by itself implies that a given observational unit's response function must be in the set $\{y_i(\cdot) : y_i(x_i) = y_i\}$, since the observed outcome is the response to the observed treatment. Shape restrictions directly restrict this set further, for instance to contain only monotone response functions or concave response functions. In that way, shape restrictions can directly result in partial identification.

In other cases, the shape restriction is placed on other quantities. Manski and Pepper (2000, 2009) use the assumption of "monotone treatment selection" that the expected response $E(y(d)|x)$ is a monotone function of $x$, where $x$ is the observed treatment. Giustinelli (2011) uses a similar assumption on a quantile of the distribution of responses. Chalak (2019) uses assumptions on the sign (and magnitude) of confounding. In a regression with interval data, Manski and Tamer (2002) use the assumption of monotonicity in the variable subject to interval observation. Lee (2009) assumes that the function that relates treatment to selection is monotone in a sample selection model. Nevo and Rosen (2012) assumes the sign of the correlation between an instrumental variable and the unobservable.

Sign restrictions are also used in the identification of structural vector autoregressions, an idea developed in Faust (1998), Canova and De Nicolo (2002), and Uhlig (2005) among other parts of a large literature. See for instance Fry and Pagan (2011) and Uhlig (2017) for reviews and perspectives. Common sign restrictions concern the slope of supply and/or demand curves. Used alongside other assumptions, a typical result is that the object of interest is partially identified. The typical empirical implementation is from a Bayesian perspective, which can result in complications related to the use of Bayesian methods in partially identified models in general, as in Poirier (1998) or Moon and Schorfheide (2012). In particular, seemingly "uninformative" priors can (even asymptotically) actually have a substantial impact on the empirical results. Results for (Bayesian or classical) inference in such settings has been developed in Rubio-Ramirez, Waggoner, and Zha (2010), Baumeister and Hamilton (2015), Granziera, Moon, and Schorfheide (2018), Arias, Rubio-Ramírez, and Waggoner (2018), Gafarov, Meier, and Olea (2018), Giacomini and Kitagawa (2021), and Giacomini, Kitagawa, and Read (2022), as representative of a larger literature.

In other cases, the identification result rather than the assumptions concerns the sign of a parameter, or whether the parameter satisfies some other shape restriction. Given any identification result for a parameter $\theta$, it is possible to use the identification result to check what can be learned about the sign of $\theta$, or indeed what can be learned about other qualitative features of $\theta$. For instance, if the identified set for a scalar $\theta$ contains only non-negative values, then it is possible to conclude that $\theta$ is non-negative. This can be particularly relevant if $\theta$ represents a treatment effect. Even if an identified set used in an empirical analysis is not sharp, it may still be possible to make "sharp" conclusions about qualitative features of the parameter; for instance, an identified set may not be sharp, and yet contain only non-negative values of the parameter, so the sign of the parameter is "sharply" identified to be non-negative.

From the point of view of conducting an identification analysis, it is worth noting that it may be difficult to think through a (partial) identification result for the underlying model parameter, but easier (in some sense) to think through the logic of an identification result directly for the sign of the parameter, or some other qualitative feature of the parameter. Alternatively, it may be useful to characterize features of the data that imply a parameter has a particular qualitative feature, for instance because of ease of interpretation of those features of the data. For instance, Manski and Tamer (2002), Abrevaya, Hausman, and Khan (2010), Kline (2016b), and Kline and Tamer (2018) have results that emphasize the identification of the sign of a parameter of interest. A similar question also arises in the context of identification of the sign of an interaction effect in a model with multiple decision makers, discussed below.

4.2. **Optimizing Behavior, Equilibrium Behavior, and Revealed Preference.** Partial identification is relevant for statistical models involving decision makers making choices with optimizing behavior or equilibrium behavior. Data from such models can be viewed as providing revealed preference information.

For a simplified unifying framework, suppose that a choice $y$ is made from among the potential options in a set $\mathcal{Y}$. This choice could be made by an individual decision maker, or this choice could be made "jointly" (in some way that is specified by a model) by multiple decision makers. Each decision maker might make a single choice, or multiple choices. Suppose that the decision maker(s) are assumed to have utility function(s) $u$ from a space $\mathcal{U}$. If there are multiple decision makers, then $u$ is a vector of the utility functions across the decision makers. Suppose that the empirical researcher has a "theory" $\mathcal{C}_{\mathcal{Y}} : \mathcal{U} \to 2^{\mathcal{Y}}$, interpreted as a generalized "choice correspondence," which implies that the choice from $\mathcal{Y}$ given a particular $u$ must be in the set $\mathcal{C}_{\mathcal{Y}}(u)$. This "choice correspondence" could be utility maximization in the case of a single decision maker, or a solution concept for a game in the case of multiple decision makers (e.g., the set of Nash equilibrium outcomes for given utility functions).

Consider then a stylized identification analysis: What can be learned about $u$ (or the distribution thereof) based on observed choices (or the distribution thereof)? As a starting point, a central idea is that for any observed choice $y$, the corresponding $u$ is "partially identified" (used informally here) to be in $\mathcal{C}_{\mathcal{Y}}^{-1}(y) = \{u : y \in \mathcal{C}_{\mathcal{Y}}(u)\}$, the set of $u$ such that $y$ could be chosen for that $u$ given the potential options in $\mathcal{Y}$. Since in general there can be multiple utility functions that result in the choice $y$, there is "partial identification" of the utility function. Often, as detailed below, the result is that observation of a particular choice $y$ implies that the corresponding utility function satisfies a particular set of inequality constraints.

