MOMENT INEQUALITIES AND PARTIAL IDENTIFICATION IN INDUSTRIAL ORGANIZATION

BRENDAN KLINE*, ARIEL PAKES&, AND ELIE TAMER&

Friday 20th August, 2021

*Department of Economics, University of Texas at Austin.
&Department of Economics, Harvard University.
E-mail addresses: brendan.kline@utexas.edu, apakes@fas.harvard.edu, elietamer@fas.harvard.edu.
We are grateful for feedback from a set of reviewers that helped improve the chapter.

1. Introduction

The idea that economic theory generates "inequalities" permeates modern economics. Its implications for empirical analysis appears at least as far back as the revealed preference theory of Samuelson (1938) who emphasized the ordinal nature of preferences and its relationship to consumer choice.[1] Samuelson (1938)'s goal of connecting preferences to observed behavior through revealed preferences is also one central theme of this chapter. Because of the role of inequalities and optimizing behavior in economic theory, and the role of theory in directing both the questions and the forms of analysis in empirical work, it is difficult to give a comprehensive review of the literature that followed. Perhaps the earliest use of inequalities from theory to direct estimation in a context that is clearly relevant for Industrial Organization is in Marschak and Andrews (1944)'s analysis of production functions. They use inequalities derived through second order optimality conditions along with sign restrictions to constrain the parameters of the production function to regions of the parameter space.[2] The Marschak and Andrews (1944) paper uses inequalities in the context of analyzing the implications of theory for choices of a continuous variable (second order conditions), whereas the literature since deals also with choice sets with some possible discreteness.

Much of the early literature emphasized testing of either important economic hypotheses about the utility and/or production function, or of the existence of a model that is compatible with the data and optimizing behavior. Aided by the advances in computing power and the associated increase in the richness of the data available, the use of moment inequalities and partially identified models has expanded to enable the researcher to do richer analysis of policy and environmental changes, particularly, but not solely, those that involve interacting

[1]Other relevant contributions to this strand of thought, often as applied to consumer and producer theory that explicitly considers inequalities and sets is Afriat (1967, 1973), Hanoch and Rothschild (1972), Diewert (1973), Varian (1982, 1983, 1984) and McFadden (2005). See also Deaton (1986), Crawford and De Rock (2014), and Chambers and Echenique (2016) for reviews of revealed preferences in general.

[2]Chapter 2 of Nerlove (1965), titled "Partial identification: the Marschak-Andrews approach," describes the contribution of that approach as one where *"Marschak and Andrews use restrictions on the parameter values obtained from profit maximizing conditions and economic interpretations of the residuals in the production function and in the profit maximizing conditions to achieve partial identification of the parameters of Cobb-Douglas production functions."*

agents. Examples which illustrate this fact are used liberally in Sections 3 and 4 below and involve: (i) the use of weaker, and hence more credible, assumptions in estimation, (ii) circumventing computational issues in problems where neither the researcher nor the decision maker could be expected to compute the action that generated the highest expected return, (iii) examining the nature of beliefs and the information sets that generated them, and (iv) allowing for disturbances that differentiate between the expectations that induced agents' actions and the measures of returns the analyst has access to.

These and other applications of moment inequalities and partially identified models raise a number of methodological issues. Empirical work that involves a theoretical structure that can be used to analyze (past or likely future) behavior requires an "identification strategy" to guide estimation. The estimator can be viewed to be a set- or a single-valued- mapping from the data to the parameters of interest. The identification strategy sets out the logical argument behind the econometrician's use of the assumptions and the process generating the data to learn about the parameters of the model. Therefore, an identification strategy should lead to parameter values that are consistent with the data and the model. These parameter values are then used to obtain answers to the questions of interest.

Point identification is the case familiar from standard textbook treatments of econometrics. However, entertaining only models that lead to a unique value of the parameters of interest restricts the class of models that one can study, and one goal of this chapter is to dispose of this binary view of identification (a parameter is either point identified or not) and enable empirical work for models that are not point identified including moment inequality models. This widens the class of models that can be analyzed for evaluation and prediction purposes. If multiple values of the parameters are consistent with the data and the model, the parameters are partially identified. Then it is possible only for the econometrician to conclude that the true value of the parameter lies within the set of possible values compatible with the assumptions and the data.[3]

[3] Within econometrics broadly there are many instances of models that are partially identified, with examples ranging from models of treatments effects that avoid making strong random assignment of treatment assumptions, to models with missing or censored or mismeasured data, and to nonlinear instrumental variable

Section 2 of this chapter formalizes the distinctions between the various notions of identification from the population of interest and their relationship to alternative estimators obtainable from a given data set, while Section 5 reviews the literature on alternative ways of forming those estimators. Among the applications referred to above, those with multiple decision makers introduce special considerations in both identification and estimation. In these models, the profit (or utility) functions depend on the decisions of other decision makers. This implies that the model can admit multiple "rest points" or equilibrium outcomes and makes the relationship between the observed data and the model parameters more complicated. Generally, in such models, if estimation is based on an equilibrium assumption, then without further assumptions there is not a unique map from the parameters to the distribution of the observed data. As a result the econometrician no longer has a standard likelihood function to guide choosing an "efficient estimator" corresponding to a point identification result. Section 2 defines a notion of "sharpness" of an identified set, the set of parameter values that are subject to all of the restrictions available from the data and the assumptions, while Section 5 considers associated estimators. Note that "efficiency" concerns the statistical properties of the estimator, whereas "sharpness" concerns the identification strategy. The multiplicity problem is also important when we consider analyzing the equilibrium likely to be generated by a new policy, as then in general we cannot use past data to predict the equilibrium chosen in the future.

Section 3 starts out by specifying a set of assumptions that enables us to use revealed preference inequalities to guide estimation for the cases considered in this chapter. The resulting framework allows for: strategic interaction, measurement and misspecification errors, and differences between the distribution the agent conditions on when it makes its decisions and the data generating process underlying the empirical distributions the analyst observes.

models. For examples see Manski (1989, 1990, 1996, 1997, 2003, 2009), Manski and Pepper (2000, 2009), Hotz, Mullin, and Sanders (1997), Molinari (2010), Okumura and Usui (2014), Kline and Tamer (2018), Nevo and Rosen (2012), Santos (2012), Kline (2016b), Chesher and Rosen (2017), Horowitz and Manski (1995), Horowitz and Manski (2000), Manski and Tamer (2002), Molinari (2008), Cross and Manski (2002), Molinari and Peski (2006), Fan, Sherman, and Shum (2014), and many others. Recent summaries of this literature include Tamer (2010), Ho and Rosen (2017), and Molinari (2020).

Examples from empirical work in Industrial Organization illustrate how the use of moment inequalities enhances both our ability to account for these phenomena and to empirically analyze problems which would be considerably less tractable using other techniques. This section also introduces alternative estimation strategies and the assumptions underlying them.

Section 4 discusses a particular implementation of revealed preference. This implementation is motivated by and derived from the standard discrete choice literature. This generalized discrete choice approach uses parametrizations of player payoffs and assumptions on behavior to derive restrictions on the parameters from the observed data. The choice models used are stochastic from the perspective of the econometrician because of variables that determine choice that are not observed by the econometrician. Sections 3 and 4 also consider alternative assumptions that enable us to analyze likely counterfactual equilibria. Given that the result of partial identification analysis is a set of values for the parameters, empirical research requires an approach for conveying the identified sets to the reader. Section 4 of this chapter also provides a detailed illustration of how this can be done, and in so doing reviews the problems that arise in analyzing interacting agent models with a complete information assumption.

Estimation and inference in partially identified models requires special statistical methodology, which is reviewed in Section 5, specifically the literature on the moment inequality approach, the criterion function approach, the random sets approach, and the Bayesian approach. Finally, Section 6 discusses some further issues surrounding implementation of partial identification strategies.

2. Definitions and background

Identification and partial identification concern what can be learned about a quantity of interest, given the assumptions used and the distribution of the data in the population of interest.[4] If the true value of the quantity of interest can be learned, there is point

[4]For the definition of point identification see (e.g., Matzkin (2007) and more recently Lewbel (2019)), while for the definition of partial identification (e.g., Manski (2003)).

identification. This happens when only the true value of the parameter could have generated the population distribution, given the assumptions. If the true value of the quantity of interest can only be learned to be restricted to be within some set, there is partial identification. This happens when multiple specifications of the parameter could have generated the population distribution, given the assumptions.

Even in the simplest empirical settings, identification is always studied in the context of a statistical model. The statistical model could be as simple as the model of i.i.d. sampling from the distribution of a random variable, with no further "model" for that random variable. In general, a statistical model is simply a formal mathematical relationship between a parameter θ and the observed data. This can be a fully parametric model, where the statistical model results in a distribution of the data for each value of θ. Alternatively, this can be a semiparametric model, where for instance the statistical model might result only in statements about the moments of the data for each value of θ (as in linear regression). A statistical model may or may not be based on an economic model of the sort studied by economic theorists. As illustrated below, the formal definition of identification depends on the setup of the model and assumptions underlying the model used by the econometrician. Throughout, suppose the parameter of the model is θ.

The sharp identified set Θ_I is the set of all values of θ that are compatible with all the assumptions and the distribution of the data in the population of interest. By definition every value of the parameter in Θ_I could have generated the population data and is compatible with the assumptions. Therefore, with "infinite data" from the population, it would be possible to learn that the true value of the parameter is in Θ_I, and further it would not be possible to learn more about the true value of the parameter. That is, learning the sharp identified set corresponds to learning the most that can possibly be learned about the true value of the parameter given the assumptions and an "infinite" data set from the population.

A non-sharp identified set $\tilde{\Theta}_I$, sometimes called an outer set, is a set of values of θ that contains all values of the parameter that are compatible with the assumptions and the distribution of the data in the population of interest. However, unlike with a sharp identified set, it is allowed that some values of the parameter in $\tilde{\Theta}_I$ may not be compatible with some of the assumptions and/or the population data. Therefore, it is possible to learn that the true value of the parameter is in $\tilde{\Theta}_I$, but some values of the parameter in $\tilde{\Theta}_I$ could not be the true value of the parameter, if indeed they are incompatible with the assumptions and/or the population data. In that sense, learning a non-sharp identified set corresponds to learning something about the true value of the parameter, but possibly not everything that can be learned given the assumptions and the population data. By definition, $\Theta_I \subseteq \tilde{\Theta}_I$. In some models, finding the sharp identified set can be difficult, specifically because it can be difficult to prove that all parameter values in the candidate sharp identified set are indeed compatible with all of the assumptions and the distribution of the data in the population. It can be much simpler to find a non-sharp identified set, which only requires a proof ruling out certain parameters values as not compatible with at least some of the assumptions and the distribution of the data in the population.

Similar definitions apply to identification of other objects of interest δ, like components or functions of a (vector-valued) θ.

For a formal definition of identification, begin with the kind of model for which defining identification is the simplest. These models result in a distribution of the data for every value of the parameter, as follows. Suppose the econometrician is working with a model that implies that the econometrician observes i.i.d. draws of W_i for $i = 1, 2, \ldots, N$ from some distribution $F(W)$. The econometric model provides the econometrician with distributions $P_\theta(W)$ for each value of the parameter θ in the parameter space Θ, so that the distribution of the observed data is $F(W) = P_{\theta_0}(W)$ with θ_0 being the true value of the parameter. For example, in a structural model of decision makers, W_i could be the observed behavior of decision maker i and $P_\theta(W)$ could be the structural model that generates the observed behavior of the decision makers on the basis of the parameter θ, which for example could determine the utility functions of the decision makers. With an “infinite sample” from the population, the econometrician can treat $F(W)$ as a known quantity. The identification problem concerns what can be learned about θ based on $F(W)$ and any other assumptions used by the econometrician. Note that the form of $P_\theta(W)$ itself reflects modeling assumptions used by the econometrician, but the econometrician may have other assumptions about θ, like non-negativity constraints or other information from prior sources. Formally, assumptions can be represented either as part of the definition of the parameter space Θ, for example non-negativity constraints on parameters, or on the model $P_\theta(\cdot)$, for example functional form assumptions or distributional assumptions.

The sharp identified set for θ is $\Theta_I = \{\theta \in \Theta : P_\theta(W) = F(W) = P_{\theta_0}(W)\}$. Therefore, in this category of model, the sharp identified set collects all values of the parameter that result in the same distribution of the observable data as does the true value of the parameter. By construction, $\theta_0 \in \Theta_I$, but there can also be other elements of Θ_I. There is point identification if $\Theta_I = \{\theta_0\}$ is a singleton, and there is partial identification otherwise. A non-sharp identified set $\tilde{\Theta}_I$ would need to satisfy the condition that $\Theta_I \subseteq \tilde{\Theta}_I$, which implies that if $\theta \notin \tilde{\Theta}_I$ then $P_\theta(W) \neq F(W)$. Thus, all values $\theta \notin \tilde{\Theta}_I$ can be ruled out as candidate values of the true value of the parameter, since such θ would result in a different distribution of the data.

Although the above captures the main idea of the definition of point identification and partial identification, some of the specific details of the kind of model considered there do not reflect typical empirical practice.

First, many models used in empirical practice are conditional models, in the sense that they concern the distribution of W conditional on some observable X. In such cases, the econometric model provides the econometrician with distributions $P_\theta(W|X = x)$ for each value of θ and each x in the support of X. Conditional models leave unspecified the distribution of X. Often, this is because the distribution of X is formally unrelated to θ. Alternatively, this could be because the econometrician allows for the distribution of X to depend on θ, yet does not use that dependence in the identification of θ. Now, with an “infinite sample,” the econometrician can treat $F(W|X = x)$ as a known quantity for all x in the support of X. The sharp identified set for θ is $\Theta_I = \{\theta \in \Theta : P_\theta(W|X = x) = F(W|X = x) = P_{\theta_0}(W|X =$

$x)$ for all $x \in \text{Supp}(X)\}$.[5] Although conditional models (along with the distribution of X) do imply a distribution for (W, X) similar to those considered above in unconditional models, this representation of Θ_I clarifies that the source of identification is how θ relates to the conditional distribution of $W|(X = x)$ and not how θ relates to the distribution of X by itself.

Second, many models used in empirical practice do not result in a distribution of the data for each value of the parameter, as did the models considered above. Rather, many models used in empirical practice result in restrictions on the distribution of the data. As follows, there are two related frameworks for considering such settings.

First, in some models there is a functional $\psi(F(W), \theta)$ of the distribution of the data and the parameter such that the model implies that the true value of the parameter θ_0 satisfies the restriction that $\psi(F(W), \theta_0) \in \Psi$ for some Ψ known by the econometrician. Although it generally does not, Ψ might (implicitly) depend on $F(W)$ – in the same way that $\psi(\cdot)$ can depend on $F(W)$ – and/or the assumptions. Consequently, if this restriction on θ_0 involving ψ is the only restriction from the model, the sharp identified set is $\Theta_I = \{\theta \in \Theta : \psi(F(W), \theta) \in \Psi\}$. Alternatively, if this restriction is only part of the restriction from the model, the non-sharp identified set is $\tilde{\Theta}_I = \{\theta \in \Theta : \psi(F(W), \theta) \in \Psi\}$. Even though $\tilde{\Theta}_I$ has the same definition as Θ_I, it differs by the supposition that the $\psi(\cdot)$ appearing in $\tilde{\Theta}_I$ represents only part of the restrictions from the model whereas the $\psi(\cdot)$ appearing in Θ_I represents all of the restrictions from the model. This formalizes that the econometrician learns that the value of the parameter satisfies the restriction implied above. If there is more than one value of the parameter that satisfies these restrictions, the parameter is partially identified. Examples of restrictions are moment equality conditions and moment inequality conditions. In moment equality condition models (e.g., Hansen (1982)), the model might imply that

[5] This definition depends on $\text{Supp}(X)$ which is an important determinant of what can be learned about the parameter of interest from the data available to the researcher. In particular, generally larger supports of X result in more restrictions on θ according to this specification of Θ_I, such that larger supports of X result in smaller identified sets for θ. Further, the observed support of variables in small data sets can be smaller than the support in large data sets (or the support in the population). If so, in empirical practice there can be effectively fewer restrictions on θ in small data sets compared to the population identified set.

$E(m(W,\theta_0)) = 0$ for some known function $m(W,\theta)$ of an observation W and the parameter θ. In that case, $\psi(F(W),\theta_0) \equiv \int m(W,\theta_0)dF(W)$ and $\Psi = 0$. For example in the context of moment conditions from a linear regression model $Y_i = X_i\beta_0 + \epsilon_i$, the econometrician can assume that $E(\epsilon|X) = 0$, implying the moment condition that $E(X'(Y - X\beta_0)) = 0$. If the distribution of the unobservable ϵ is left unspecified, the model does not result in a distribution of the data but rather just the moment condition that is a statement about the data that depends on the value of the parameter. In moment inequality condition models, see also Section 5.3, the model might say that $E(m(W,\theta_0)) \geq 0$ for some known function $m(W,\theta)$ of an observation W and the parameter θ. In that case, $\psi(F(W),\theta_0) \equiv \int m(W,\theta_0)dF(W)$ and $\Psi = [0,\infty)$. If there is more than one value of the parameter that satisfies these moment inequality conditions, the parameter is partially identified. Moment inequality conditions represent an important category of partially identified models and, as discussed in the following sections, typically care should be taken in how to derive these inequalities from the underlying economic theory model.