4.2.1. *Single Decision Maker.* The main idea that observed choice of a single decision maker can be used to learn about utility or other fundamental objects is due to Samuelson (1938), with an important subsequent literature. In many cases, this idea is used in economic theory. This idea is also central to a related literature in identification and econometrics. Crawford and De Rock (2014) and Echenique (2020) review revealed preference.

As a simplified but unifying example, Manski (2009, Section 7.5) considers identification in a potential outcomes framework under the assumption that the decision maker makes a choice that maximizes the outcome. This implies the partial identification result that, for each decision maker, each other choice results in a potential outcome that is weakly less than the observed (maximized) outcome. By the same logic, if each decision maker has a utility function $u(y, x, \epsilon)$ that depends on the choice $y$, other observed determinants of utility $x$ ("explanatory variables"), and unobservable determinants of utility $\epsilon$, then the assumption of utility maximization implies that the observed choice $y_i$ will satisfy the restriction $u_i(y_i, x_i, \epsilon_i) \geq u_i(t, x_i, \epsilon_i)$ for all possible alternative choices $t$. Therefore, an observed choice $y$ implies that the utility function satisfies these inequality constraints. Specifically, for each possible choice $\tilde{y}$, it must be that $P_0(y = \tilde{y}|x) \leq \tilde{P}(u(\tilde{y}, x, \epsilon) \geq u(t, x, \epsilon) \text{ for all } t \in \mathcal{Y}|x)$ w.p.1. $x$. Here, $\tilde{P}$ is a distribution for $\epsilon|x$, which might be assumed to be known or which might be an unknown parameter of the statistical model. Thus, as illustrated here to make a more general point, identification strategies based on the assumption of optimizing behavior (revealed preference) are likely to result in partial identification.

Because of the centrality of optimizing behavior in economics, such revealed preference ideas have been the source of partial identification results in a series of specific settings in Blundell, Browning, and Crawford (2008), Pakes (2010), Blundell, Kristensen, and Matzkin (2014), Manski (2014), Cherchye, De Rock, Lewbel, and Vermeulen (2015), Pakes, Porter, Ho, and Ishii (2015), Barseghyan, Molinari, and Teitelbaum (2016), Hausman and Newey (2016), Cherchye, Demuynck, and De Rock (2019), Cattaneo, Ma, Masatlioglu, and Suleymanov (2020), Mourifie, Henry, and Meango (2020), Barseghyan, Coughlin, Molinari, and Teitelbaum (2021), and Allen and Rehbeck (2022) among other examples. The settings range from demand models, to models with limited attention, to Roy models, among other settings. Although not all of these results fit directly into the basic outline above, the basic outline above suggests why revealed preference arguments result in partial identification in general. In some instances, the identification analysis concerns objects of interest concerning optimizing behavior itself, as in tests of "consumer rationality," with specific results in Hoderlein and Stoye (2014), Kitamura and Stoye (2018), and Aguiar and Kashaev (2021) among other examples. In early results, Marschak and Andrews (1944) and Leamer (1981) use economic theory to develop results that partially identify the parameter of a production function and/or supply and demand functions.

4.2.2. *Multiple Decision Makers.* Statistical models involving multiple decision makers (e.g., "games") are commonly used to illustrate new ideas in the partial identification literature, and many of the frameworks papers cited above discuss the applications to this setting. Different aspects of this literature, often with reference also to point identification results, have been the focuses of reviews in de Paula (2013), Aradillas-López (2020), Kline and Tamer (2020), and Kline, Pakes, and Tamer (2021).

An important consideration in an identification analysis of a model involving multiple decision makers is the assumption about the solution concept. The solution concept provides the theory for how the interaction among the decision makers "works" and can include the assumption of Nash equilibrium ("mutual best response") or weaker assumptions like iterated best response or level-$k$ rationality. Different solution concepts imply different "choice correspondences." The basic idea of Nash equilibrium is that each decision maker "best responds" to the other decision makers, based on what the decision maker knows. Other solution concepts are based on other descriptions of how decision makers behave. Every solution concept must address the fact that the decision makers cannot simply maximize utility because they do not know the choices of the other decision makers when they make their choice (at least in a simultaneous move game).

Suppose for instance the utility function for decision maker $i$ is of the form $u_i(y, x, \epsilon)$, where $y = (y_i, y_{-i})$ is the vector of choices for all decision makers, $x$ are observed determinants of utility ("explanatory variables"), and $\epsilon$ are unobservable determinants of utility. The decision makers can be individual people, in which case this basic model reflects the idea of "peer effects" or "social interactions," whereby the choices of some people have an impact on the utility of other people. The decision makers can be firms, with utility understood to be profit, in which case this basic model reflects ideas like "competitive effects," whereby the choices of some firms influence the profits of other firms.

If $\epsilon$ is common knowledge among the decision makers, as in a complete information game, and the decision makers have no ex post regret as in a pure strategy Nash equilibrium (so that observed choices are best responses to observed choices, rather than to a mixture over choices as in a mixed strategy Nash equilibrium), then the choice of each decision maker $y_i$ will satisfy the restriction that $u_i(y_i, y_{-i}, x, \epsilon) \geq u_i(t_i, y_{-i}, x, \epsilon)$ for all possible alternative choices $t_i$. Consequently, an observation of $y$ implies that the utility function satisfies these inequality constraints. Specifically, for each possible vector of choices $\tilde{y}$, it must be that $P_0(y = \tilde{y}|x) \leq \tilde{P}(u_i(\tilde{y}_i, \tilde{y}_{-i}, x, \epsilon) \geq u_i(t_i, \tilde{y}_{-i}, x, \epsilon)$ for all $i$ and for all $t_i|x)$ w.p.1. $x$. Here, $\tilde{P}$ is a distribution for $\epsilon|x$, which might be assumed to be known or which might be an unknown parameter of the statistical model. It is worth noting that a pure strategy Nash equilibrium may not even exist in some games, and so the "assumption" that observed choices come from a pure strategy Nash equilibrium can be provably false in some games. In other words, although these inequality constraints are intuitively appealing properties of a choice, the question of whether such a choice actually exists with multiple decision makers is related to the existence of a pure strategy Nash equilibrium. In a model with a single decision maker, the "assumption" that a utility maximizing choice exists is much more innocuous in general.