Another example of restrictions are models characterized by criterion functions, see also Section 5.4. In such models, there is a function $Q_0(\theta)$ that depends on the distribution $F(W)$, such that θ_0 can be characterized as some property of $Q_0(\cdot)$. For example, it could be that θ_0 is characterized as solving $Q_0(\theta_0) = 0$. In that case, $\psi(F(W),\theta) = Q_0(\theta)$ and $\Psi = 0$. Or, it could be that θ_0 is characterized as solving $\theta_0 = \arg\max Q_0(\theta)$. In that case, $\psi(F(W),\theta) = Q_0(\theta) - \max Q_0(\theta)$ and $\Psi = 0$. In either case, if there is more than one value of the parameter that solves $\psi(F(W),\theta) = 0$, the model is partially identified.

Second, and alternatively, many models used in empirical Industrial Organization are incomplete, in the sense that the model results in a set of possible distributions of the data for each value of the parameter. In such cases, the econometric model provides the econometrician with a set of distributions of the observed data $\mathcal{P}_\theta$ for each value of the parameter θ. This means that the model says only that the distribution of the data will be one of the distributions in $\mathcal{P}_\theta$ when the parameter is θ. In particular, the distribution of the observed data is an element of $\mathcal{P}_{\theta_0}$ so that $F(W) \in \mathcal{P}_{\theta_0}$. For example, in a structural model with multiple decision makers based on a game theory model with multiple equilibrium outcomes, the model including the assumption of equilibrium predicts only that one of the equilibria is selected, but not which equilibrium is selected. See also Section 4. This results in a set of possible distributions of the data, corresponding to all of the different ways of selecting among the multiple equilibrium outcomes. Consequently, the identified set is $\Theta_I = \{\theta \in \Theta : F(W) \in \mathcal{P}_\theta\}$. This formalizes that the econometrician learns that the value of the parameter satisfies the restriction implied above.

A restriction on the distribution of the data, like a moment condition, can be translated to saying that the model results in multiple distributions of the data, namely all distributions of the data that are compatible with the restrictions on the distribution of the data. Conversely, multiple distributions of the data being consistent with the model can be translated to saying that the model results in restrictions on the distribution of the data, namely the restriction that the distribution of the data is within the set of distributions predicted by the model. Therefore, the previous discussion is simply two different perspectives on the same underlying phenomenon. Depending on the specific model, one or the other perspective can result in a more straightforward identification analysis.

Finally, in some cases, the object of interest is a function $\delta(\theta)$ of θ. For one example, it can be that $\delta(\theta) = \theta_k$ is a particular component of $\theta = (\theta_1, \theta_2, \ldots, \theta_K)$. For another example, it can be that $\delta(\theta)$ represents a marginal effect in a non-linear model. Identification of $\delta(\theta)$ is related to but distinct from identification of θ. Building on the simplest representation of an identified set from above, the identified set for $\delta(\theta)$ is $\Delta_I = \{\delta : \exists\theta \in \Theta \text{ s.t. } \delta = \delta(\theta) \text{ and } P_\theta(W) = F(W) = P_{\theta_0}(W)\}$. Or building on the representation of an identified set based on models that imply restrictions on the distribution of the data, the identified set for $\delta(\theta)$ is $\Delta_I = \{\delta : \exists\theta \in \Theta \text{ s.t. } \delta = \delta(\theta) \text{ and } \psi(F(W), \theta) \in \Psi\}$. In general, the identified set for $\delta(\theta)$ can be written as $\Delta_I = \{\delta : \exists\theta \in \Theta_I \text{ s.t. } \delta = \delta(\theta)\}$. If an identification strategy provides an identified set for θ, then it is trivial to determine the identified set for $\delta(\theta)$ by simply applying $\delta(\cdot)$ to all elements of the identified set for θ. However, it is possible that an identification strategy directly provides an identified set for $\delta(\theta)$ without necessarily providing (directly) an identified set for θ. For example, an identification strategy might provide an identified set for the finite-dimensional elements of θ, while not providing an identified set for the infinite-dimensional elements of θ, like the distributions of unobservables. Note however that any identified set Δ_I for $\delta(\theta)$ implies a corresponding (possibly non-sharp) identified set for θ, $\{\theta \in \Theta : \delta(\theta) \in \Delta_I\}$. Indeed, it is possible that $\delta(\theta)$ is point identified even if θ is partially identified.

3. Revealed preference

Revealed preference is a basic building block of economic theory and underlies much of the empirical work in Industrial Organization. The formal literature dates back to the classic article by Samuelson (1938), and the underlying idea is both simple and powerful. If an economic agent chose d from a feasible set of choices which included both d and d', then the agent expected to be better off as a result of choosing d then it would have been had it chosen d'. There is no restriction on either the nature of the choice set or on whether the underlying returns depend on the actions of competitors. This makes the logic of revealed preference particularly attractive to the Industrial Organization community. Note however that the revealed preference logic applies to expectations rather than realizations and is silent on the relationship between the ex ante expectations and the observable variables that the applied researcher has at their disposal. To use the insights from revealed preference in empirical work these relationships must be specified.

3.1. **Primitive assumptions.** Pakes (2010) and Pakes, Porter, Ho, and Ishii (2015) provide conditions which enable one to go from revealed preference to a consistent estimate of a partially identified set. We begin with a definition of revealed preference which depends on the agent's expectation operator and the model for the returns from its actions that the agent believes is true and so is used by the agent in forming its expectations. The empirical probabilities of outcomes need not conform to the probabilities the agent assigns to those outcomes, and the econometrician's specification for the model that determine the returns from the agent's actions need not conform to the model the agent used in making its decisions. The goal of this subsection is to provide a set of assumptions that connect the agent's expectations and perceived model of returns, to the data generating process and the econometrician's model of returns. The next subsection considers the implications of these assumptions for estimation, while the final subsection provides examples which illustrate how they have been used to illuminate issues of importance to Industrial Organization.

We begin with two assumptions that are needed to derive the inequality in the agent's expectations that revealed preference generates. An additional two conditions are then needed to: (i) relate the probabilities of outcomes implicit in the agent's expectation operator to the empirical probabilities the analyst observes, and (ii) relate the model the agent used to determine the returns from those outcomes and the analyst's model of returns. As we shall see, it is the latter assumptions that differentiate the econometric approaches that have been used in Industrial Organization.

Two caveats before we start. First the focus will be on parametric models. Second though special cases of the assumptions presented below underlie virtually all past applied work in Industrial Organization, they are not the only assumptions that could be used, and we will comment on alternatives. Hopefully our occasional references to non-parametric work and weaker assumptions will induce colleagues to pursue these avenues further.

The agent's perceptions. We start with the best response condition. If we assume a parametric model and let $\pi(d_i, d_{-i}, y_i, \theta)$ be the profit agent i would earn as a result of choosing d_i when its competitors chose d_{-i} and y_i is any other determinant of profits, and let boldface variables denote variables that can be random to the decision maker then

$$C1: \quad \mathcal{E}[\pi(d_i, \mathbf{d}_{-i}, \mathbf{y}_i, \theta)|I_i] = \sup_{d \in \mathcal{D}_i} \mathcal{E}[\pi(d, \mathbf{d}_{-i}, \mathbf{y}_i, \theta)|I_i]$$

where I_i is the information set available to the agent when the decision is made, $\mathcal{E}(\cdot)$ provides the *agent's* expectations conditional on that information, and $\mathcal{D}_i$ is a set of feasible choices that the agent considered.

The distributions of $\mathbf{y}_i$ and $\mathbf{d}_{-i}$ might change were the agent to make a different decision. For example, this would occur in a two period model if d_i represented a first period decision (say on the location of a product or a choice of a contracting partner), and $\mathbf{y}_i$ represented the prices chosen in the second period. Then y_i is "endogenous" in the sense that when we compare d_i to the counterfactual d_i' we need to predict what the agent perceived y would have been had it chosen d'. When this occurs we need a model for the response of y_i to a change in d_i. So our second condition, the counterfactual condition, is

$$C2 : (i)\ \mathbf{d}_{-i} = \mathbf{d}^{-}(d_i, z_i) \ \text{ and } \ (ii)\ \mathbf{y}_i = \mathbf{y}(d_i, \mathbf{d}_{-i}, z_i),$$

where $z_i \in I_i$ does not respond to changes in d_i, or d_{-i}; that is z_i is "exogenous"[6].

Part (i) of C2 is always satisfied in both simultaneous move games and in single agent "games against nature", as in these cases d_{-i} does not change when we assign a counterfactual policy to agent i. In contrast, in dynamic games in which the returns to an agent's action depends on the responses of competitors both conditions (i) and (ii) require a specification of the agent's model for its competitors' responses to counterfactual behavior. Condition (ii) is typically also needed in "two-period" models with simultaneous moves in the first period. These models often focus on an initial simultaneous move game (e.g. a location choice), followed by a price setting game that conditions on the outcome of the initial period decision (which depends on the counterfactual chosen) and determines profits (see Subsection 3.3 for examples)[7].

[6]Note that this definition of exogeneity does not refer to the relationship between variables that are, and those that are not, observed by the analyst. The components of the model that are not observed by the analyst are explicitly introduced below where their relationship to observable variables are explained.

[7]It is understood that if a model of competitor behavior is needed, as in the dynamic game example, it may depend on additional parameters. Very little work has been done on the use of inequalities in dynamic games. For a demonstration of their potential use in full information Nash equilibrium models see Berry and Compiani (2020). Modelling frameworks for dynamics which allow for asymmetric information, such as those in Fershtman and Pakes (2012) and Asker, Fershtman, Jeon, and Pakes (2020), are now being adapted for empirical work, and lead naturally to the use of inequalities.

The implication of combining C1 and C2 is that if $d' \in \mathcal{D}_i$ then

$$\mathcal{E}\left[\Delta\pi(d_i, d_i', \mathbf{d}_{-i}, z_i, \theta)|I_i\right] \equiv \mathcal{E}\left[\pi(d_i, \mathbf{d}_{-i}, \mathbf{y}_i, \theta) - \pi\left(d', \mathbf{d}^-(d_i', z_i), \mathbf{y}(d_i', \mathbf{d}_{-i}, z_i), \theta\right)|I_i\right] \geq 0. \tag{1}$$

To go from Equation (1) to a conditional moment inequality we can take to data we need: (i) the relationship between the agent's expectation operator and the expectations generated by the data generating process (the DGP), and (ii) the relationship between the difference in profits in Equation (1) and the econometrician's approximation to that difference.

If $E(\cdot|\cdot)$ is the expectation operator emanating form the DGP then all that is required for (i) is

$$C3: \quad \mathcal{E}\left[\Delta\pi(d_i, d_i', d_{-i}, z_i, \theta)|I_i\right] \geq 0 \Rightarrow E\left[\Delta\pi(d_i, d_i', d_{-i}, z_i, \theta)|I_i\right] \geq 0.$$

So "rational expectations", that is $\mathcal{E}\left[\Delta\pi(d_i, d_i', d_{-i}, z_i, \theta)|I_i\right] = E\left[\Delta\pi(d_i, d_i', d_{-i}, z_i, \theta)|I_i\right]$, will do but we can suffice with the "weak rationality" assumption from Pakes (2010), that agents do not err on average. Indeed the yet weaker assumptions that agent's are not pessimistic, but could be overoptimistic, regarding the returns from the choice they made relative to feasible alternatives would also do.

Given this assumption, there are two approaches to obtaining an empirical analogue to Equation (1); we could either specify the distribution of primitives and attempt to calculate the difference in expectations directly, or we could attempt to calculate the realizations of the profit difference and then insure that our estimating equation is robust to expectational error. If uncertainty can be ignored the two approaches are identical, while if there is significant uncertainty in the environment the former approach requires additional assumptions, including a selection mechanism when there is the possibility of multiple equilibria, and is typically less computationally convenient. So the literature has focused on working with realizations.

With that in mind define

$$\nu_1(d_i, d_i', d_{-i}, z_i, \theta) \equiv \Delta\pi^o(d_i, d_i', d_{-i}, z_i, \theta) - E\left[\Delta\pi^o(d_i, d_i', \mathbf{d}_{-i}, z_i, \theta)|I_i\right],$$

where the $\Delta\pi^o(\cdot)$ is the observable approximation to the difference in profits that would be constructed *if all components of z_i were observed.* Then by construction

$$E[\nu_1(d_i, d_i', d_{-i}, z_i, \theta)|I_i] = 0.$$

$\nu_1(d_i, d_i', d_{-i}, z_i, \theta)$ will contain; (i) expectational error generated by randomness in the environment (that is from the difference between the information the agent has at its disposal when making its decision and the determinants of the returns from that decision), (ii) measurement error in profits, and (iii) that part of functional form misspecification that is mean independent of the agent's information sets. Depending on the issue studied and data available, any one of these sources of error can be dominant. Expectational errors are likely to occur in modeling decisions which generate returns after the decision is made, misspecified profit functions occur when we use approximations to the true objective functions, and measurement errors are thought to be pervasive in the measures of profits we have access to and in cases where publication of the precise values of variables are limited by proprietary contracts.

In addition not all of z_i need be observed to (or even known to) the econometrician. If we let x_i be the variables that the econometrician can condition on, and define

$$\Delta\pi^o(d_i, d_i', d_{-i}, x_i, \theta) \equiv \pi^o(d_i, d_{-i}, x_i, \theta) - \pi^o\left(d_i', \mathbf{d}^-(x_i, d_i'), \mathbf{y}(d_i', d_{-i}, x_i), \theta\right)$$

then there is a second disturbance defined by

$$\nu_2(d_i, d_i', d_{-i}, z_i, x_i, \theta) \equiv \Delta\pi^o(d_i, d_i', d_{-i}, z_i, \theta) - \Delta\pi^o(d_i, d_i', d_{-i}, x_i, \theta). \tag{2}$$

This $\nu_2(d_i, d_i', d_{-i}, z_i, x_i, \theta)$ is a result of variables that the agent has access to but the analyst does not, and/or specification errors in the objective function that are correlated with the information the agent conditions its decision on.

The point to stress here is that it is z, and not x, which determines the agent's decision, so

$$E[\nu_2(d_i, d_i', d_{-i}, z_i, x_i, \theta)|I_i] \neq 0.$$

This implies that if not accounted for, $\nu_2(\cdot)$ will cause a bias in the estimates (this is often referred to as a selection bias).

Note that the components of x_i may or may not be a subset of the components of z_i. In the case where the agent's utility is a function of $z_i = (x_i, \epsilon_i)$ then the x_i are components of z_i, and if either the agent's or its competitors' decisions depend on ϵ_i then $\nu_2(\cdot)$ need not be mean zero conditional on I_i. However in a model with measurement error in a component of z_i the agent observes the true value of that component but the analyst only observes an error prone measure of it. Then the measured component, that is x, is not contained in z. As shown in the examples in Section 3.3.1 the structure of a consistent estimation algorithm can depend on the source of $\nu_2(\cdot)$.

Collecting terms we have

$$\Delta\pi^o(d_i, d_i', d_{-i}, x_i, \theta) = E\left[\Delta\pi(d_i, d_i', d_{-i}, z_i, \theta)|I_i\right] + \nu_1(d_i, d_i', d_{-i}, z_i, \theta) - \nu_2(d_i, d_i', d_{-i}, z_i, x_i, \theta). \tag{3}$$

That completes the discussion of the framework that underlies the estimators discussed below. Note that it avoids the specification of a full model of the decision making environment. In particular it does not require specifications for; information sets, how the agents formulate their expectations given their information, or for the complete choice sets of the agents. It also gives the analyst flexibility to choose counterfactuals that give it the best chance of obtaining precise information on the parameters of interest for the applied problem and data at hand. As the examples will illustrate these properties have enabled researchers to analyze complex problems with limited assumptions. Finally note that when applied to game theoretic problems with multiple possible equilibria, the framework does not require the analyst to specify an equilibrium selection mechanism, or to assume the selection mechanism is the same across markets, in order to obtain consistent set estimators[8]. Since there are many cases where the equilibrium selection mechanism is difficult if not impossible for the analyst to know, this is fortunate.