Different information conditions on $\epsilon$ or different solution concepts would mean that a given observed choice would have different logical implications for the utility functions. From the point of view of identification analysis, the general idea is that economic theory can be used to conclude that observation of a particular vector of choices $\tilde{y}$ implies that the utility functions must have satisfied certain restrictions in order to have possibly generated that choice. This is a source of partial identification of the utility functions.

For a variety of specific models, and with a variety of specific identification strategies, partial identification results have been derived in Aradillas-López and Tamer (2008), Ciliberto and Tamer (2009), Grieco (2014), Aradillas-López and Gandhi (2016), Ciliberto, Murry, and Tamer (2021), and Aradillas-López and Rosen (2022). Further, this setting is a common example used in the frameworks papers mentioned above. Of course, under sufficient conditions, (features of) the utility functions are point identified, as in Tamer (2003), Aradillas-López (2010, 2012), Bajari, Hong, Krainer, and Nekipelov (2010), Bajari, Hong, and Ryan (2010), Wan and Xu (2014), Xu (2014), Kline (2015, 2016a), Liu, Vuong, and Xu (2017), and Xiao (2018). Often, the point identification results rely on support conditions for the observable determinants of utility, or assumptions on the selection mechanism (how a choice is realized when $|\mathcal{C}(u)| \geq 2$), that are not used in the partial identification results.

This approach to identification analysis accommodates the possibility that the model does not make a unique prediction about the choice that results from a given utility function, as in "multiple equilibria." In the "choice correspondence" setup this means the identification analysis accommodates the possibility that $|\mathcal{C}(u)| \geq 2$ for some $u$. This is because the identification analysis is based on what an observed choice logically implies about the utility function, but not what a utility function implies about the resulting choice. The latter would involve dealing with multiple choices potentially arising from a given utility function.

It is possible to conduct an identification analysis for other features of the model, including specific features of the utility functions. For instance, de Paula and Tang (2012) for incomplete information and Kline (2016a) for complete information establish how it is possible to identify whether the utility functions are increasing ("strategic complements") or decreasing ("strategic substitutes") functions of the choices of the other decision makers, under particularly weak assumptions. This can be viewed as partial identification of the utility function, and specifically partial identification of the interaction effect feature of the utility function, where the identified sets are either the utility functions with positive interaction effects or the utility functions with negative interaction effects. Aradillas-López (2011) partially identifies the probability that a given outcome is a Nash equilibrium. Kline and Tamer (2012) partially identifies the best response functions.

Also, partial identification approaches have been used to examine the question of robustness to information structures. In most empirical work in games, strong assumptions are made on what the decision makers know about their own utility functions and also about the utility functions of the other decision makers. For instance, most work on auctions makes assumptions on what bidders know about the value of the object and what other bidders know. Also, bidders obtain signals that may be informative about valuations and modeling these signals affects the identification analysis. Building on the recent progress in the mechanism design literature (see for instance Bergemann and Morris (2013)), Syrgkanis, Tamer, and Ziani (2021) characterize the identified sets for valuation distributions, welfare counterfactuals and other relevant objects in auctions that are robust to information structures. These identified sets are solutions to linear programs and hence are simple to compute. See also a similar exercise in the context of entry games in Magnolfi and Roncoroni (2022).

In many applications, each decision maker has a single choice. In empirical research, the specific choices that are analyzed often parallel the sorts of choices typically analyzed by standard discrete choice models but allowing for the possibility of spillovers. However, in other applications, each decision maker has multiple choices. A particularly important example is the case of network formation models.

In a typical network formation model, the choices of each pair of decision makers determines whether or not that pair is connected in the network. Thus, each decision maker has multiple choices, corresponding to potentially connecting to each of the other decision makers. A connection can be a friendship, some other social relationship, or some sort of economic relationship. In these models, the solution concept of Nash equilibrium is often viewed as unreasonable. For instance, Nash equilibrium can allow for the possibility that a pair is not connected despite a connection being mutually beneficial (and therefore "expected" to form), in that both decision makers best respond to the other in the pair having "not connected" by "not connecting" in return. Therefore, alternative solution concepts are used, but the basic logic of the identification strategy can follow that outlined above.

For instance, the solution concept of pairwise stability from Jackson and Wolinsky (1996) requires that: if a connection exists between two decision makers, then both get utility with that connection that is weakly greater than the utility they would get without that connection; and, if a connection does not exist between two decision makers, then at least one would get less utility with that connection compared to the utility without that connection. In other words, each observed choice implies that the utility functions must have satisfied a certain set of inequality constraints. As with the example of pure strategy Nash equilibrium, this can be used to derive a partial identification result for the utility functions. For more on these solution concepts, see for instance Jackson (2010, Chapter 6). Related partial identification results have been derived in Miyauchi (2016), de Paula, Richards-Shubik, and Tamer (2018), Sheng (2020), and Gualdani (2021).