On the other hand there are drawbacks to limiting ourselves to not specifying a complete model. It should be obvious that use of the restrictions from a full model of the decision making environment when that model is correct could produce more powerful estimators in the sense that the sharp identified will be (weakly) smaller. Perhaps more important, the lack of a full specification and the possibility of multiple equilibria raises the issue of how to analyze counterfactuals. Though we may be able to use data to determine which equilibrium was played in the past, there generically will not be any data which tells us which outcome will be chosen once we change the environment. That is if we do not know what underlies the choice of equilibria that we see in the data, we are unlikely to know if that selection is likely to change when we change the environment. We come back to a more complete description of counterfactuals in models with multiple equilibrium in Section 6.3.2 below.

One final note before proceeding to the assumptions on the disturbance terms in Equation 3 which, together with assumptions C1 to C3, will lead directly to estimation strategies. Sometimes it will be both possible and useful to use data to delve deeper into the primitives that underlie C1 to C3. For example Dickstein and Morales (2018) (see example 8 below) use the data to clarify the information available to agents when they make their decisions (i.e. the contents of I_i above), while Doraszelski, Lewis, and Pakes (2018) investigate how initially uninformed agents develop their beliefs on the actions of their competitors[9].

Where possible, analysis of this sort is likely to lead to a far richer analysis of market outcomes[10]. For example one way to resolve the multiplicity issue discussed above is to

[8]See Tamer (2003) for an early discussion of the impact of multiple equilibria on appropriate estimation techniques for the special case of C1 to C3 we return to in Section 4 below.

[9]A closely related literature analyzes information sets and how they evolve in consumption theory and in the analysis of education and labor market decisions. For early contribution see Flavin (1981) and Carneiro, Hansen, and Heckman (2003).

[10]More fundamentally the theory literature has delved deeper into the conditions needed for different equilibria. For example Aumann and Brandenburger (1995) show that full information Nash equilibrium (which is a special case of our C1 to C3) requires mutual knowledge of the profit functions, common knowledge of

structure how agents learn to adapt to the counterfactual environment. This would modify C3 above to provide an explicit model for how beliefs on likely outcomes from different policy choices in the counterfactual environment are formed[11]. It would also have the added advantage of predicting the transition path from one equilibrium to another.

3.2. **Paths to estimators.** Given C1 to C3 the different estimation strategies differ in their assumptions on the two disturbance terms in Equation 3. We begin with the two extremes. One is for cases where the dominant sources of error are those that determine $\nu_1(d_i, d_i', d_{-i}, z_i, \theta)$, and assumes $\nu_2(d_i, d_i', d_{-i}, , z_i, x_i, \theta)$'s influence on weighted averages of $\Delta\pi^o(d_i, d_i', d_{-i}, z_i, \theta)$ can be ignored. The other is for cases where the dominant sources of error are those that determine $\nu_2(d_i, d_i', d_{-i}, z_i, x_i, \theta)$, and assumes $\nu_1(d_i, d_i', d_{-i}, z_i, \theta)$'s influence on weighted average of $\Delta\pi^o(d_i, d_i', d_{-i}, z_i, \theta)$ can be ignored. Much of the applied work to date has made one of these two assumptions. The examples in Section 3.3.1 provide ways of adding structure to C1 to C3 that enable consistent set estimators when both types of disturbances are important.

To obtain set estimators when $\nu_2(d, d', d_{-i}, z_i, x_i, \theta) = 0$ we form *positive valued* functions of variables known to the agent and observed by the analyst when it made its decision, say $\{h_j(x_i)\}_j$, and search for values of θ that made a metric in the j moments

$$\sum_i \Big(\Delta\pi^o(d_i, d_i', d_{-i}, z_i, \theta)\Big)_- h_j(x_i) \geq 0, \tag{4}$$

beliefs/conjectures, and a common prior. One could attempt to use data to determine the extent to which these assumptions are an adequate approximation to reality in different circumstances.

[11]There is an extensive experimental literature on learning (e.g. Erev and Roth (1998)), which might have implications for empirical work. For example Feltovich (2000) concludes that belief based learning (e.g. fictitious play) and the reinforcement learning models have been used intensively in computational work (e.g. Pakes and McGuire (2001)) seem to perform about equally in a laboratory setting. Relatedly Salmon (2001), using laboratory data on constant sum games, finds that standard econometric techniques have difficulty in distinguishing between various learning models. These papers might suggest that the structure of the learning model that a researcher imposes need not have a substantial effect on predicted outcomes. On the other hand Lee and Pakes (2009) find that this was not the case in evaluating a possible merger in the banking industry.

where if $f(\cdot)$ is any function, $f(\cdot)_- = min(f(\cdot), 0)$, or the negative values of $f(\cdot)$[12]. Minimizing a metric in these moments penalizes values of θ that violate the revealed preference inequalities and can generate set estimators that cover the true value of θ with sufficiently high probability. Details of this, and the other estimation algorithms outlined in this section are given in Section 5.

The path from the assumption that $\Delta\nu_1(d, d', d_{-i}, z_i, \theta)$, does not affect weighted averages of the profits but $\Delta\nu_2(d, d', d_{-i}, z_i, x_i, \theta)$ does to estimation differs with the nature of the problem being investigated. If this is a game against nature, or a single agent problem, the outcome is discrete, and $z = (x, \epsilon)$ we are back to the assumption of standard discrete choice estimation algorithms, and there is a robust literature on estimators for that problem.

In game theoretic problems, the estimation algorithm depends on the extent to which the ϵ_i, the variable that is unknown to the analyst but is known by the agent making the i^{th} decision, is known to its competitors. If it is fully revealed d_{-i} will depend on ϵ_i and we must treat the necessary conditions for the optimal choice of each agent as a system of equations. A second alternative that has been used extensively in the applied literature on auctions is that the competitors' only know realizations of ϵ_i are random draws a particular distribution (for an early creative use of inequalities to analyze auction see Haile and Tamer (2003)). These two cases are both considered full information Nash equilibrium in most of the literature, and we abide by that terminology. If competitors only receive a noisy signal on the value of ϵ_i, say through past choices, the underlying game is one of asymmetric information. Then we need to add a model of belief formation to C1 to C3 above and formulate policies and estimators that are consistent with it.

The case that has been fully worked out and used in the empirical literature is the full information Nash equilibrium case, typically with a parametric assumption on the distribution of the vector of ϵ's. Then to see if a particular value of θ is acceptable the analyst checks

[12]Notice that despite the assumption that $\nu_2(\cdot) = 0$, components of z_i that are not observed by the analyst can be determinants of profits. Models with errors in variables are consistent with this set of assumptions (for details see example 7 below). For another example say profits are additively separable in an error which is mean independent of x_i, z_i is an exogenous determinant of the variance of that error, and agents are expected profits maximizers. Then z_i is a determinant of realized profits but does not effect expected profits.

whether the vector of necessary conditions for equilibrium emanating from C1 to C3 are satisfied. Details are given in Section 4 below, where it is noted that often one can add power by adding information from the sufficient conditions for an equilibrium[13].

Though the assumptions used in these two special cases might seem extreme, they should be familiar from prior applied work. The assumptions used in rational expectation models satisfy the condition that $\Delta\nu_2(d, d', d_{-i}, \tilde{z}_i, z_i, \theta) = 0$[14], and the assumptions used in the static single agent discrete choice literature satisfy the assumption that $\Delta\nu_1(d, d', d_{-i}, z_i, \theta) = 0$. On the other hand it should be clear from the sources of these disturbances that both these assumptions can be problematic in contexts of interest to Industrial Organization. We often do not have access to all the information management keys off of in making their decisions, our measures of profits (or discounted returns) are at best approximations, and since the model's controls maximize expected returns over some interval of time there are also likely to be differences between expectations and realizations.

However we do not know of an estimator that can take account of both types of disturbances without additional assumptions (additional to C1 to C3). Pakes (2010) and Pakes, Porter, Ho, and Ishii (2015) did provide "high level" conditions which enable one to allow for both types of disturbances, but they are subtle and not easy to check[15]. Since then others have added to these conditions in ways that have proved effective for a number of classic Industrial Organization problems.[16]

[13] A similar procedure works if the agent i only has partial information on ϵ_{-i} but then checking equilibrium conditions requires a specification for what agent i does know, and a way of integrating ϵ_{-i} out of the best response condition in C1. The latter becomes more complicated and requires more assumptions in models where the y_i in C2 is endogenous, that is responds to differences in d_i, as then we must check all possible equilibrium for each draw on the value of ϵ.

[14] For an early use of rational expectations to rationalize an estimator based on these assumptions see Hansen and Singleton (1982)'s first order conditions estimator for models with continuous controls.

[15] Their most general condition provides cases which generate an observable positive valued instrument, say $h_j(z_i) \in I_i$, which is known to the agent at the time decisions are made and not positively correlated with $\Delta\nu_2(d, d', d_{-i}, \tilde{z}_i, z_i, \theta)$, and uses them to construct the averages in equation (4) that would be greater than zero. It is essentially insuring the existence of valid instrumental functions that generate monotone instrumental variables along the lines of Manski and Pepper (2000) and Manski and Pepper (2009).

[16] See Ho and Rosen (2017) for a more extensive review of empirical work using moment inequalities.

As a result we proceed as follows. This subsection concludes with two examples that have used the assumption that $\nu_2(d_i, d_i', d_{-i}, \tilde{z}_i, z_i, \theta) = 0$ to analyze problems that would have been difficult, if not impossible, to analyze without the use of moment inequalities. Section 3.3.1, reviews seven examples that have added extra assumptions to enable the use of moment inequalities of the same form as those in Equation 4 to generate consistent set estimators when there are both $\nu_1(\cdot)$ and $\nu_2(\cdot)$ disturbances. Aside from their empirical importance, each example focuses on a different econometric problem that appears frequently in Industrial Organization. Then Section 4 considers algorithms that set $\nu_1(d, d', d_{-i}, z_i, \theta) = 0$, and uses an example to illustrate how to construct identified sets from partially identified models.

3.3. **Examples.** We begin by looking at two examples that used the revealed preference assumptions in C1 to C3, assume $\nu_2(\cdot) = 0$, and provide information on I.O. topics that would have been difficult to analyze without the use of moment inequalities. The first example is a single agent problem, and the second involves a market of interacting agents.

We note that with hindsight both of these examples could have been modified to include disturbances with $\nu_2(\cdot)$ properties were we willing to restrict the distribution of that disturbance in one or more of the ways introduced in the examples in Section 3.3.1. So the contributions of these papers, in both methodology and in empirical issues we can address, can be adapted to a wide variety of applied problems.

Example 1: Single Agent Dynamic Discrete Choice With Large Choice Sets. Moment inequalities have been used to circumvent computational problems in dynamic discrete choice problems in which evaluating all possible sequences of choices is simply infeasible. These models adapt the logic of the Euler perturbations used to circumvent computational problems in the estimation continuous choice problem (see Hansen and Singleton (1982)'s classic article) to problems with discrete choices[17]. That is they assume C1 to C3, use revealed preference inequalities to compare the actual choice to alternative feasible choices, and employ the

[17] A related adaptation for problems with boundaries that are hit with positive probabilities is provided in Pakes (1994).

resultant moment inequalities in an estimation algorithm. We illustrate with influential examples from the literature.

The first is due to Holmes (2011) who studies the sequence of location choices of Wal-Mart stores. His model is a dynamic with rich geographic detail on the locations of Wal-Mart's stores and distribution centers. Given the enormous number of possible combinations of store-opening sequences, it is not feasible to directly solve the dynamic programming problem. Instead he uses a revealed preference approach to infer the magnitude of the density economies resulting from closeness to distribution centers. In particular he analyzes how much sales cannibalization of closely packed stores Wal-Mart is willing to suffer to achieve density economies.

Holmes conditions on the number of stores and distribution centers that Wal-Mart opens in different years, and analyzes the choices for the location of the stores that they open in subsequent years. He obtains his revealed preference inequalities by constructing unbiased estimate of the difference between the expected discounted value of profits resulting from the realized sequence of store openings and the expected discounted value that Wal-Mart would have obtained had it chosen an alternative sequence of store openings. Importantly he restricts himself to "pairwise re-sequencing"; that is of changing the timing of opening of two stores that did open. This eliminates the need for specifying candidate locations for stores. It also implies that one need not compute returns from the two sequences for the years after the year that the last of the couple being re-sequenced opened, as the profits after that are the same for the actual and the counterfactual choice[18]. The actual perturbations are chosen to focus in on the trade-off between store density, population size, and distance to a distribution center. The analysis shows how to make one source of the economics of density explicit, the cost of deliveries from the distribution center to local Wal-Marts, and quantifies

[18]The two sources of the $\nu_1(\cdot)$ disturbance likely in Holmes' model are (i) measurement error in profits, and (ii) the difference between expectations and realizations. If the perturbations coincide with a decision by management today that determines opening over the period of the perturbation, then the estimator is consistent if both sources of $\nu_1(\cdot)$ disturbances are present. If not, one has to rely on the assumption that the $\nu_1(\cdot)$ disturbance is wholly measurement error as future decisions condition on realizations of the expectational errors in the subsequent periods.

its effects. His estimates show that the economics of density, measured in this way, are an important part of Wal-Mart's profitability.

A paper by Morales, Sheu, and Zahler (2019) uses related ideas in their analysis of "extended gravity"; i.e. they show how firms current choices of export markets depend on the markets they have serviced in the past, and analyze the implications of their finding. Given the number of possible countries to export to the the full solution to their dynamic multinomial discrete choice problem is not feasible. Instead they employ revealed preference and moment inequalities to compare the returns from the observed and alternative sample paths to obtain estimates of the needed parameters[19]. A somewhat different argument was used to analyze "switching cost" in dynamic consumer choice by Illanes (2017) in his analysis of the choice of pensions in Chile. He uses the necessary condition that the optimal choice must lead to a higher discounted value than feasible alternatives, and an approximation to the difference in discounted values from the two alternatives to develop inequalities that allow him to estimate the parameters of the model. □

Example 2: Product Repositioning. Product repositioning refers to a change in the characteristics of the products marketed by an incumbent firm. Recent work has shown that there are a number of industries in which firms already in the market can change the characteristics of their products as easily as they can change prices. As a result an analysis of pricing responses to environmental change (for e.g. to a merger) that does not take repositioning into account is likely to be seriously misleading, even in the very short run. Examples include; Nosko (2010) analysis of the response of the market for CPU's to innovation, Eizenberg (2014)'s analysis of the impact of the introduction of the Pentium 4 chip in PC's and notebooks, and Wollmann (2018)'s analysis of the impact of the bailout of GM and Chrysler's truck divisions (which we come back to below).

[19]The perturbations in this paper are chosen on the basis of purely exogenous factors, so provided the sequence chosen for the perturbation was in the original choice set, the estimators will be consistent when both measurement and expectational error are embedded in the $\nu_1(\cdot)$ disturbance.

These models require estimates of the fixed costs of adding and of deleting products. For simplicity assume these are constant across products (typically they would depend on product characteristics) and equal to F, a parameter to be estimated. To estimate bounds on F go back to Equation 1 and consider two sets of counterfactual choices;

- First construct the profits from counterfactual d' that consist of all the products that were marketed but a particular product, so that $\mathcal{E}\left[\Delta\pi(d_i, d_i', d_{-i}, z_i, \theta)|I_i\right]$ represents the perception of what the difference in profits would have been had the firm deleted a product that was marketed.
- Then construct the profits from counterfactual d' that consist of all the products that were marketed plus a product that could have been marketed but was not, so that $\mathcal{E}[\Delta\pi(d_i, d_i', d_{-i}, z_i, \theta)|I_i]$ represents the perception of what the difference in profits would have been had the firm added a product that was not marketed.

Assuming $C1$ and $C2$ and the weak rationality assumption that suffices for $C3$ above, we expect the average difference from the first set of counterfactuals, that of deleting a product that was marketed, to be larger then the fixed costs of marketing a product. Analogously the difference in profits from adding a product that was not marketed should be less than the fixed costs. So the two different sets of inequalities generate bounds on F.

Wollmann (2018) notes that the fact that trucks are "modular" in that different cabs can be connected to different trailers, makes it easy to reposition products. His goal is to compare ways of analyzing what would have happened to the truck market had there not been a bail out of GM and Chrysler's truck divisions. One way of analyzing the counterfactual assumes only that prices would have changed, the second also allows the remaining firms to reposition their products in response to the change in their competitors. To compute what the counterfactual outcome would be he needs to select out which, among many possible counterfactual equilibria, are likely to be realized. For this he uses a simple reinforcement learning model (similar to that used in Lee and Pakes (2009)).

The results below are taken from Wollmann (2018)'s Table 5. They look at two possible responses to the bail out; one where the truck divisions of GM and Chrysler are simply liquidated, and another where they are bought out by Ford. The rows provide changes in the average markup, output, and compensating variation between the counterfactual and the actual bailout. The left hand side of the table provides the counterfactual results when only the prices of the remaining products are allowed to change. The right hand side provides the results when the remaining incumbents firms are allowed to reposition their products when GM and Chrysler leave the market. The differences between the two ways of calculating counterfactuals are large, suggesting that analyzing counterfactuals by just looking at the induced price changes and ignoring product repositioning would lead to seriously misleading results, even in the very short run. □

Counterfactual Outcomes

	Repositioning Ignored		With Repositioning	
	Ford Acq.	Liq.	Ford Acq.	Liq.
1. Markups (%)	10.9	4.0	0.8	0.0
2. Quantities (%)	-5.1	-11.2	-1.3	-1.6
3. Comp. Variation (Mill 2005$)	119	253	22	28

We now go on to models that use special structure to allow for both types of disturbances.

3.3.1. *Richer assumptions on disturbances.* As noted earlier the progress that has been made in allowing for sources of disturbance that generate significant variance in both $\Delta\nu_2(d, d', d_{-i}, \tilde{z}_i, z_i, \theta)$ and $\Delta\nu_1(d, d', d_{-i}, z_i, \theta)$ is through special structure on one or the other disturbance. We illustrate with examples which contain methodology for doing so which is likely applicable to many Industrial Organization problems. All of these examples abide by conditions $C1$ to $C3$ in Section 3 and generate moment conditions of the form in Equation 4.

Examples 3 and 4 are cases where for each observation we can construct an alternative choice, a d', that is linear in $\nu_{2,i}$ for every d_i chosen, a fact which allows them to use various forms of differencing to circumvent the selection problems. They can do this because the choice sets are ordered in particular ways. The two examples differ in that Example 3 is a single agent discrete choice problem, and Example 4 involves a market of interacting agents.

Example 3. This is from Katz (2007) who studies the costs that shoppers assign to driving to a supermarket. This is important to the determination of zoning laws, public transportation, ..., but has been difficult to analyze because it is a two stage decision process (first choose a supermarket and then products to purchase) and the second stage has a complex choice set (all possible bundles at the chosen store).

The agent's decision, or d_i, consists of the store chosen, say s_i, and the basket of goods bought, say b_i, so that $d_i = (s_i, b_i)$. Katz assumes that the agents' utility functions are additively separable functions of;

- utility from basket of goods bought,
- expenditure on that basket, and
- drive time to the supermarket.

So if z_i are individual's characteristics then the agent's utility is

$$\pi(d_i, z_i, \theta) = U(b_i, z_i) - e(b_i, s_i) - \theta_i dt(s_i, z_i),$$

where $e(\cdot)$ provides expenditure, $dt(\cdot)$ provides drive time, and the free normalization is used on expenditures (so the cost of drive time are in dollars).

To obtain revealed preference inequalities, Katz compares the actual choice, d_i, to the alternative $d'(d_i)$ of purchasing

- the *same basket* of goods,
- at a store which is *further away* from the consumer's home then the store the consumer shopped at.

This eliminates both the need to specify the choice set and the need for a two stage problem. Notice that the fact that the agent might have chosen a different basket at d' then at d, just reinforces the inequality.

Let $\mathcal{E}(\cdot)$ be the *agent's* expectation operator. Then we know that

$$\mathcal{E}\Big[\Delta\pi(d_i, d'(d_i), z)\Big] =$$

$$-\mathcal{E}\Big[\Delta e(d_i, d'(d_i))\Big] - \theta_i \, \mathcal{E}\Big[\Delta dt(d_i, d'(d_i))\Big] \geq 0.$$

Case 1: $\theta_i = \theta_0$. An assumption which suffices for C3 above here is

$$N^{-1}\sum_i \mathcal{E}\Big[\Delta e(d_i, d'(d_i))\Big] - N^{-1}\sum_i \Delta e(d_i, d'(d_i)) \to_P 0,$$

$$N^{-1}\sum_i \mathcal{E}\Big[\Delta dt(d_i, d'(d_i))\Big] - N^{-1}\sum_i \Delta dt(d_i, d'(d_i)) \to_P 0$$

which would be true if, for e.g., agents were weakly rational. Then

$$-\mathcal{E}\Big[\Delta e(d_i, d'(d_i))\Big] - \theta \, \mathcal{E}\Big[\Delta dt(d_i, d'(d_i))\Big] \geq 0 \;\Rightarrow\; -\frac{\sum_i \Delta e(d_i, d'(d_i))}{\sum_i \Delta dt(d_i, d'(d_i))} \;\to_p\; \underline{\theta} \;\leq\; \theta_0,$$

where the change of sign in the last inequality is a result of the fact that $dt(\cdot) < 0$. Had we taken an alternative store which was closer to the individual

$$-\frac{\sum_i \Delta e(d_i, d'(d_i))}{\sum_i \Delta dt(d_i, d'(d_i))} \;\to_p\; \overline{\theta} \;\geq\; \theta_0,$$

and when he uses both inequalities he gets an interval estimate of θ.

Case 2: $\theta_{2,i} = (\theta_0 + \nu_{2,i})$, $\sum \nu_{2,i} = 0$. This case allows for a component of the cost of drive times ($\nu_{2,i}$) that is known to the agent (since the agent conditions on it when it makes its decision) but not to the econometrician. Then provided $dt(d_i)$ and $dt(d'(d_i))$ are *known to the agent*

$$\mathcal{E}\left[\frac{\Delta e(d_i, d'(d_i))}{\Delta dt(d_i, d'(d_i))} - (\theta_0 + \nu_{2,i})\right] \leq 0, \;\Rightarrow\; \frac{1}{N}\sum_i \left(\frac{\Delta e(d_i, d'(d_i))}{\Delta dt(d_i, d'(d_i))}\right) \;\to_P\; \underline{\theta} \;\leq\; \theta_0.$$

and as above we get bounds for θ.

The first case does not allow for unobservables the agent knows but the researcher does not generates its estimates as the ratio of averages, while in the second which does allow for such unobservables obtains its bounds from an average of ratios.

Katz's model is richer than this and allows for additional observable determinants of supermarket choice and of drive time. He also compares his results to the drive times obtained from a standard multinomial choice model taken from the relevant literature. Here we provide the median estimates of the drive time coefficient he obtains from analyzing the shopping behavior of Massachusetts shoppers.

The median estimate from the multinomial model was \$240 per hour (the median wage in this region is \$17 per hour). The inequality estimators generated point estimates but tests indicated that the model was accepted. Point estimates and very conservative confidence sets for the two cases are

$$\theta_i = \theta_0 : \quad .204 \ \ [.126, .255].$$

which translates to a point estimate of \$4 per hour and

$$\theta_{2,i} = \theta_0 + \nu_i : \quad .544 \ [.257, .666],$$

which translates to a point estimate \$14 per hour.

Clearly both estimates are more reasonable than the multinomial estimate, but from the fact that the median wage was \$17 per hour and the more detailed results that use observable store characteristics that are in the paper, the estimate that allows for a ν_2 seem more accurate. □

Example 4. This is taken from Joy Ishii (2004)'s thesis. She analyzes the welfare implications of alternative market designs for the placement of ATM machines. Her model of the choices of the number of ATM's by firms in a given market is a semi-parametric ordered choice problem. She constructs an estimate of bank i's revenue should it chose d_i machines when its competitors have d_{-i}, given the exogenous variables z_i; say $r^o(d_i, d_{-i}, z_i)$. In her base case she assumes the cost of servicing a machine is independent of the number of machines, but varies over banks. Since the realized profit from the ATM's is its revenues minus its costs we have

$$\pi(d_i, d_{-i}, z_i) = r^o(d_i, d_{-i}, z_i) - (\theta_0 + \nu_{2,i})d_i + \nu_{1,i,d}$$

where; $\theta_0 + \nu_{2,i}$ are the costs of servicing an ATM by firm i, with θ_0 equal to the mean of those costs across firms so $\sum_i \nu_{2,i} = 0$, and $\nu_{1,i,d}$ contains measurement and expectational error both of which are orthogonal to z_i. Similar to Example 3 but working in a market environment Ishii compares the choice actually made to alternative choices, say d_i'. Revealed preference implies

$$E\left[(r^o(d_i, d_{-i}, z_i) - r^o(d_i', d_{-i}, z_i)) - (\theta_0 + \nu_{2,i})(d_i - d_i') + (\nu_{1,i,d} - \nu_{1,i,d'}) \mid z_i\right] \geq 0$$

Provided $\forall i,\ d_i - d_i' = c$ for some constant c, when we take the average of this over the N firms in the market we find

$$E\left[N^{-1} \sum_i \Big((r^o(d_i, d_{-i}, z_i) - r^o(d_i', d_{-i}, z_i)) + (\nu_{1,i,d} - \nu_{1,i,d_i c})\Big) - c\theta_0 \mid z_i\right] \geq 0$$

This equation with $c = 1$ and $c = -1$ would generate upper and lower bounds for θ_0. However some of the banks have no ATM's so $c = -1$ is not a feasible counterfactual for those banks. Moreover dropping the observations with $d_i = 0$ before forming the inequalities would generate a selection problem since those banks are likely to be banks with high costs of servicing ATMs. To deal with the boundary problem an additional assumption, that the $\nu_{2,i}$ are i.i.d. with a distribution that is symmetric (about zero) is used. Extending the argument of Powell (1986), the symmetry assumption allows for the use of the information from the direction that is not truncated (for $c > 0$) to obtain a bound in the truncated direction ($c < 0$). Details are provided in Pakes, Porter, Ho, and Ishii (2015). □

The next two examples use the special structure that ν_2 is constant over a group of observations. This assumption has been investigated in the panel data discrete choice literature where multiple choices for a given individual are observed and the ν_2 is treated as an individual and choice specific fixed effect (see Chamberlain (1980) for conditional logits and Manski (1987) for non-parametric binary choice). The allowance for the fixed effect generates the analogue to the within-between distinction that has been helpful in analyzing models with continuous dependent variables. This because the factors that drive difference in choices among different individuals (or firms) are often different those that drive changes in the choices of a single individual (firm) over time. Since panel data has been increasingly available to use for issues in I.O. we begin with the panel data examples, and then use the same structure to revisit the entry literature taking explicit account of the fact that the two period models are approximations meant to summarize the available data.

Example 5. We begin with a result in Pakes and Porter (2019), and in a modified form in Shi, Shum, and Song (2018). They analyze choices allowing for "i" (i.e. individual or firm) by product fixed effects. We then note that extensions of their results are able to analyze two other problems that have hampered empirical literature for some time; distinguishing between unobserved heterogeneity from state dependence (or switching costs) in demand analysis (noted in this example), and interpreting the results from two period entry models (the next example).

Consider decision makers with feasible choices $d \in \mathcal{D}$ in each of $t = 1, \dots T$ periods. The profit from each choice in period t is given by

$$\pi(d_{i,t}, d_{-i,t}, z_{i,t}, \theta_0) = r^o(d_{i,t}, d_{-i,t}, z_{i,t}; \theta_0) + \nu_{2,d,i} + u_{i,t,d}. \tag{5}$$

Notice that though each agent has a different choice specific fixed effect (the $\nu_{2,i} \equiv \{\nu_{2,i,d}\}_{d \in \mathcal{D}}$), those effects do not vary over time. Also we have used the notation $u_{i,t,d}$ here as a second disturbance to make it clear that this disturbance could have either the properties of a $\nu_1(\cdot)$ or an additional $\nu_2(\cdot)$ disturbance, provided that

- its *marginal distribution is constant over time*, and
- is independent of $z_i \equiv \{z_{i,t}\}_t$.

The $\{u_{i,t,d}\}_{t,d}$ can be freely correlated both across time and across choices. Restricted versions of these assumptions are used in the panel data discrete choice literature that emanated from the articles cited above.

We begin by comparing choices across t for a given i assuming that individuals (i) are independent draws from a larger population. Note that since neither the ν_2 nor the distribution of the $u(\cdot)$ differ across time periods, any difference in the probability of a given choice between period s and period t is determined solely by comparing $r^o(d_{i,s}, d_{-i,s}, z_i; \theta_0)$ to $r^o(d_{i,t}, d_{-i,t}, z_i; \theta_0)$. With this in mind let $d^1(\cdot)$ be the choice with the largest difference in $r^0(\cdot)$ between the two periods

$$r^o(d^1 = d_{i,t}, d_{-i,t}, z_i; \theta_0) - r^o(d^1 = d_{i,s}, d_{-i,s}, z_i; \theta_0) \geq \max_{d \neq d^1} \Big[r^o(d = d_{i,t}, d_{-i,t}, z_i; \theta_0) - r^o(d = d_{i,s}, d_{-i,s}, z_i; \theta_0) \Big]$$

while for $j = 2, \ldots D$, the choice with j^{th} largest difference is

$$r^o(d^j, \cdot)_t - r^o(d^j, \cdot)_s = \max_{d \notin \{d_1, \ldots d_{j-1}\}} \Big[r^o(d = d_{i,t}, d_{-i,t}, z_i; \theta_0) - r^o(d = d_{i,s}, d_{-i,s}, z_i; \theta_0) \Big].$$

In words, d_1 is the choice whose observed returns improves most between periods s and t, d_2 improves the next most, and so on.

Now form sets of choices as follows: $D_i^1(\theta) = d_i^1(\theta), D_i^2(\theta) = \{d_i^1(\theta), d_i^2(\theta)\}, \cdots, D_i^{D-1} = \{d_i^1(\theta), d_i^2(\theta), \cdots, d_i^{D-1}(\theta)\}$. Pakes and Porter (2019) show that at $\theta = \theta_0$

$$Pr(d_{i,t} \in D_i^j(\theta_0) | z_i, d_{-i}) \geq Pr(d_{i,s} \in D_i^j(\theta_0) | z_i, d_{-i}), \quad \forall j.$$

That is if the model's profits for all options $d \in D^j$ improves by more than all options $c \notin D^j$ between periods s and t, then i will be more likely to choose an option $d \in D^j$ at time t than at time s[20]. This is true regardless of the value of the unobserved fixed effects. Moreover given these assumptions, these inequalities are sharp.

Note that the model above allows the $\{u_{i,t,d}\}_{t,d}$ to be freely correlated across t for a given i. That is one way of allowing for correlation over time in choices that is not accounted

[20]Since these probabilities can be expressed as the expectation of indicator functions, estimation can proceed as it does for Equation 4, and is explained after that equation.

for observables. Pakes, Porter, Shepard, and Calder-Wang (2021) show that if we replace the assumption that the marginal distribution of $\{u_{d,i,t}\}_{d,i,t}$ is constant over time with the stronger assumption that its distribution is i.i.d. over t then a model that allows for both individual by product fixed effects (the $\nu_{2,d,i}$) and state dependence can be analyzed by extending the result above. So this is a second way of modeling correlation over time that is not accounted for by observables.

The model analyzed in Pakes, Porter, Shepard, and Calder-Wang (2021) is

$$\pi(d_{i,t}, d_{-i,t}, z_{i,t}, \theta_0) = r^o(d_{i,t}, d_{-i,t}, z_{i,t}; \theta_0) - \kappa\chi\{d_{i,t-1} \neq d\} + \nu_{2,d,i} + u_{i,t,d},$$

where $\chi\{\cdot\}$ is the indicator function which takes the value of one if the condition inside is satisfied and zero elsewhere. The interest in this model is due to the fact that, given its assumptions, it can distinguish correlation over time that is due to unobserved components of preferences (the $\{\nu_{2,i,d}\}_{i,d}$) from a "causal" effect of last period's choice on the current choice ($\kappa\mathcal{I}\{d_{i,t-1} \neq d\}$).

The distinction between state dependence and unobserved heterogeneity has been in the econometric literature at least since the work of Heckman (1978b, 1981) and in many applications has significant policy implications. Examples of interest to Industrial Organization include demand models with switching costs and unobserved heterogeneity. Pakes, Porter, Shepard, and Calder-Wang (2021)'s empirical work analyzes demand for health insurance choices and finds that allowing for state dependence has a major effect on price coefficients. Another application is in the study of the impact of monitoring devices in (car or health) insurance; there the researcher might want to distinguish whether the increase in compliance is a result of adverse selection or moral hazard (see Abbring, Heckman, Chiappori, and Pinquet (2003)). Finally several network models can be modelled in a similar way; that is do two individuals in the same network make similar choices because they have similar tastes, or is it because they communicate among themselves (see, for e.g. Conley and Udry (2010)). Note that in the latter case i indexes a network and t indexes individuals within that network.

Example 6 is from Pakes (2014) who considers what we can learn from two period entry models. The early literature on entry worried about unobservable differences in profitability across markets. Since higher market profitability was serially correlated it led to both more incumbents and a larger incentive to enter. The resulting correlation between the profitability differences and the extent of incumbency would lead to estimates of the impact of the number and types of incumbents on the likelihood of entry that were significantly biased downward. Given the importance of entry as an equilibrating force in markets it should not be surprising that this worry generated a series of influential papers including those of Bresnahan and Reiss (1990, 1991a,b) and Berry (1992).[21] These papers were based on the two period models used by our theory colleagues to develop intuition for what could happen in dynamic situations.