There are also non-strategic network formation models that emphasize different considerations, and de-emphasize the choice considerations emphasized above. Overall, network formation models have been a focus of reviews in Graham (2015, 2020a,b), Chandrasekhar (2016), de Paula (2017, 2020a,b), and Bonhomme (2020).

4.3. **Misspecification and Sensitivity Analysis.** This review primarily takes the perspective of an identification analysis that "believes" the assumptions. Another important perspective within the partial identification literature concerns the case of misspecification analysis or sensitivity analysis.

In one version of this perspective, the empirical researcher conducts a baseline analysis under a certain set of assumptions, and then investigates the "sensitivity" of the empirical results to the assumptions. This can result in a sort of partial identification result with respect to the relaxed assumptions. Many partial identification results can be viewed from either of these perspectives. Partial identification results that intentionally take some version of this latter perspective include Kline and Santos (2013) and Masten and Poirier (2018, 2020, 2021). Assumptions that are relaxed vary from assumptions about missing data to exogeneity assumptions to exclusion restrictions.

A closely related question concerns the behavior of empirical results when the assumptions used in a partial identification result themselves are wrong. Ponomareva and Tamer (2011) show that misspecification can essentially be a "source" of identification, in the sense that estimation of a misspecified model can result in the false appearance of learning more about the parameter than actually is possible.

4.4. **Data Limitations and Semiparametric Models.** This review primarily focuses on cases where the assumptions are the focus of the partial identification analysis. However, there are many important cases where the data is the focus of the partial identification analysis. Obviously, the identification result depends on both the assumptions and the data, so this is a somewhat arbitrary distinction.

Partial identification analysis is commonly used in cases of missing data or mismeasured data or contaminated data or interval data or combined data, with examples in Manski (1989, 2003, 2009), Horowitz and Manski (1995), Manski and Tamer (2002), Khan and Tamer (2009), Stoye (2010), Fan, Sherman, and Shum (2014), and Khan, Ponomareva, and Tamer (2016) among other contributions to this large part of the partial identification literature. Often, the results are in cases where the absence of these specific data limitations would imply the point identification of the parameter of interest. In that sense, the focus of the partial identification analysis is on the data. In fact, some of the earliest examples of partial identification results in Gini (1921) and Frisch (1934), among others, are for a linear model with measurement error. Also an early example, Hoeffding (1940) and Fréchet (1951) develop results that partially identify a joint distribution for given marginal distributions, which can be viewed as a sort of data combination question.

One way to think about these settings is that there could be "complete" data that, if observed by the empirical researcher, would suffice for point identification. The "complete" data would not be subject to missing data, or measurement error, or other data limitations. However, the observed data is subject to data limitations. As such, the observed data is compatible with multiple "complete" data. For instance, observed data subject to missing data is compatible with any "complete" data that can result from "filling in" the missing data with particular values. Consequently, there is partial identification. The partial identification result accounts for the fact that the distribution of the "complete" data cannot be uniquely determined by the distribution of the observed data.

In particular, identification in semiparametric models is often based on conditions on the data that may not be credible. Consider the binary choice model where $y = 1[\epsilon \leq x\beta]$ and $(y, x)$ is observed. A sufficient condition for point identification of $\beta$ under a (conditional) median independence assumption (i.e., $\text{Med}(\epsilon|x) = 0$) is the existence of an explanatory variable with large support as in Manski (1975). Without this condition on the data, for instance allowing for the case that $x$ takes on finitely many values, the median independence assumption can be used to construct the identified set for $\beta$. This identified set uses the median independence assumption to get the representation $\{\beta : x\beta \geq 0 \text{ iff } P(y = 1|x) \geq \frac{1}{2} \text{ w.p.1. } x\}$ as in Komarova (2013). Also, in discrete outcomes models with fixed effects that includes dynamic models, there is a set of results that derive identified sets on slope parameters relaxing the distributional assumptions that are typically used, such as the logit restriction on the unobservables. See for instance Khan, Ponomareva, and Tamer (2020) that construct the sharp identified set under stationary errors and Pakes, Porter, Shepard, and Calder-Wang (2021) that construct identified sets under mean restrictions. For bounds on partial effects in discrete choice models see Chernozhukov, Fernández-Val, Hahn, and Newey (2013) and Torgovitsky (2019). For results on identification in censored models (that can allow for endogeneity) in panel models, see for instance Khan, Ponomareva, and Tamer (2016).

## 5. CONCLUSIONS

The partial identification approach emphasizes an approach to empirical research that begins with data, a set of assumptions (motivated by an economic model, for example), and a statistical model, and then asks what can be logically concluded about the value of the parameter that is relevant for answering an empirical question. Practically, it allows for empirical research in settings in which it is not possible to learn the exact value of the parameter (as in point identification), and therefore the partial identification approach substantially expands the universe of empirical questions that can be addressed. The partial identification approach substantially increases the credibility of empirical research, because it eliminates the need to make not-credible assumptions in order to fit an empirical analysis into a particular identification result. Unlike the point identification framework that necessarily only applies to sets of assumptions that result in point identification, the partial identification framework applies to any set of assumptions.

Partial identification still requires the empirical researcher to use a set of assumptions. The assumptions to use and how far to weaken these assumptions is an important question that empirical research either in the partial identification or point identification approaches needs to take a stand on. Standard models are sets of assumptions that the community of researchers are comfortable with. The strength of the partial identification approach is that it is robust to the distributional, functional form, and other assumptions on which economic theory is silent and where existing data are not helpful.