The resulting two period models were designed to produce a summary of incentives to enter as a function of the number and type of incumbents that *conditioned* on market profitability. However they; (i) abstracted from the fact that there was a history prior to the first period and a future after the second, (ii) they typically used rough approximations to the static profit functions, and (iii) they did not account for expectational errors. These are the reasons why it is not correct to interpret the estimates obtained from these models as structural models of the entry process. However if we are trying to produce summary statistics on the relationship between entry and the number and type of incumbents *conditional on* the profitability of different markets, we would want to have both a ν_2 and a ν_1 disturbance, since the latter would allow for the approximation and expectational errors noted above.

To illustrate how much of a difference allowing for both types of error can make this example computes the solution to a sequential Markov Perfect entry game, uses it to generate data, and then uses that data to regress the value of entry on the number and type of incumbents allowing for market size variables that differ by market. This is the true reduced form regression which summarizes the relationship between entry and the number and type of incumbents. The paper then uses the same data in two estimation algorithms, one that

[21]Related subsequent work includes Ellickson, Houghton, and Timmins (2013) who study differences between large retailers, Aradillas-López and Rosen (2018) who study the number of stores different retailers open in markets, and Mazzeo (2002) and Seim (2006) who study geographic location of entry decisions.

allows for both ν_2 and ν_1 and another algorithm that allows for only ν_2. The results are compared to the reduced form regression described above. Note that here the ν_2 variable is common across firms in the same market, which gives us the structure needed to allow for both types of disturbances.

This example uses the Pakes and McGuire (1994) algorithm to compute equilibrium to Markov Perfect games in which potential entrants determine whether to enter in a given location (East or West). Those that enter chose in each active period thereafter whether to exit and if not whether to invest in a quality variable which evolves over time as a controlled Markov process. Consumers are either located west or east and differ in their sensitivity to price. Period profits are determined by a structural model of demand and costs, and a Nash pricing equation. A panel of markets were simulated. The markets differed in a market size variable which distributes as an exogenous Markov process with normally distributed increments. At a given point in time this is common to the firms in a market. Incumbents received i.i.d. cost and selloff value shocks and potential entrants received i.i.d. entry cost shocks. Each market could have up to six incumbents, and those not active were potential entrants.

The value of being active is calculated for each market in each time period and is regressed against the exogenous variable and the following endogenous market characteristics; the number of a firm's competitors in the given location and quality, $n_{l,q}$, the number with the given quality n_{-q} and the number in the same location and quality n_{-l}. This generates the summary we hope our estimators will replicate.

The model is then estimated two different ways neither of which have access to the market size variable. One estimator assumes there are only ν_2 errors and uses the algorithm described in the next section (labelled GDC for generalized discrete choice in the table), the other uses the algorithm described in Example 5 (labelled the P&P algorithm in the following table). When we use the P&P algorithm of Example 5 we compare choices of different firms in the same market at a given point in time.[22] This because the ν_2 fixed effect now has the

[22] Formally the i index in Equation 5 is now a market, and the t indexes firms.

same value for all agents in a given market at a particular point of time. The ν_1 are, by construction, orthogonal to the current included variables as they are constructed as the residual from a projection of the true values on those variables. So for the properties of the P&P estimator to hold we have to additionally assume that the ν_1 disturbances are distributed independently of those variables and are identically distributed across firms at a given time.

The coefficients from the summary regression and the confidence sets from the two estimators are provided in the following table[23]. The underlying data set was large enough to provide estimates of what is essentially the identified set.

Estimate of	$n_{l,q}$	n_{-l}	n_{-q}
True Data Summary; $R^2 = .74$			
1. This is OLS with Mkt Size*	-0.72 (.01)	-0.58 (.01)	-0.07 (.01)
Analyst does not know market size			
2. P-P Inequalities	[-1.02,-0.68]	[-0.81,-0.44]	[-0.27,-0.03]
3. GDC Inequalities	[-0.86,-0.70]	[-0.32,-0.07]	[0.12,0.24]

*The market size variable's coefficient was 1.59 (.01)

Row 1 provides the true reduced form coefficients for the column variables. The first point to note is that the "state variables" on right hand side of the reduced form regression only generate an R^2 of .74. So the approximations generated by the regression of the true entry value on those state variables is not nearly perfect. This imperfection is the source of the ν_1 disturbance.

The confidence intervals for the two estimators (rows 2 and 3) both cover the true value of the $n_{l,q}$ coefficient. However the confidence interval for the GDC estimator, the estimator which ignores approximation error, does not cover the n_{-l} coefficient and estimates a confidence

[23]The GDC estimator requires distributional assumptions; it was assumed that the analyst knew that the firm specific shocks were exponential (as is true). The P&P estimator is non-parametric in the distribution of the disturbances but does require a normalization.

interval for the n_{-q} estimator that is entirely in the wrong orthant. The P&P confidence intervals cover all three coefficients. So just as not accounting for the ν_2 disturbance in Example 3 led to a misleading estimates in the supermarket example, not accounting for the ν_1 error in this example leads to misleading estimates of the true reduced form of this example.

The P&P estimator does rely on knowing that the value of ν_2 is common across a number of observations. On the other hand just as in the within-between case for consumer demand, this is a case where the purpose is to condition on a difference which is common to incumbents and potential entrants, and consider the impact of number and type of incumbents conditional on those differences. □

The next two examples incorporate ν_1 disturbances generated by errors in either expectation or measurement in a right hand side variable into models with a ν_2 disturbance. The first considers a discrete choice setting, the allocation of patients to different hospitals. It allows for the ν_1 disturbances to be in price, since the prices actually paid are not reported, and controls for ν_2 by allowing for thousands of fixed effects. The second "plugs in" an error-prone measure of observed returns for the expected returns required by the model. This generates a ν_1 disturbance. The paper also allows for a determinant of expectations which is observed by the agent but not be the analyst (the source of their ν_2 disturbance). The latter generates a selection term which depends on expected returns in a non-linear way. Their bounds require only that the ν_2 error have a log-concave distribution. The paper provides a creative way of handling a problem (the difference between realized and expected returns) which is likely to appear in many I.O. settings.

Example 7 Ho and Pakes (2014) study the impact of capitation contracts on the price sensitivity of doctors who allocate women giving birth to alternative hospitals. They show that the analysis requires hospital × severity fixed effects, and given that there were over one hundred and fifty hospitals and a hundred severity groups, it was not practical to include them in a standard multinomial choice model. Since patients with different comorbidities (the severity indicator), were allocated to different hospitals, the missing hospital $\times$ severity interactions become ν_2 errors.

In addition the actual prices charged to insurers were not observed. Prices were proprietary information set out in contracts negotiated between the hospitals and health insurers. The analysts observed only list prices which, on average, were about three times actual prices and varied significantly across insurers for the same treatments at a given hospital. This generated measurement error in price, in addition to the expectational error resulting from the allocation being made without knowing the realized price which depended on the treatments given at the hospital.

Inequalities from revealed preference are used to analyze this problem. The specification for the expected welfare of woman i with insurer π and severity s who was allocated to hospital h, say $W(i, \pi, h, s)$, was linear in separate severity by hospital fixed effects, distance from home to the hospital and a price term which was constructed from the list price and the hospital's average discount interacted with insurer specific dummies. Consider two women (i and i') who were; allocated to h and $h' \neq h$ hospitals, had access to both h and h', had the same insurance plan, and had the same highest comorbidity (which determines severity). Note that if they differed in the extent of other comorbidities and they would differ in expected price. Revealed preference implies

$$W(i, \pi, h, s) - W(i, \pi, h', s) \geq 0, \text{ and } W(i', \pi, h', s) - W(i', \pi, h, s) \geq 0.$$

The analysis adds these two inequalities together, interacts the sum with positive valued instruments, sums over couples of women, and finds the values of the parameter vector that maintain the inequality. Since the inequality for i contains the difference between the h and h' fixed effects for severity s and that for i' contains the difference between the h' and h fixed effect, the two fixed effects cancel in the sum. Moreover the averaging over (i, i') couples averages out the expectational and measurement error.

The table below provides the percent capitation for the for-profit plans, the price coefficients from a multinomial logit discrete choice model with hospital fixed effects and fifteen interactions between hospital and patient characteristics (the ML column), and the inequality estimates of the price coefficients. The inequality estimates are for the whole sample but the ML column is for only the least sick patients (lowest severity score).

When Ho and Pakes (2014) ran the ML column for the whole sample all price coefficients were positive. When Ho and Pakes (2014) limit ML to the least sick patients three coefficients are insignificant and the others all are more than an order of magnitude less than the inequality estimates. The inequality estimates are strictly ordered by the capitation rates of plans (the confidence intervals do not overlap) and are both statistically and economically significant. The paper concludes that capitation generates allocations that lower hospital costs, and goes on to quantify its effects and consider its relationship to observable measures of quality. □

Price Coefficients

	%	$\hat{\theta}_{p,\pi}$		$[CI_{LB},\quad CI_{UB}]$	
π	capit	ML	Inequalities		
Pac/care	0.97	-.08 (.01)	-1.50	[-1.68,	-1.34]
Aetna	0.91	-.01 (.02)	-0.92	[-0.95,	-0.86]
HNet	.80	-.04 (.01)	-0.78	[-0.80,	-0.44]
Cigna	0.75	-.02 (.02)	-0.35	[-0.40,	-0.33]
BC	.38	.01 (.01)	-0.29	[-0.31,	-0.25]

Example 8. This example is taken from Dickstein and Morales (2018). That paper shows how weak assumptions on what the analyst knows about how agents form their expectation of returns can be used in discrete choice problems if one is willing to suffice with moment inequalities. In particular they only require that the analyst knows some of the variables in the decision maker's information set when their expectations are formed and the weak

rationality assumption version of C3. Added benefits of their approach are that they can allow for measurement error in observed returns (provided that the instruments used are independent of the measurement error), and they can test whether particular observable variables were known to the decision marker when it made its decision.

The paper investigates the determinants of decisions to export. A firm i is assumed to export to destination j if its expected returns from exporting there, $\mathcal{E}[r_{i,j,t}|I_{i,t}]$, is greater than its fixed cost, $f_{i,j,t} = x_{i,j,t}\theta + \nu_{2,i,j,t}$, where $\nu_{2,i,j,t} \sim N(0, \sigma^2)$. $I_{i,j,t}$ is not known to the analyst, but the analyst does know a subset of these variables, $Z_{i,j,t} \subset I_{i,j,t}$. $x_{i,j,t} \in Z_{i,j,t}$. $\nu_{2,i,j,t} \in I_{i,j,t}$, but it is not known to the analyst; i.e. $\nu_{2,i,j,t} \notin Z_{i,j,t}$ and distributes independent of it. The paper derives two sets of two inequalities.

Letting $d_{i,j,t} = 1$ if the firm exports and zero if not, and $\phi(\cdot)$ be the normal density, revealed preference implies

$$d_{i,j,t} \left(\mathcal{E}[r_{i,j,t} - x_{i,t}\theta|I_{i,j,t}]\right) - \int_{\nu_{2,i,j,t}=-\infty}^{(\mathcal{E}[r_{i,j,t}|I_{i,j,t}]-x_{i,j,t}\theta)} \phi(\nu_2)d\nu_2 \geq 0$$

where, as above, $\mathcal{E}(\cdot|\cdot)$ provides the agent's expectations. Given C3, if we replace the first $\mathcal{E}[r_{i,j,t}|I_{i,j,t}]$ on the right hand side of this equation with the observed returns, say $r^o_{i,j,t}$, and evaluate we generates a ν_1 error which is orthogonal to $Z_{i,j,t}$. The paper shows that Jensen's inequality implies that if we do the same replacement in the upper limit of the integral, evaluate at $\theta = \theta_0$, and take expectation with respect to the DGP, we bound that integral from below, so that the DGP's expectation for the whole equation is positive condition on $Z_{i,j,t}$ The revealed preference equation for not exporting is analogous and use of both inequalities enable the analyst to obtain upper and lower bounds to coefficients.

The paper also derives odds based moments. That is the model implies that if $\Phi(\cdot)$ is the cumulative normal distribution

$$\mathcal{E}\Big[(1 - d_{i,j,t})\frac{\Phi\left(\sigma^{-1}(\mathcal{E}[r_{i,j,t}|I_{i,j,t}] - x_{i,j,t}\theta])\right)}{1 - \Phi\left(\sigma^{-1}(\mathcal{E}[r_{i,j,t}|I_{i,j,t}] - x_{i,j,t}\theta])\right)} - d_{i,j,t} \mid Z_{i,j,t}\Big] = 0,$$

and if we replace $\mathcal{E}[r_{i,j,t}|I_{i,I,t}]$ in this expression with $r^o_{i,I,t}$, evaluate it at $\theta = \theta_0$, Jensen's inequality implies that this expression holds as an inequality. In this case the second inequality is

$$\mathcal{E}\Big[d_{i,j,t}\left(\frac{1-\Phi\left(\sigma^{-1}[r^o_{i,j,t} - x_{i,j,t}\theta]\right)}{\Phi\left(\sigma^{-1}[r^o_{i,j,t} - x_{i,j,t}\theta]\right)}\right) - [1-d_{i,j,t}] \mid Z_{i,j,t}\Big] \geq 0.$$

A couple of final points on this example. First they did not require a normal distribution for the $\nu_{2,i,j,t}$ for their results; any log-concave distribution would do. Second the empirical results in the paper are rather dramatic. They compare their estimates of fixed costs in different Chilean industries to those estimated from; (i) a model with perfect foresight and (ii) from a complete rational expectations model which conditions on the observable $x_{i,j,t}$. The perfect foresight estimates were often a factor of ten larger than the upper bound of the inequality estimates and the rational expectations estimates were often a factor of four or more larger than it. Given the importance of expectations in our models, this suggests we should pay closer attention to how we model them. $\square$

4. Generalized discrete choice approaches

This section provides an approach to identification in models with multiple interacting decision makers that is motivated by and derived from the literature on discrete choice modeling. The starting point is a set of observations on observable outcomes, covariates, and possibly other variables. The analysis then uses the structure of discrete choice that is derived from a well-specified game to relate these data to the payoff functions of the decision makers (like the profit functions of firms) using varied sets of assumptions. These assumptions vary in range and scope and include behavioral assumptions to functional form and distributional assumptions. A key insight here is that payoffs are random from the point of view of the econometrician, and hence in each market the econometrician gets a draw of a payoff function from a distribution of payoff functions. The objective is to learn about this population distribution of payoff functions for the purpose mainly of evaluation but also policy simulations. The main analysis we do here is under complete information in a stylized framework which allows us to focus on intuition. We also point to generalizations to nonparametric settings, weaker assumptions on information, and other modeling decisions.

4.1. **Models of discrete games with complete information.** There is another approach to studying the econometrics of models with strategic interactions that follows at its core the usual discrete choice approaches used initially in single agent choice problems. As in the multiple agent models in the previous section, the defining feature of the models is that the utility function of each of the decision makers also depends on the decisions made by the other decision makers. For example, in a model of market entry, the utility functions are the profit functions of the potential entrants, which depend on the entry decisions of the other potential entrants, reflecting the impact of competition on profits. This approach here takes as its initial objective identification of the utility functions (or payoff functions) of the decision makers. Often, the ultimate objective, as in the previous section, is to use these identification results to produce counterfactual predictions.

In various forms, econometric models based around game theory models have been studied beginning with early work in Bjorn and Vuong (1984), Jovanovic (1989), Bresnahan and Reiss (1990, 1991a,b), Berry (1992) and Tamer (2003), along with more standard models of discrete choice with simultaneity in Heckman (1978a) and Blundell and Smith (1994). The literature on the "econometrics of games" has been recently reviewed also in de Paula (2013) and Aradillas-López (2020). Although this part focuses on the case of discrete action spaces, Section 4.4 on auction models focuses on an importance instance of continuous action spaces.

In these models it is not possible to use the assumption that the decision makers maximize utility, as would be familiar from standard models with a single decision maker, because the utility functions of the decision makers depend on decisions made by the other decision makers. The econometrician must make assumptions concerning how the decision makers deal with this feature of their utility functions, which can involve making assumptions about how the decision makers predict and respond to how the other decision makers will make their decisions (relating to the solution concept), and information assumptions about what the decision makers know (relating to the information structure). In many ways, these assumptions are interrelated, as the details of the solution concept will depend on the details of the information structure as highlighted throughout this chapter.