## References


Abrevaya, J., J. A. Hausman, AND S. Khan (2010): "Testing for causal effects in a generalized regression model with endogenous regressors," *Econometrica*, 78(6), 2043–2061.

Aguiar, V. H., AND N. Kashaev (2021): "Stochastic revealed preferences with measurement error," *The Review of Economic Studies*, 88(4), 2042–2093.

Allen, R., AND J. Rehbeck (2022): "Latent complementarity in bundles models," *Journal of Econometrics*, 228(2), 322–341.

Aradillas-López, A. (2010): "Semiparametric estimation of a simultaneous game with incomplete information," *Journal of Econometrics*, 157(2), 409–431.

Aradillas-López, A. (2011): "Nonparametric probability bounds for Nash equilibrium actions in a simultaneous discrete game," *Quantitative Economics*, 2(2), 135–171.

Aradillas-López, A. (2012): "Pairwise-difference estimation of incomplete information games," *Journal of Econometrics*, 168(1), 120–140.

Aradillas-López, A. (2020): "The econometrics of static games," *Annual Review of Economics*, 12, 135–165.

Aradillas-López, A., AND A. Gandhi (2016): "Estimation of games with ordered actions: An application to chain-store entry," *Quantitative Economics*, 7(3), 727–780.

Aradillas-López, A., AND A. M. Rosen (2022): "Inference in ordered response games with complete information," *Journal of Econometrics*, 226(2), 451–476.

Aradillas-López, A., AND E. Tamer (2008): "The identification power of equilibrium in simple games," *Journal of Business & Economic Statistics*, 26(3), 261–283.

Arias, J. E., J. F. Rubio-Ramírez, AND D. F. Waggoner (2018): "Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications," *Econometrica*, 86(2), 685–720.

Artstein, Z. (1983): "Distributions of random sets and random selections," *Israel Journal of Mathematics*, 46(4), 313–324.

Bajari, P., H. Hong, J. Krainer, AND D. Nekipelov (2010): "Estimating static models of strategic interactions," *Journal of Business & Economic Statistics*, 28(4), 469–482.

Bajari, P., H. Hong, AND S. P. Ryan (2010): "Identification and estimation of a discrete game of complete information," *Econometrica*, 78(5), 1529–1568.

Barseghyan, L., M. Coughlin, F. Molinari, AND J. C. Teitelbaum (2021): "Heterogeneous choice sets and preferences," *Econometrica*, 89(5), 2015–2048.

Barseghyan, L., F. Molinari, AND J. C. Teitelbaum (2016): "Inference under stability of risk preferences," *Quantitative Economics*, 7(2), 367–409.

Baumeister, C., AND J. D. Hamilton (2015): "Sign restrictions, structural vector autoregressions, and useful prior information," *Econometrica*, 83(5), 1963–1999.

Beresteanu, A., I. Molchanov, AND F. Molinari (2011): "Sharp identification regions in models with convex moment predictions," *Econometrica*, 79(6), 1785–1821.

Beresteanu, A., I. Molchanov, AND F. Molinari (2012): "Partial identification using random set theory," *Journal of Econometrics*, 166(1), 17–32.

Beresteanu, A., AND F. Molinari (2008): "Asymptotic properties for a class of partially identified models," *Econometrica*, 76(4), 763–814.

Bergemann, D., AND S. Morris (2013): "Robust predictions in games with incomplete information," *Econometrica*, 81(4), 1251–1308.

Blundell, R., M. Browning, AND I. Crawford (2008): "Best nonparametric bounds on demand responses," *Econometrica*, 76(6), 1227–1262.

Blundell, R., D. Kristensen, AND R. Matzkin (2014): "Bounding quantile demand functions using revealed preference inequalities," *Journal of Econometrics*, 179(2), 112–127.

Bonhomme, S. (2020): "Econometric analysis of bipartite networks," in *The Econometric Analysis of Network Data*, ed. by B. Graham, and A. de Paula, chap. 5, pp. 83–121. Elsevier.

Bontemps, C., AND T. Magnac (2017): "Set identification, moment restrictions, and inference," *Annual Review of Economics*, 9, 103–129.

Canay, I. A., AND A. M. Shaikh (2017): "Practical and theoretical advances in inference for partially identified models," in *Advances in Economics and Econometrics*, ed. by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, vol. 2, chap. 9, pp. 271–306. Cambridge University Press.

Canova, F., AND G. De Nicolo (2002): "Monetary disturbances matter for business fluctuations in the G-7," *Journal of Monetary Economics*, 49(6), 1131–1159.

Cattaneo, M. D., X. Ma, Y. Masatlioglu, AND E. Suleymanov (2020): "A random attention model," *Journal of Political Economy*, 128(7), 2796–2836.

Chalak, K. (2019): "Identification of average effects under magnitude and sign restrictions on confounding," *Quantitative Economics*, 10(4), 1619–1657.

Chandrasekhar, A. (2016): "Econometrics of network formation," in *The Oxford Handbook of the Economics of Networks*, ed. by Y. Bramoullé, A. Galeotti, and B. W. Rogers, chap. 13, pp. 303–357. Oxford University Press.

Chen, X., T. M. Christensen, and E. Tamer (2018): "Monte Carlo confidence sets for identified sets," *Econometrica*, 86(6), 1965–2018.

Cherchye, L., B. De Rock, A. Lewbel, and F. Vermeulen (2015): "Sharing rule identification for general collective consumption models," *Econometrica*, 83(5), 2001–2041.

Cherchye, L., T. Demuynck, and B. De Rock (2019): "Bounding counterfactual demand with unobserved heterogeneity and endogenous expenditures," *Journal of Econometrics*, 211(2), 483–506.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and quantile effects in nonseparable panel models," *Econometrica*, 81(2), 535–580.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and confidence regions for parameter sets in econometric models," *Econometrica*, 75(5), 1243–1284.

Chesher, A., and A. M. Rosen (2017): "Generalized instrumental variable models," *Econometrica*, 85(3), 959–989.

Chesher, A., and A. M. Rosen (2020): "Generalized instrumental variable models, methods, and applications," in *Handbook of Econometrics*, ed. by S. N. Durlauf, L. P. Hansen, J. J. Heckman, and R. L. Matzkin, vol. 7, chap. 1, pp. 1–110. Elsevier.

Chesher, A., A. M. Rosen, and K. Smolinski (2013): "An instrumental variable model of multiple discrete choice," *Quantitative Economics*, 4(2), 157–196.

Chetverikov, D., A. Santos, and A. M. Shaikh (2018): "The econometrics of shape restrictions," *Annual Review of Economics*, 10, 31–63.

Ciliberto, F., C. Murry, and E. Tamer (2021): "Market structure and competition in airline markets," *Journal of Political Economy*, 129(11), 2995–3038.

Ciliberto, F., and E. Tamer (2009): "Market structure and multiple equilibria in airline markets," *Econometrica*, 77(6), 1791–1828.

Crawford, I., and B. De Rock (2014): "Empirical revealed preference," *Annual Review of Economics*, 6, 503–524.

Echenique, F. (2020): "New developments in revealed preference theory: decisions under risk, uncertainty, and intertemporal choice," *Annual Review of Economics*, 12, 299–316.

Fan, Y., R. Sherman, and M. Shum (2014): "Identifying treatment effects under data combination," *Econometrica*, 82(2), 811–822.

Fang, Z., A. Santos, A. M. Shaikh, and A. Torgovitsky (2021): "Inference for large-scale linear systems with known coefficients."

Faust, J. (1998): "The robustness of identified VAR conclusions about money," *Carnegie-Rochester Conference Series on Public Policy*, 49, 207–244.

Fréchet, M. (1951): "Sur les tableaux de corrélation dont les marges sont données," *Ann. Univ. Lyon Sect. A*, 14, 53–77.

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

Fry, R., and A. Pagan (2011): "Sign restrictions in structural vector autoregressions: A critical review," *Journal of Economic Literature*, 49(4), 938–60.

Gafarov, B., M. Meier, and J. L. M. Olea (2018): "Delta-Method inference for a class of set-identified SVARs," *Journal of Econometrics*, 203(2), 316–327.

Galichon, A. (2018): *Optimal Transport Methods in Economics*. Princeton University Press.

Galichon, A., AND M. Henry (2009): "A test of non-identifying restrictions and confidence regions for partially identified parameters," *Journal of Econometrics*, 152(2), 186–196.

Galichon, A., AND M. Henry (2011): "Set identification in models with multiple equilibria," *The Review of Economic Studies*, 78(4), 1264–1298.

Giacomini, R., AND T. Kitagawa (2021): "Robust Bayesian inference for set-identified models," *Econometrica*, 89(4), 1519–1556.

Giacomini, R., T. Kitagawa, AND M. Read (2022): "Robust Bayesian inference in proxy SVARs," *Journal of Econometrics*, 228(1), 107–126.

Gini, C. (1921): "Sull'interpolazione di una retta quando i valori della variabile indipendente sono affetti da errori accidentali," *Metron*, 1(3), 63–82.

Giustinelli, P. (2011): "Non-parametric bounds on quantiles under monotonicity assumptions: with an application to the Italian education returns," *Journal of Applied Econometrics*, 26(5), 783–824.

Graham, B. S. (2015): "Methods of identification in social networks," *Annual Review of Economics*, 7, 465–485.

Graham, B. S. (2020a): "Dyadic regression," in *The Econometric Analysis of Network Data*, ed. by B. Graham, and A. de Paula, chap. 2, pp. 23–40. Elsevier.

Graham, B. S. (2020b): "Network data," in *Handbook of Econometrics*, ed. by S. N. Durlauf, L. P. Hansen, J. J. Heckman, and R. L. Matzkin, vol. 7, chap. 2, pp. 111–218. Elsevier.

Granziera, E., H. R. Moon, AND F. Schorfheide (2018): "Inference for VARs identified with sign restrictions," *Quantitative Economics*, 9(3), 1087–1121.

Grieco, P. L. (2014): "Discrete games with flexible information structures: An application to local grocery markets," *The RAND Journal of Economics*, 45(2), 303–340.

Gualdani, C. (2021): "An econometric model of network formation with an application to board interlocks between firms," *Journal of Econometrics*, 224(2), 345–370.

Hausman, J. A., AND W. K. Newey (2016): "Individual heterogeneity and average welfare," *Econometrica*, 84(3), 1225–1248.

Ho, K., AND A. M. Rosen (2017): "Partial Identification in applied research: benefits and challenges," in *Advances in Economics and Econometrics*, ed. by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, vol. 2, chap. 10, pp. 307–359. Cambridge University Press.

Hoderlein, S., AND J. Stoye (2014): "Revealed preferences in a heterogeneous population," *Review of Economics and Statistics*, 96(2), 197–213.

Hoeffding, W. (1940): *Masstabinvariante korrelationstheorie*. Schriften des Mathematischen Instituts und Instituts fur Angewandte Mathematik der Universitat Berlin.