We illustrate these ideas using a simple stylized two player entry game with complete information where both players know their own and their opponent's payoffs. This assumption can be relaxed, as was a key highlight of the approach in the previous section. The structure of this game is such that the payoffs are normalized so that each player i gets 0 payoff if it takes action $d_i = 0$. Thus, $\pi^1(0, y_2) = 0 = \pi^2(y_1, 0)$. This game is summarized in Table 1.

	$d_2 = 0$	$d_2 = 1$
$d_1 = 0$	$(0, 0)$	$(0, \pi^2(0, 1))$
$d_1 = 1$	$(\pi^1(1, 0), 0)$	$(\pi^1(1, 1), \pi^2(1, 1))$

Table 1. A nonparametric two-player, two-action game in normal form

In this setting, we assume that the econometrician has access to an i.i.d. sample of realized entry decisions on T observations of the game (d_{1t}, d_{2t}), for $t = 1, \ldots, T$, where one can think of the t subscript as market t. The identification problem is how to link these data to the model of behavior based on the above game. At this point, the game is nonparametric in that no further assumptions are made on the π's. For now, the objects of interest are the best response functions. The best response of firm i when the entry decision of firm $-i$ is d_{-i} is denoted $Br^i(d_{-i})$. The argument d_{-i} of the best response function refers to an entry decision conjectured by the econometrician, not a realized entry decision observed in the data.[24] The best response functions are:

$$Br^1(d_2) = 1[d_2 = 1]1[\pi^1(1, 1) > 0] + 1[d_2 = 0]1[\pi^1(1, 0) > 0]$$

[24]Payoffs are assumed to be in general position, in the sense that firm i is never indifferent between entering and not entering in response to an entry decision of firm $-i$. This is equivalent to $\pi^1(1, 1) \neq 0$, $\pi^1(1, 0) \neq 0$, $\pi^2(1, 1) \neq 0$ and $\pi^2(0, 1) \neq 0$.

and

$$Br^2(d_1) = 1[d_1 = 1]1[\pi^2(1,1) > 0] + 1[d_1 = 0]1[\pi^2(0,1) > 0].$$

We view the payoff function π across markets as random from the perspective of the econometrician, and so the best response functions are also random from the perspective of the econometrician. Then, this motivates one to consider the best response probabilities (where this randomness is generated by the i.i.d. distribution over markets):

$$P(Br^1(d_2) = 1) = P\left(1[d_2 = 1]1[\pi^1(1,1) > 0] + 1[d_2 = 0]1[\pi^1(1,0) > 0] = 1\right)$$

and

$$P(Br^2(d_1) = 1) = P\left(1[d_1 = 1]1[\pi^2(1,1) > 0] + 1[d_1 = 0]1[\pi^2(0,1) > 0] = 1\right).$$

These are the probability distributions over the response to the decision of the other firm. Notice here that $P(Br^i(d_{-i}) = 1)$ is the probability from the perspective of the econometrician that firm i would enter the market if firm $-i$ *were mandated to have entry decision* d_{-i}, and firm i were allowed to re-optimize its entry decision. The econometrician then asks what can be learned about these best response probabilities given observations on *realized entry* decisions. This requires assumptions on behavior and here we illustrate with the requirement of Nash equilibrium play[25] allowing for mixed strategies. This requires the following condition. If (d_1^*, d_2^*) is a Nash equilibrium choice for players 1 and 2, then it must be true that

$$\pi^1(d_1^*, d_2^*) \geq \pi^1(d_1', d_2^*) \quad \text{and} \quad \pi^2(d_1^*, d_2^*) \geq \pi^2(d_1^*, d_2') \tag{6}$$

Here, this allows for mixed strategies in that if d_1^* (d_2^*) are mixed strategies, then the π^1 (π^2) are interpreted as expected payoffs with respect to the mixed strategy profile induced by (d_1^*, d_2^*). A similar condition, termed a primitive condition, is used in Section 3 and in Pakes (2010). To get an understanding of how realized entry data relates to the best responses, consider the following question. If we observe $(d_{1t} = 1, d_{2t} = 0)$ in market t, does it mean

[25]For more here on other forms of solution concepts such as rationalizability, see Kline and Tamer (2012).

that $(1,0)$ lie on the best response curves defined above? In other words, does it mean that

$$Br_t^1(d_{2t}=0) = 1[\pi_t^1(1,0)>0] = 1 \quad \text{and} \quad Br_t^2(d_{1t}=1) = 1[\pi_t^2(1,1)>0] = 0$$

or equivalently that $\pi_t^1(1,0)>0$ and $\pi_t^2(1,1)<0$? The answer is negative in general, unless we are assuming Nash equilibrium in pure strategies.[26] This is because, with mixed strategies even with Nash equilibrium, it is not necessarily the case that data lie on best response functions. The actions realized from mixed strategies are not necessarily best responses to each other.

The data that we observe allow us to obtain the probabilities of the four outcomes $(1,1),(0,0),(1,0)$ and $(0,1)$. Identification explores the link between these observed probabilities and the π's. This randomness - that in different markets we observe different outcomes - is a result of the fact that from the point of view of the econometrician, the π's are nondegenerate random variables that have a probability induced by a model of a game which is the object of interest. To proceed forward, we require a specification of this model.

4.2. **Simple game example.** We illustrate some approaches to identification when we assume equilibrium in pure strategies in the game above under a specification that only includes "structural errors." Let the payoff functions depend on a vector $z=(z_1',z_2')'$ that is observed by both players. In addition, assume that for $z_1=(x_1',\epsilon_1)'$ and $z_2=(x_2',\epsilon_2)'$, the econometrician observes both x_1 and x_2 but does not observe the scalars ϵ_1 and ϵ_2. As in Pakes (2010), we term such unobservables "structural errors." These are similar to the standard unobservables in the discrete choice literature and are interpreted as part of the payoff functions that are observed by the players but not by the econometrician. Therefore, from the perspective of the econometrician, the payoff functions are random.

Often, standard assumptions from economic theory and the observable data imply only weak restrictions on the utility functions. In particular, it can happen that there are multiple specifications of the utility functions of the decisions makers that are compatible with the

[26] As a reminder, in games there are situations where the only Nash equilibrium uses mixed strategies.

assumptions and that result in the same behavior of the decision makers, and therefore that result in the same observable data. In such cases, the model is partially identified. The main issues exist even in the simplest possible non-trivial game, with two players each of which makes a binary decision. Within empirical Industrial Organization, this can be used as a model of market entry where Kline and Tamer (2012) show that the model is not point identified. This game can also be the building block for more complicated models, like models of product positioning as mentioned in Section 4.3.1. In other subfields of economics, this can be used for other purposes, including as a model of social interactions (e.g., Kline and Tamer (2020)).

Here, we focus on using the above Nash equilibrium condition (Equation 6) in the context of a parametric specification of the payoff functions. This utility function has a parametric form that is assumed to depend in a linear and additively separable way on observables x_i and unobservables ϵ_i. It is useful to define $x = (x_1, x_2)$ and $\epsilon = (\epsilon_1, \epsilon_2)$. Some observables can be in both x_1 and x_2, reflecting market-level observable characteristics, for example some measure of the "size" of the market available to the potential entrants. It is often important to allow correlation between ϵ_1 and ϵ_2, to allow market-level unobservables. Linking this to Section 3 above, this is where we do not have uncertainty from the perspective of the firms, both firms know the payoff functions (and there is no misspecification), and the data are not subject to any measurement error. So, the error ν_1 is not present. But, ν_2 represents (ϵ_1, ϵ_2) and this ν_2 is usually assumed to be statistically independent[27] of x. The game can be represented in normal form as in Table 2.

	$d_2 = 0$	$d_2 = 1$
$d_1 = 0$	$(0, 0)$	$(0, x_2\beta_2 + \epsilon_2)$
$d_1 = 1$	$(x_1\beta_1 + \epsilon_1, 0)$	$(x_1\beta_1 + \Delta_1 + \epsilon_1, x_2\beta_2 + \Delta_2 + \epsilon_2)$

Table 2. A parametric two-player, two-action game in normal form

[27]It is possible to weaken the independence to quantile or median independence. However, mean independence is not possible since it will not lead to any restrictions. This was shown in the context of binary choice models in Manski (1988b).

In this model applied to market entry decisions, $x_i\beta_i + \epsilon_i$ is the monopoly profits earned by firm i, and $x_i\beta_i + \Delta_i + \epsilon_i$ is the duopoly profits earned by firm i, so that Δ_i is the effect that firm $-i$ entering the market has on the profits earned by firm i. The discussion here will take Δ_i to be a fixed parameter, but Δ_i can be modeled as a random parameter, depending either on observable and/or unobservable characteristics of the firms. The condition that firms earn 0 profits when not entering the market is both a reasonable approximation and a normalization because only differences in utility functions have observable implications. The typical goal of the econometrician is to identify the parameters $\beta = (\beta_1, \beta_2)$ and the distribution of $\epsilon = (\epsilon_1, \epsilon_2)$, although it is possible to consider other objects of interest with different identification strategies. Linking back to the nonparametric specification of the profit functions (and suppressing the functional dependence of the left hand sides of the following equations on ϵ and x):

$$\begin{aligned}
\pi^1(d_1, d_2) &= d_1 \times (x_1\beta_1 + \Delta_1 d_2 + \epsilon_1) \\
\pi^2(d_1, d_2) &= d_2 \times (x_2\beta_2 + \Delta_2 d_1 + \epsilon_2)
\end{aligned}$$

Using the assumption of Nash equilibrium in pure strategies, observing (d_{1t}, d_{2t}) in market t implies the following inequalities:

$$\begin{aligned}
d_{1t} \times (x_{1t}\beta_1 + \Delta_1 d_{2t} &+ \epsilon_{1t}) - (1 - d_{1t}) \times (x_{1t}\beta_1 + \Delta_1 d_{2t} + \epsilon_{1t}) \geq 0 \qquad (7) \\
d_{2t} \times (x_{2t}\beta_2 + \Delta_2 d_{1t} &+ \epsilon_{2t}) - (1 - d_{2t}) \times (x_{2t}\beta_2 + \Delta_2 d_{1t} + \epsilon_{2t}) \geq 0
\end{aligned}$$

This means that

$$\text{observe} \quad (d_{1t}, d_{2t}) \implies \begin{array}{l} (2d_{1t} - 1) \times (x_{1t}\beta_1 + \Delta_1 d_{2t} + \epsilon_{1t}) \geq 0 \\ (2d_{2t} - 1) \times (x_{2t}\beta_2 + \Delta_2 d_{1t} + \epsilon_{2t}) \geq 0 \end{array}$$

Further, if we take expectations[28] of the above right hand side with respect to the population distribution ("adding up" across markets), we get that

$$E[(2d_{1t}-1)\times(x_{1t}\beta_1+\Delta_1 d_{2t}+\epsilon_{1t})]=E[(2d_{1t}-1)(x_{1t}\beta_1+\Delta_1 d_{2t})]+E[(2d_{1t}-1)\epsilon_{1t}]\geq 0$$
$$E[(2d_{2t}-1)\times(x_{2t}\beta_2+\Delta_2 d_{1t}+\epsilon_{2t})]=E[(2d_{2t}-1)(x_{2t}\beta_2+\Delta_2 d_{1t})]+E[(2d_{2t}-1)\epsilon_{2t}]\geq 0$$

Notice here that the above *moment inequalities* are an *implication* of the model as we may have multiple equilibria and so one can expect these inequalities to not necessarily use all the information in the model assumptions (more on sufficient conditions below). The two terms $E[(2d_{it}-1)\epsilon_{it}]$, $i=1,2$ above can be written as

$$E_x\left[E[\epsilon_{it}|d_{it}=1,x]P(d_{it}=1|x)-E[\epsilon_{it}|d_{it}=0,x]P(d_{it}=0|x)\right].$$

These are the *selection terms* that are similar to ones in the usual selection models. It is possible that if the ϵ's are pure measurement errors that d and ϵ are independent. But, in this case with structural errors, this selection term will not vanish and presents a serious identification issue[29] that would need to be dealt with. This is the fundamental identification problem that the structural error method faces. When a decision is made based on variables that are payoff relevant and observed by the agent and unobserved to the econometrician, the econometrician must confront this selection problem.

Remark 1. The approach above that targets necessary conditions for Nash equilibrium behavior is simpler to implement. For instance, when the data are assumed to lie on the best

[28]One does not need to use expectations only. Any functional that respects first order stochastic dominance can be used here, such as a quantile.

[29]It is possible to try to use semiparametric econometric methods that are used for selection correction in single agent models such as Ahn and Powell (1993). We do not pursue this here as using those matching approaches -where one would need to match a nonparametric function- along with moment inequalities is an open question.

response functions, then at a given observed $\mathbf{d} = (d_1, \ldots, d_K)$ for K players, we have

$$\text{observe} \quad (d_{1t}, \ldots, d_{Kt}) \implies \begin{array}{l} (2d_{1t} - 1) \times (x_{1t}\beta_1 + S(\mathbf{d}_{-1t}; \theta_1) + \epsilon_{1t}) \geq 0 \\ \vdots \\ (2d_{Kt} - 1) \times (x_{Kt}\beta_K + S(\mathbf{d}_{-Kt}; \theta_K) + \epsilon_{Kt}) \geq 0 \end{array}$$

where $S(\mathbf{d}_{-1t}; \theta_1)$ is the impact of the observed actions of all players $j \neq 1$, and similarly for $S(\mathbf{d}_{-Kt}; \theta_K)$. These necessary conditions can be generalized in a natural way to cover games where the action space is nonbinary (such as deciding on the number of locations to enter in a given market rather than just binary enter or not enter). This is attractive from a practical perspective since these inequalities derived from necessary conditions of Nash equilibrium behavior do not require that the econometrician solves for the equilibria of the game. They also hold in general discrete games with non-binary actions and many players.

Finally, it would also be important to use the *joint* distribution of outcomes (given covariates) to be able to address the question of inference on the correlation between ϵ_1 and ϵ_2 which would represent market-specific unobservables. This in principle would be possible if the above were treated as a system of moment inequalities where the covariance matrix of these errors is part of the estimation procedures.

Remark 2 (Comments on the ν_2 errors in Equation 2 above:)**.** In general, as in the traditional discrete choice literature, the analysis following the above proceeds with the assumption that the "ν_2" errors (the ϵ) are independent of the covariates x but are otherwise unrestricted.[30] In particular, no restrictions are made on their support. Support restrictions in principle can lead to identification power but these types of support restrictions are hard to motivate. The independence of these errors and the covariates is an important and restrictive assumption. For instance, it rules out the "$y(\cdot)$" variables from the previous section, so we cannot have for example prices among the explanatory variable x's that could react to various actions d.

[30] For instance, in binary choice models, $E[\epsilon|d = 1, x] = E[\epsilon|x\beta + \epsilon \geq 0, x]$. When $\epsilon \perp x$, this quantity is equal to $E[\epsilon|\epsilon \geq -x\beta]$ which is the usual selection term. Otherwise when ϵ and x are not independent, one would usually need to model the joint distribution of ϵ and x.

One way out of this is to model the determination of prices[31]. Other types of variables that are ruled out by this independence assumption arise in "games played on networks" where the game is played among firms that have decided to (endogenously, based on the x and ϵ) connect together. This endogenous network formation game can lead to endogeneity of variables that impact firm payoffs[32].

Remark 3 (Comments on allowing for ν_1 errors:)**.** It is possible to explore the identification problem when we allow for a particular source of ν_1 error as in the last section. In particular, first allow for only ν_1 classical measurement error and no ν_2 errors in the model of the simple game above. A caveat in this model is that without ν_2, we are assuming that the variables the players observe are the same ones that are observed by the econometrician.[33] In the extreme case without ν_1 and ν_2 errors then markets with the same x should have the same set of equilibrium outcomes, though with multiple equilibria, it is possible that the realized equilibrium differs across markets with the same x. On the other hand, with only measurement error (and no ν_2) of classical kind, the model we have is one where we do not observe x, rather we observe $x_i^* = x_i + \eta_i$ for $i = 1, 2$, where η_i is uncorrelated with x_i (or statistically independent of x_i). The inequalities in Equation 7 become

$$
\begin{aligned}
(2d_{1t} - 1) \times (x_{1t}^* \beta_1 + \Delta_1 d_{2t} + \eta_{1t}\beta_1) &\geq 0 \\
(2d_{2t} - 1) \times (x_{2t}^* \beta_2 + \Delta_2 d_{1t} + \eta_{2t}\beta_2) &\geq 0
\end{aligned}
$$

By construction, the measurement error $\eta_t = (\eta_{1t}, \eta_{2t})$ is correlated with x_t^*. This is a discrete choice model where measurement error in the covariate is the only error in the model. It is possible here to explore an instrumental variable assumption that requires a random variable

[31]An example of this approach in the context of a fully parametric model is in Ciliberto, Murry, and Tamer (2021).
[32]An example of games played on networks is the work of Kline (2015).
[33]This is clearly restrictive in these settings as the model may appear deterministic.

z that is independent of η but correlated with x (and hence x^*). We leave further development of such methods for future research.