Honoré, B. E., AND E. Tamer (2006): "Bounds on parameters in panel dynamic discrete choice models," *Econometrica*, 74(3), 611–629.

Horowitz, J. L., AND C. F. Manski (1995): "Identification and robustness with contaminated and corrupted data," *Econometrica*, 63(2), 281–302.

Ishihara, T. (2021): "Partial identification of nonseparable models using binary instruments," *Econometric Theory*, 37(4), 817–848.

Jackson, M. O. (2010): *Social and Economic Networks*. Princeton University Press.

Jackson, M. O., AND A. Wolinsky (1996): "A strategic model of social and economic networks," *Journal of Economic Theory*, 71(1), 44–74.

Khan, S., M. Ponomareva, AND E. Tamer (2016): "Identification of panel data models with endogenous censoring," *Journal of Econometrics*, 194(1), 57–75.

Khan, S., M. Ponomareva, AND E. Tamer (2020): "Identification of dynamic binary response models."

Khan, S., AND E. Tamer (2009): "Inference on endogenously censored regression models using conditional moment inequalities," *Journal of Econometrics*, 152(2), 104–119.

Kim, W., K. Kwon, S. Kwon, AND S. Lee (2018): "The identification power of smoothness assumptions in models with counterfactual outcomes," *Quantitative Economics*, 9(2), 617–642.

Kitamura, Y., AND J. Stoye (2018): "Nonparametric analysis of random utility models," *Econometrica*, 86(6), 1883–1909.

Kline, B. (2011): "The Bayesian and frequentist approaches to testing a one-sided hypothesis about a multivariate mean," *Journal of Statistical Planning and Inference*, 141(9), 3131–3141.

Kline, B. (2015): "Identification of complete information games," *Journal of Econometrics*, 189(1), 117–131.

Kline, B. (2016a): "The empirical content of games with bounded regressors," *Quantitative Economics*, 7(1), 37–81.

Kline, B. (2016b): "Identification of the direction of a causal effect by instrumental variables," *Journal of Business & Economic Statistics*, 34(2), 176–184.

Kline, B., A. Pakes, AND E. Tamer (2021): "Moment inequalities and partial identification in industrial organization," in *Handbook of Industrial Organization*, ed. by K. Ho, A. Hortaçsu, and A. Lizzeri, vol. 4, chap. 5, pp. 345–431. Elsevier.

Kline, B., AND E. Tamer (2012): "Bounds for best response functions in binary games," *Journal of Econometrics*, 166(1), 92–105.

Kline, B., AND E. Tamer (2016): "Bayesian inference in a class of partially identified models," *Quantitative Economics*, 7(2), 329–366.

Kline, B., AND E. Tamer (2018): "Identification of treatment effects with selective participation in a randomized trial," *The Econometrics Journal*, 21(3), 332–353.

Kline, B., AND E. Tamer (2020): "Econometric analysis of models with social interactions," in *The Econometric Analysis of Network Data*, ed. by B. Graham, and A. de Paula, chap. 7, pp. 149–181. Elsevier.

Kline, P., AND A. Santos (2013): "Sensitivity to missing data assumptions: Theory and an evaluation of the US wage structure," *Quantitative Economics*, 4(2), 231–267.

Komarova, T. (2013): "Binary choice models with discrete regressors: Identification and misspecification," *Journal of Econometrics*, 177(1), 14–33.

Leamer, E. E. (1981): "Is it a demand curve, or is it a supply curve? Partial identification through inequality constraints," *The Review of Economics and Statistics*, 63(3), 319–327.

Lee, D. S. (2009): "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *The Review of Economic Studies*, 76(3), 1071–1102.

Lewbel, A. (2019): "The identification zoo: Meanings of identification in econometrics," *Journal of Economic Literature*, 57(4), 835–903.

Liao, Y., and W. Jiang (2010): "Bayesian analysis in moment inequality models," *The Annals of Statistics*, 38(1), 275–316.

Liao, Y., and A. Simoni (2019): "Bayesian inference for partially identified smooth convex models," *Journal of Econometrics*, 211(2), 338–360.

Liu, N., Q. Vuong, and H. Xu (2017): "Rationalization and identification of binary games with correlated types," *Journal of Econometrics*, 201(2), 249–268.

Magnolfi, L., and C. Roncoroni (2022): "Estimation of discrete games with weak assumptions on information."

Manski, C. F. (1975): "Maximum score estimation of the stochastic utility model of choice," *Journal of Econometrics*, 3(3), 205–228.

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

Manski, C. F. (1990): "Nonparametric bounds on treatment effects," *The American Economic Review*, 80(2), 319–323.

Manski, C. F. (1997): "Monotone treatment response," *Econometrica*, 65(6), 1311–1334.

Manski, C. F. (2003): *Partial Identification of Probability Distributions*, Springer Series in Statistics. Springer.

Manski, C. F. (2009): *Identification for Prediction and Decision*. Harvard University Press.

Manski, C. F. (2014): "Identification of income–leisure preferences and evaluation of income tax policy," *Quantitative Economics*, 5(1), 145–174.

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

Manski, C. F., and J. V. Pepper (2009): "More on monotone instrumental variables," *The Econometrics Journal*, 12, S200–S216.

Manski, C. F., and E. Tamer (2002): "Inference on regressions with interval data on a regressor or outcome," *Econometrica*, 70(2), 519–546.

Marschak, J., and W. H. Andrews (1944): "Random simultaneous equations and the theory of production," *Econometrica*, 12(3/4), 143–205.