Another possibility for the existence of ν_1 is specification error, but allowing for such an error may be nontrivial. For instance, consider the nonparametric normal form game in Table 1 above. Without allowing for the ν_2 error, we can have that the econometrician uses a misspecified version of the payoffs (the πs). For instance, the econometrician uses $x_1\beta_1 + \Delta_1$ instead of $\pi^1(1,1)$ such that

$$\pi^1(1,1) = x_1\beta_1 + \Delta_1 + [\pi^1(1,1) - x_1\beta_1 - \Delta_1] = x_1\beta_1 + \Delta_1 + \epsilon_1$$

and similarly for the other payoffs in Table 1. To proceed further, assumptions must be made on ϵ_1, reflecting the specification error. What is required here to allow for this kind of misspeficication is to require that ϵ be independent of x. This is a strong assumption in this setup as it would seem that this misspecification error is unlikely to satisfy this independence assumption. In addition, from the point of view of the econometrician, the identification problem is the same as the one above. One caveat of this approach to including only specification error is that, for given covariate values, firms have the same payoffs in every market.

4.3. **Using both necessary and sufficient conditions for Nash equilibrium.** The generalized discrete choice approach uses both necessary and sufficient conditions of Nash equilibrium behavior to obtain the sharp identified set. We explain this insight first and then derive the sharp inequalities. The inequalities in Equation 7 use *necessary* conditions for pure strategy Nash equilibrium: For instance, if we see the outcome $(1,0)$ in market t then, this implies that

$$(y_{1t}, y_{2t}) = (1,0) \implies \quad x_{1t}\beta_1 + \epsilon_{1t} \geq 0 \quad \text{and} \quad x_{2t}\beta_2 + \Delta_2 + \epsilon_{2t} \leq 0$$

But, it is possible to also have the reverse logical implication in some cases: For instance (and ignoring probability zero events when the payoffs are equal to zero), when $x_{1t}\beta_1 + \epsilon_{1t} \geq 0$

and $x_{2t}\beta_2 + \epsilon_{2t} \leq 0$ then $(1, 0)$ is a dominant strategy and hence we get

$$x_{1t}\beta_1 + \epsilon_{1t} \geq 0 \quad \text{and} \quad x_{2t}\beta_2 + \epsilon_{2t} \leq 0 \implies \quad (y_{1t}, y_{2t}) = (1, 0)$$

The event $(1, 0)$ is sandwiched between two events in terms of $(\epsilon_{1t}, \epsilon_{2t})$. Indeed, the inequality from the left hand side (sufficient condition) is a region for ϵ where the only Nash equilibrium (in pure strategies) is $(1, 0)$ while the inequality from the right (necessary condition) is one where $(1, 0)$ is one of the equilibria. If for all realizations of the structural errors ϵ there is uniqueness then necessary and sufficient conditions would yield to moment equalities. To implement the above insight in a general model, we need to solve the game for essentially all values (or many draws) of the ϵs (given covariates) and enumerate the equilibria for each one of these draws.

We now analyze in detail the above game by using all the information. We use a parametric assumption on the distribution of $\epsilon = (\epsilon_1, \epsilon_2)$ to obtain a set of moment conditions.[34] Using the assumption that the solution concept is pure strategy Nash equilibrium and the assumption of complete information (i.e., the firms have common knowledge of the profit functions), and the assumption that the unobservables are independent from the observables $\epsilon \perp x$, the data generating process from this model can be characterized by the following three equations relating the observed probabilities of the outcomes on the left of the equation to the utility functions on the right of the equation. This analysis assumes that $\Delta_1 \leq 0$ and $\Delta_2 \leq 0$, so that there is a negative effect of competition on profits in the entry model. A similar analysis can be conducted when it is assumed that $\Delta_1 \geq 0$ and $\Delta_2 \geq 0$

Figure 1 shows the mapping between the value of (ϵ_1, ϵ_2) and the outcome of the game, for given specification of (x_1, x_2). Let $S_\beta(x_1, x_2) = \{\epsilon : -x_1\beta_1 \leq \epsilon_1 \leq -x_1\beta_1 - \Delta_1, -x_2\beta_2 \leq \epsilon_2 \leq -x_2\beta_2 - \Delta_2\}$. If $(\epsilon_1, \epsilon_2) \in S_\beta(x_1, x_2)$ for given (x_1, x_2), then there are two pure strategy Nash

[34] In principle, this approach remains possible without imposing this parametric distributional assumption on the joint distribution of the errors, as one can then replace this joint distribution flexibly with a sieve approximation. Methods of inference in partially identified models with unknown functions (densities in this case) is a current topic of research in econometric theory. See for example Chen, Tamer, and Torgovitsky (2011).

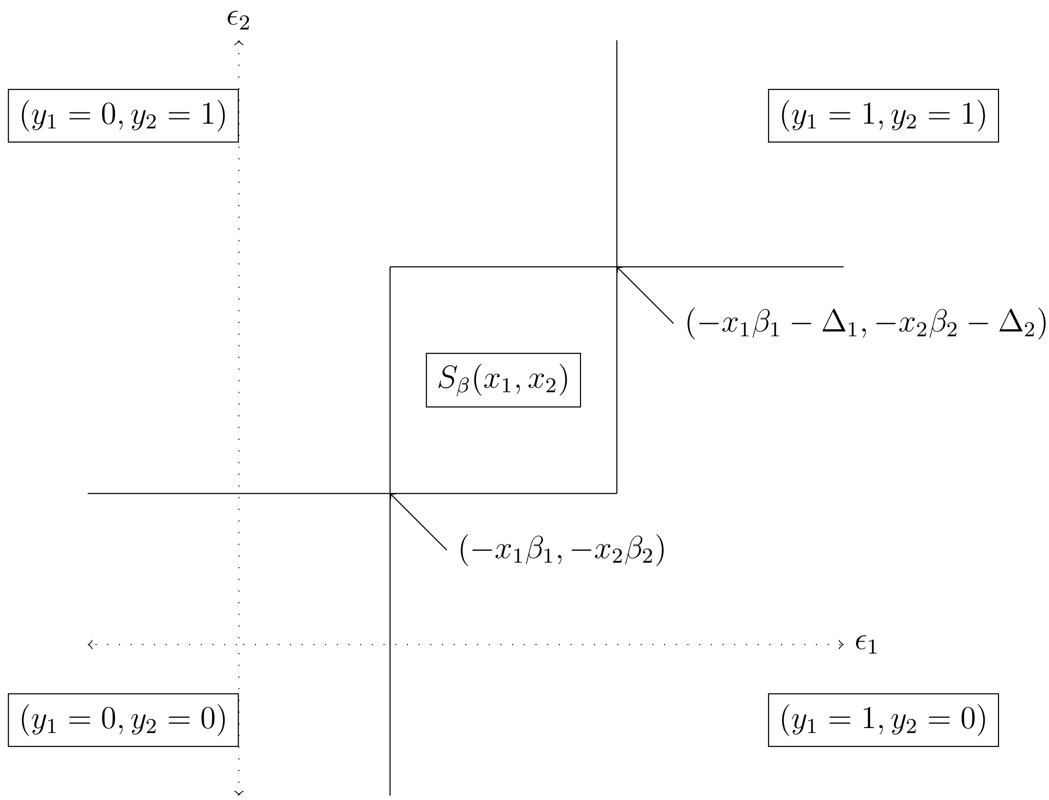

FIGURE 1. Mapping from ϵ to outcomes

equilibria of the game. It is a Nash equilibrium for firm 1 to be a monopolist entrant, and it is a Nash equilibrium for firm 2 to be a monopolist entrant. The realized entrant depends on which equilibrium is selected.

The following equations characterize the model.

$$
\begin{aligned}
P(y_1 = 0, y_2 = 0|X = x) &= P(\epsilon_1 \leq -x_1\beta_1, \epsilon_2 \leq -x_2\beta_2) \qquad (8) \\
P(y_1 = 1, y_2 = 1|X = x) &= P(\epsilon_1 \geq -x_1\beta_1 - \Delta_1, \epsilon_2 \geq -x_2\beta_2 - \Delta_2) \\
P(y_1 = 1, y_2 = 0|X = x) &= P(\epsilon_1 \geq -x_1\beta_1, \epsilon_2 \leq -x_2\beta_2 - \Delta_2, \epsilon \notin S_\beta(x_1, x_2)) \\
&\quad + P(y_1 = 1, y_2 = 0|X = x, \epsilon \in S_\beta(x_1, x_2)) \times P(\epsilon \in S_\beta(x_1, x_2))
\end{aligned}
$$

The left hand sides of these equations are observed by the econometrician, so the identified set consists of all specifications of β_1, β_2, Δ_1, Δ_2, and distributions of ϵ that are such that the right hand sides of these equations match the left hand sides. There is no equation for

$P(y_1 = 0, y_2 = 1|X = x)$ displayed above because it would be redundant: by construction, $P(y_1 = 0, y_2 = 1|X = x) = 1 - P(y_1 = 0, y_2 = 0|X = x) - P(y_1 = 1, y_2 = 1|X = x) - P(y_1 = 1, y_2 = 0|X = x)$, so the model predictions for $P(y_1 = 0, y_2 = 1|X = x)$ do not place any additional restrictions on the parameters of the utility function.

The first and second equations are straightforward. The first equation indicates that the probability that neither firm enters the market is equal to the probability that neither firm is profitable as a monopolist. This reflects that neither firm entering the market in a pure strategy Nash equilibrium is equivalent to neither firm being profitable as a monopolist. The second equation indicates that the probability that both firms enter the market is equal to the probability that both firms are profitable as duopolists. This reflects that both firms entering the market in a pure strategy Nash equilibrium is equivalent to both firms being profitable as duopolists. These equations are not so different from equations that arise in models with a single decision maker.

The third equation is the place that this model distinguishes itself from standard models with a single decision maker. The third equation indicates that the probability that firm 1 enters the market as a monopolist is equal to the sum of two probabilities. The first probability on the right hand side is the probability of the event that ϵ satisfies $\epsilon_1 \geq -x_1\beta_1, \epsilon_2 \leq -x_2\beta_2 - \Delta_2, \epsilon \notin S_\beta(x_1, x_2)$, in which case firm 1 being a monopolist is the unique pure strategy Nash equilibrium. The second probability on the right hand side is the product of the probability of the event that $\epsilon \in S_\beta(x_1, x_2)$ times the conditional probability that firm 1 is a monopolist given that condition on profits. This reflects that when $\epsilon \in S_\beta(x_1, x_2)$ there are two pure strategy Nash equilibria in which one or the other firm enters the market as a monopolist. In this particular model, this can be known as the region of multiple equilibria, without any ambiguity. The assumption of pure strategy Nash equilibrium does not uniquely predict the outcome of the game when the utility functions are in the region of multiple equilibria, so empirical models must be augmented by a selection mechanism that does select which of the potential equilibrium outcomes is realized in the data. In this particular model, there is just one selection mechanism $P(y_1 = 1, y_2 = 0|X = x, \epsilon \in S_\beta(x_1, x_2))$ per value of X, because there is just one region of multiple equilibria, but in other models, there could be multiple regions of multiple equilibria with different sets of potential equilibrium outcomes, requiring multiple selection mechanisms.

In this specification of the model, the equilibrium selection mechanism conditions on $X = x$ and the event that $\epsilon \in S_\beta(x_1, x_2)$. This suffices to write down an equation for $P(y_1 = 1, y_2 = 0|X = x)$ that can be used for identification analysis. In principle, however, the equilibrium selection mechanism could condition on $X = x$ and particular values of ϵ, because the players know ϵ and therefore equilibrium selection can depend on ϵ. Therefore, if assumptions are placed on the equilibrium selection mechanism, it can be more natural to work with the selection mechanism that conditions on particular values of ϵ. Any such assumptions would imply restrictions on $P(y_1 = 1, y_2 = 0|X = x, \epsilon \in S_\beta(x_1, x_2))$ after integrating over ϵ.

Thus, it turns out that the event that firm 1 is a monopolist is not equivalent to an event simply in terms of the utility functions, because of the existence of multiple equilibria. This complicates the identification analysis.

In short, the equations relating the utility functions and the observed outcomes of the game depends on the selection mechanism that selects the realized outcome from the set of outcomes compatible with the assumptions. Without any further assumptions on the selection mechanism, the econometrician knows only that the selection mechanism is a valid probability, so $P(y_1 = 1, y_2 = 0|X = x, \epsilon \in S_\beta(x_1, x_2)) \in [0, 1]$. This results in the inequalities that $P(\epsilon_1 \geq -x_1\beta_1, \epsilon_2 \leq -x_2\beta_2 - \Delta_2, \epsilon \notin S_\beta(x_1, x_2)) \leq P(y_1 = 1, y_2 = 0|X = x) \leq P(\epsilon_1 \geq -x_1\beta_1, \epsilon_2 \leq -x_2\beta_2 - \Delta_2, \epsilon \notin S_\beta(x_1, x_2)) + P(\epsilon \in S_\beta(x_1, x_2))$. In models with multiple regions of multiple equilibria, similar inequalities can be derived for each outcome that can arise as part of a region of multiple equilibria, using the condition that the selection mechanisms necessarily satisfy the condition of being a valid distribution, as in Ciliberto and Tamer (2009).

Remark 4. It is worth having an example of the identified sets in this setting, which can also be used to discuss some common considerations when working with partially identified models. In particular, this discussion shows how different displays of the identified set helps with the understanding of what is being learned (or assumed) about the model.

Consider a case with $\beta_{11} = 1 = \beta_{21}$ and $\beta_{12} = 0.75 = \beta_{22}$ and $\Delta_1 = -1.75 = \Delta_2$. Correspondingly, $x_{11} = 1 = x_{21}$, reflecting an intercept. And, x_{12} and x_{22} are binary (0/1) variables. Given the definition of identification in a conditional model, from the perspective of identification, the specific distribution of x_{12} and x_{22} is not relevant, although the support is relevant. In the region of multiple equilibria, there is equal probability that one or the other potential entrant will be the monopolist entrant. Finally, ϵ is normally distributed with mean zero, unit variances, and covariance of 0.3 between ϵ_1 and ϵ_2.

With this setup, the probability of a no-entry market outcome in which no firm enters ranges between approximately 1% and approximately 5%, for different values of x. The probability of a duopoly market outcome in which both firms enter ranges between approximately 8% and approximately 30%, for different values of x. The probability of a monopoly market outcome in which one firm enters comprises the rest of the probability, ranging between approximately 70% and 87%, for different values of x. In other words, this setup models a setting in which most markets are served by a single entrant.

The analysis assumes that Δ_1 and Δ_2 are non-positive, reflecting a negative effect of competition on profits. The analysis also assumes that the unobservables have a normal distribution with mean zero, unit variances, and unknown correlation. The correlation is therefore a parameter of the model. Further, the analysis assumes that the correlation between ϵ_1 and ϵ_2 is positive, and less than 0.95, reflecting a positive correlation between the unobservables affecting the profits of the two potential entrants, allowing in particular market-level unobservables. Because the assumption of unit variance of the unobservables is the scale normalization, there are not any assumptions or scale restrictions on β.