Masten, M. A., and A. Poirier (2018): "Identification of treatment effects under conditional partial independence," *Econometrica*, 86(1), 317–351.

Masten, M. A., and A. Poirier (2020): "Inference on breakdown frontiers," *Quantitative Economics*, 11(1), 41–111.

Masten, M. A., and A. Poirier (2021): "Salvaging falsified instrumental variable models," *Econometrica*, 89(3), 1449–1469.

Matzkin, R. L. (1994): "Restrictions of economic theory in nonparametric methods," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, chap. 42, pp. 2523–2558. Elsevier.

Matzkin, R. L. (2007): "Nonparametric identification," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6B, chap. 73, pp. 5307–5368. Elsevier.

Matzkin, R. L. (2013): "Nonparametric identification in structural economic models," *Annual Review of Economics*, 5(1), 457–486.

Miyauchi, Y. (2016): "Structural estimation of pairwise stable networks with nonnegative externality," *Journal of Econometrics*, 195(2), 224–235.

Molchanov, I. (2005): *Theory of Random Sets*, Probability and Its Applications. Springer.

Molchanov, I., and F. Molinari (2014): "Applications of random set theory in econometrics," *Annual Review of Economics*, 6(1), 229–251.

Molchanov, I., and F. Molinari (2018): *Random Sets in Econometrics*, vol. 60 of *Econometric Society Monographs*. Cambridge University Press.

Molinari, F. (2020): "Microeconometrics with partial identification," in *Handbook of Econometrics*, ed. by S. N. Durlauf, L. P. Hansen, J. J. Heckman, and R. L. Matzkin, vol. 7, chap. 5, pp. 355–486. Elsevier.

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

Mourifie, I., M. Henry, and R. Meango (2020): "Sharp bounds and testability of a Roy model of STEM major choices," *Journal of Political Economy*, 128(8), 3220–3283.

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

Okumura, T., and E. Usui (2014): "Concave-monotone treatment response and monotone treatment selection: With an application to the returns to schooling," *Quantitative Economics*, 5(1), 175–194.

Pakes, A. (2010): "Alternative models for moment inequalities," *Econometrica*, 78(6), 1783–1822.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment inequalities and their application," *Econometrica*, 83(1), 315–334.

Pakes, A., J. R. Porter, M. Shepard, and S. Calder-Wang (2021): "Unobserved heterogeneity, state dependence, and health plan choices."

de Paula, Á. (2013): "Econometric analysis of games with multiple equilibria," *Annual Review of Economics*, 5, 107–131.

de Paula, Á. (2017): "Econometrics of network models," in *Advances in Economics and Econometrics*, ed. by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, vol. 1, chap. 8, pp. 268–323. Cambridge University Press.

de Paula, Á. (2020a): "Econometric models of network formation," *Annual Review of Economics*, 12, 775–799.

de Paula, Á. (2020b): "Strategic network formation," in *The Econometric Analysis of Network Data*, ed. by B. Graham, and A. de Paula, chap. 3, pp. 41–61. Elsevier.

de Paula, Á., S. Richards-Shubik, and E. Tamer (2018): "Identifying preferences in networks with bounded degree," *Econometrica*, 86(1), 263–288.

de Paula, Á., and X. Tang (2012): "Inference of signs of interaction effects in simultaneous games with incomplete information," *Econometrica*, 80(1), 143–172.

Poirier, D. J. (1998): "Revising beliefs in nonidentified models," *Econometric Theory*, 14(4), 483–509.

Ponomareva, M., and E. Tamer (2011): "Misspecification in moment inequality models: Back to moment equalities?," *The Econometrics Journal*, 14(2), 186–203.

Romano, J. P., and A. M. Shaikh (2008): "Inference for identifiable parameters in partially identified econometric models," *Journal of Statistical Planning and Inference*, 138(9), 2786–2807.

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

Rubio-Ramirez, J. F., D. F. Waggoner, AND T. Zha (2010): "Structural vector autoregressions: Theory of identification and algorithms for inference," *The Review of Economic Studies*, 77(2), 665–696.

Samuelson, P. A. (1938): "A note on the pure theory of consumer's behaviour," *Economica*, 5(17), 61–71.

Sheng, S. (2020): "A structural econometric analysis of network formation games through subnetworks," *Econometrica*, 88(5), 1829–1858.

Stoye, J. (2010): "Partial identification of spread parameters," *Quantitative Economics*, 1(2), 323–357.

Syrgkanis, V., E. Tamer, AND J. Ziani (2021): "Inference on auctions with weak assumptions on information."

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

Tamer, E. (2010): "Partial identification in econometrics," *Annual Review of Economics*, 2(1), 167–195.

Torgovitsky, A. (2019): "Nonparametric inference on state dependence in unemployment," *Econometrica*, 87(5), 1475–1505.

Twinam, T. (2017): "Complementarity and identification," *Econometric Theory*, 33(5), 1154–1185.

Uhlig, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52(2), 381–419.

Uhlig, H. (2017): "Shocks, sign restrictions, and identification," in *Advances in Economics and Econometrics*, ed. by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, vol. 2, chap. 4, pp. 95 – 127. Cambridge University Press.

Wan, Y., AND H. Xu (2014): "Semiparametric identification of binary decision games of incomplete information with correlated private signals," *Journal of Econometrics*, 182(2), 235–246.

Xiao, R. (2018): "Identification and estimation of incomplete information games with multiple equilibria," *Journal of Econometrics*, 203(2), 328–343.

Xu, H. (2014): "Estimation of discrete games with correlated types," *The Econometrics Journal*, 17(3), 241–270.