Figure 2 displays the identified set for (β_{11}, β_{12}), the "non-strategic" parameters of the utility function of player 1 associated with the intercept and the explanatory variable. Even

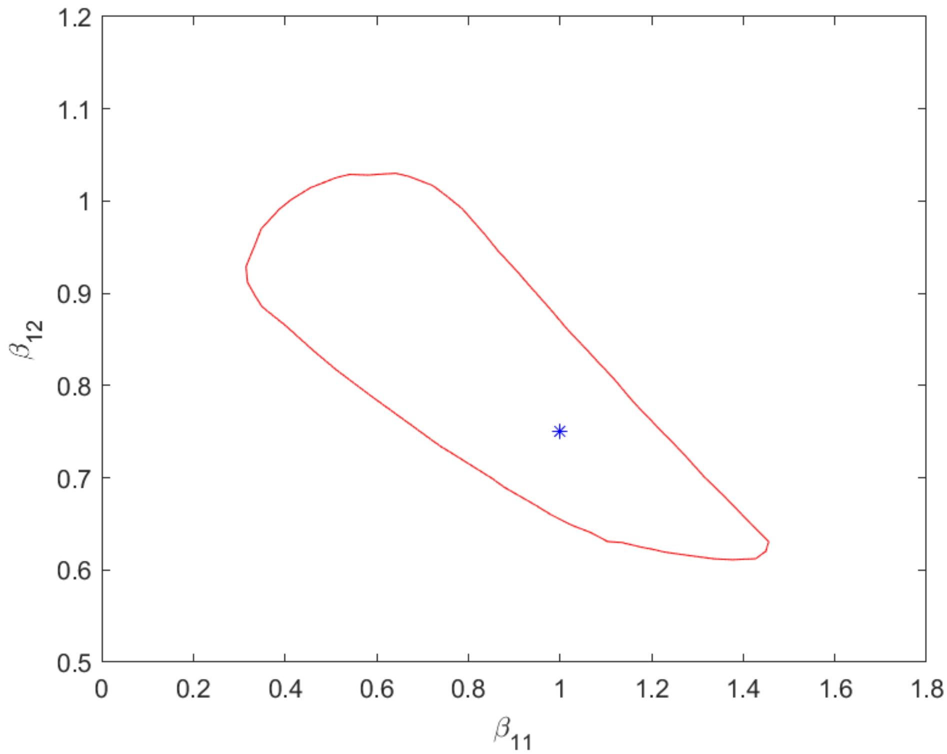

FIGURE 2. Identified set for (β_{11}, β_{12}) without restrictions on the parameters across players

with the parametric distributional assumption on the unobservables, in this setup there are only relatively weak restrictions placed on the parameters given the assumptions and the data. Although the true value of the parameter $(1, 0.75)$ is necessarily contained in the identified set, as represented by the blue star in the figure, perhaps contrary to intuition the true value is not in the "middle" of the identified set.[35]

With partial identification, it is possible to explore how different assumptions influence what can be learned about the parameters. One reasonable assumption in this setting is that

[35]Particularly when computing identified sets for a component θ_1 of the parameter vector $\theta = (\theta_1, \theta_2)$ rather than the entire parameter vector θ, there is some unavoidable numerical error. This is because, in order to determine whether a particular value θ_1^* of θ_1 is in the identified set, it is necessary to somehow profile over possible values θ_2^* of θ_2, to determine whether there is some such θ_2^* such that (θ_1^*, θ_2^*) is in the identified set. Because it is impossible to check literally every possible value of θ_2^*, some numerical tolerance must be allowed, such that θ_1^* is determined to be part of the identified set if there can be found a θ_2^* such that (θ_1^*, θ_2^*) is at least approximately (up to numerical tolerance) in the identified (e.g., satisfies the restrictions up to a numerical tolerance). The result is that the computed identified sets might contain some values of the parameter that are not actually in the identified set; therefore, these computed identified sets can be viewed as (slightly) non-sharp identified sets due to this numerical tolerance.

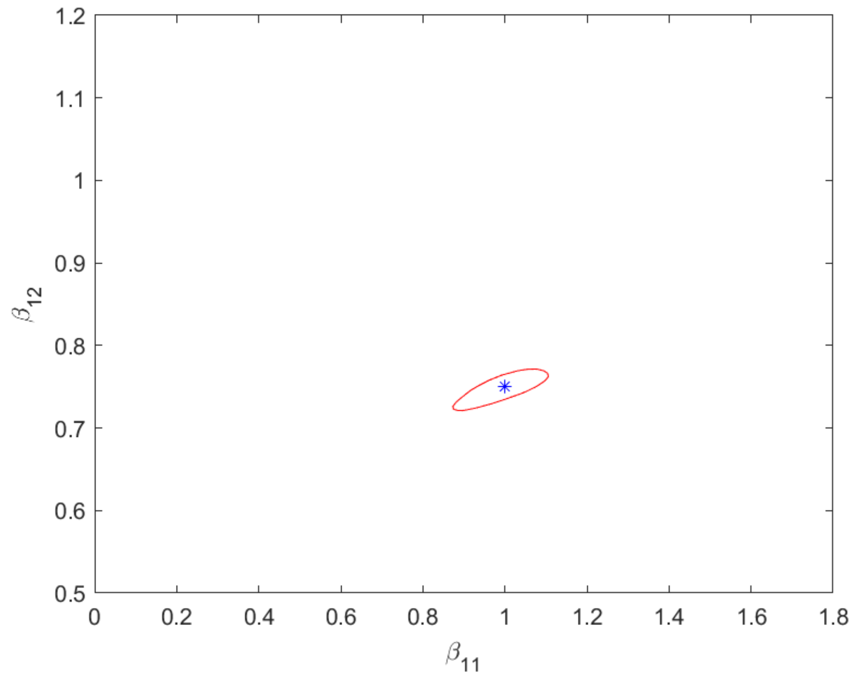

FIGURE 3. Identified set for (β_{11}, β_{12}) with parameters restricted to be equal across players

the parameters of the utility function of player 1 are equal to the parameters of the utility function of player 2. This assumption is true in the data generating process. Figure 3 displays the identified set for (β_{11}, β_{12}) with this additional assumption. In Figure 3 the identified set is substantially smaller than the identified set in Figure 2, reflecting the additional restrictions on the parameter implied by this additional assumption.

One way to see why there is such identifying power of the assumption of equality of parameters across players is to display the identified set for the parameters across players, without the assumption of equality across players. Figure 4 displays the identified set for (β_{12}, β_{22}), which is the identified set for the slopes with respect to the explanatory variables, across players. Much of this identified set would violate the assumption that the parameters are equal across players. This explains why adding this assumption substantially shrinks the size of the identified set.

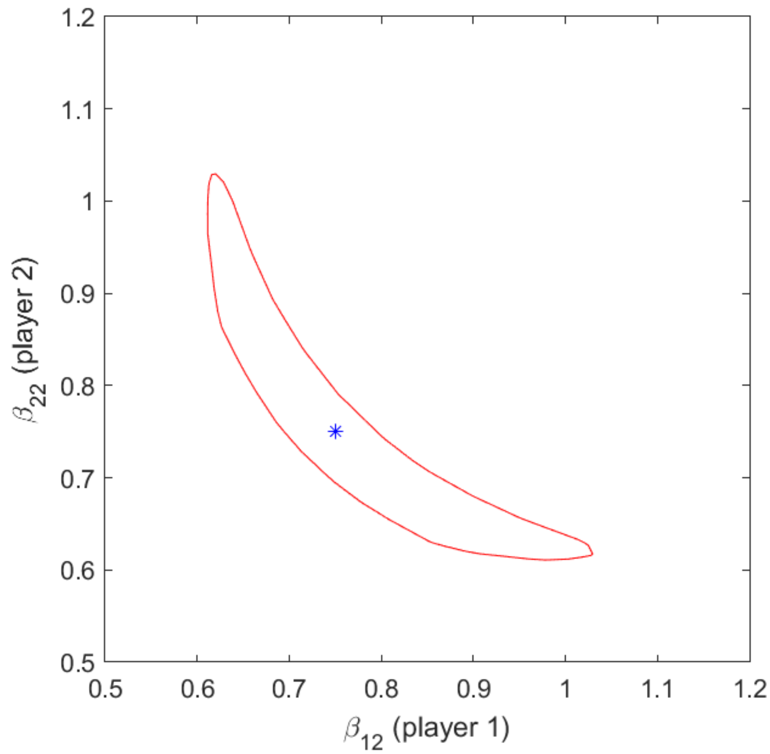

FIGURE 4. Identified set for (β_{12}, β_{22}) without restrictions on the parameters across players

Often, the interaction effect parameters (Δ_1, Δ_2) are of particular interest. Figure 5 displays the identified set for (Δ_1, Δ_2) without restrictions on the parameters across players. If the parameters were restricted to be equal across players, then the identified set for $\Delta_1 = \Delta_2$ would be approximately $[-1.93, -1.53]$.

In general with partially identified models, even with the specification of some of the parameters of the model, there is still partial identification of the other parameters of the model. This is evident from the figures displayed here, where for example even with the specification of β_{11} there is still partial identification of β_{12}.

However, in some models, the specification of some of the parameters of the model corresponds to a unique specification of the rest of the parameters of the model that is simultaneously in the identified set and compatible with that partial specification of the parameter of the model. In general, the specification of some of the parameters of the

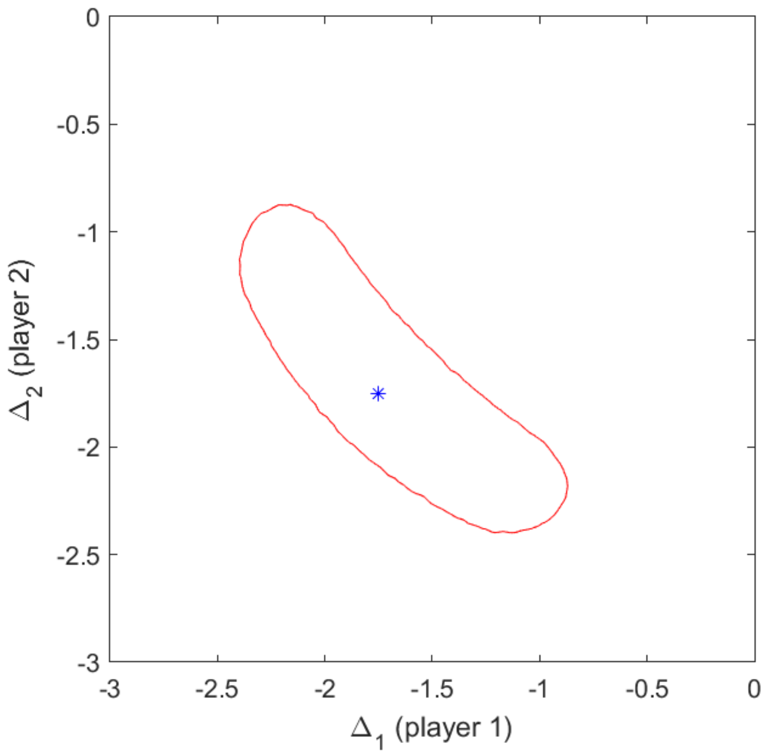

FIGURE 5. Identified set for (Δ_1, Δ_2) without restrictions on the parameters across players

model corresponds to a restricted set of specifications of the rest of the parameters of the model that are simultaneously in the identified set and compatible with that partial specification of the parameter of the model. For example, in this model, with the assumption of equality of parameters across players, and using the distributional assumption on ϵ, any given specification of the β parameters is compatible with a unique specification of the other parameters (e.g., Δ, the correlation of the unobservables, and the selection mechanisms).[36] This means that restrictions on some model parameters directly implies restrictions on other model parameters, given the identified set. In an opposite extreme case where the identified

[36]From the equation for the $(0, 0)$ outcome, it is possible to learn the correlation of the unobservables, the only remaining unknown in that equation. This is because the bivariate normal CDF is strictly monotone in the correlation parameter (e.g., Kotz, Balakrishnan, and Johnson (2000, page 255)). Then, from the equation for the $(1, 1)$ outcome, it is possible to learn Δ, the only remaining unknown in that equation. This is because, evidently, the right hand side of this equation is strictly monotone in Δ. Then given that, it is possible to learn the selection mechanisms $P(y_1 = 1, y_2 = 0|X = x, \epsilon \in S_\beta(x_1, x_2))$ from the equation for the $(1, 0)$ outcome.

set is the Cartesian product of the identified sets for each component of the parameter of the model, restrictions on some model parameters would imply no restrictions on other model parameters.

A characteristic of the partial identification approach is the ease with which additional assumptions can be used. Here, for example, it would be straightforward to introduce assumptions on the selection mechanism, or further assumptions about the parameters. The assumptions on the selection mechanism could be atheoretic *ex ante* bounds on the selection mechanism or theoretical restrictions on the selection mechanism corresponding to the imposition of refinements of Nash equilibrium.

There are two main kinds of identification strategies that establish conditions under which the β and Δ parameters are point identified. Correspondingly, under modest relaxations of those conditions, those parameters can be expected to be partially identified with relatively small identified sets.

The first kind of point identification strategy minimizes the assumptions on the selection mechanism, and looks for other sources of identification. One source of identification is the existence of values of firm-specific observables such that, when firms have those values, the firms have dominant strategies to either enter or not enter the market regardless of the decision of the other firms (e.g., Tamer (2003)). For example, in the above model, as $x_2\beta_2 \to \infty$, firm 2 will enter the market with probability approaching 1, so firm 1 essentially faces a decision problem with a single decision maker in which firm 1 chooses between being a duopolist and not entering the market. Similarly, as $x_2\beta_2 \to -\infty$, firm 2 will enter the market with probability approaching 0, so firm 1 essentially faces a single decision maker problem in which firm 1 chooses between being a monopolist and not entering the market. Therefore, looking at markets with $x_2\beta_2 \gg 0$ and separately with $x_2\beta_2 \ll 0$, firm 1's profit function can be learned using standard identification strategies for models with a single decision maker. If the observables cannot drive the probability of decisions to 1 or 0, there can be partial identification with tail restrictions on the distribution of the unobservables (e.g., Kline (2015)). Another source of identification are assumptions about the unobservables. In particular, the assumption that the density of the unobservables is unimodal, along with an assumption that the observables have enough variation but allowing for less variation than required by the previous identification strategy, is sufficient for point identification (e.g., Kline (2016a)).

The second kind of identification strategy uses further assumptions on the selection mechanism, which can include assumptions that certain variables are excluded from influencing the selection mechanism (e.g., Bajari, Hahn, Hong, and Ridder (2011)) or assumptions that the selection mechanism has a specific known form. Because this effectively reduces the number of unknown parameters in the model, this is a source of identification. These kinds of assumptions can be appealing when economically motivated. Among assumptions on the form of the selection mechanism are assumptions that the selection mechanisms selects among the potential equilibrium outcomes uniformly at random as in Bjorn and Vuong (1984) and Kooreman (1994), or assumptions that the selection mechanism selects the potential equilibrium outcome that is most favorable for a particular player as in Berry (1992) or Jia (2008) or Nishida (2015). Assumptions on the selection mechanism can often be viewed as equivalent to assumptions on the solution concept. For example, the common assumption of a pure strategy equilibrium can be viewed as an assumption on the selection mechanism relative to a model that would allow for mixed strategy equilibria.

This section has focused on the case of a game in which each player takes a single action, but in some games each player takes multiple actions. One such setting is network formation, where each player takes many actions corresponding to linking decisions with each of the other players. This literature has been summarized in de Paula (2017).

4.3.1. *Empirical applications.* The literature has seen many empirical applications of these kinds of models – albeit not always exactly the models discussed here, and not always from a partial identification perspective due to the use of stronger assumptions. When these models are used from a point identification perspective, in many cases that is because of additional assumptions on features of the model, like assumptions on the selection mechanism and/or stronger distributional assumptions on the unobservables. The core model is often nevertheless similar to described above. It is important to keep in mind that a common characteristic of this literature is that each individual paper tends to work with a bespoke model for the specific empirical application, sharing some commonalities reflected by the general models discussed above. The following discussion inevitably glosses over the specific features of the models used in each empirical application.

One common application is to entry behavior in airline markets, as in Reiss and Spiller (1989), Berry (1992), Ciliberto and Tamer (2009), Kline and Tamer (2016), and Kline (2018). Each instance of the game represents an airline market. An individual airline market is defined to be a pair of airports, and the players are typically particular air carriers. In some empirical exercises, because it can be difficult to deal with many players, some air carriers are aggregated according to a shared size. For example, low cost carriers may be modeled to be a single firm. Then, entry decisions concern whether particular air carriers provide regular service between those airports.

The profit functions determine the post-entry profitability of airlines. Typically, profits are modeled to depend both on airline-specific characteristics and market-specific characteristics. A common airline-specific characteristic is the “presence” of a particular airline at the airports in the market, the idea being that an airline that already serves a particular airport (from third airports) is likely to find it more profitable to serve other airports from that airport (e.g., Berry (1992)). This is plausibly excluded from the profit functions of rival airlines. Common market-specific characteristics concern the characteristics of the two areas that the airports serve, for example population. These market-specific characteristics may measure the “size” of the market available to the potential entrants. The interaction effect parameters are key objects of interest, reflecting the impact on profits of the entry of rival firms. For specific empirical results, for example, Berry (1992) finds that airport prescence has an important impact on profitability and Ciliberto and Tamer (2009) find that the policy impact of repealing the Wright Amendment would be to increase the number of markets served from Dallas Love Field Airport.

Another common empirical application is to models of entry of "big box" retailers into geographic markets. Jia (2008) studies decisions of Wal-Mart and Kmart, allowing for both a competitive effect of entry on the profits of rivals and a chain effect of a particular firm having stores in multiple markets. See also Nishida (2015) for allowing competition in the number of stores, in addition to binary entry decisions. Holmes (2011) shows significant benefits to profitability of Wal-Mart from their strategy of a dense network of stores. Ellickson, Houghton, and Timmins (2013) finds important difference between retailers in terms of characteristics like competitive effects and chain effects. Aradillas-López and Rosen (2018) study decisions of Lowe's and Home Depot, focusing on an ordered response decision of how many stores a particular retailer opens in a particular market.

A final set of common applications concerns entry decisions of firms that are smaller than "big box" retailers. Many papers study entry of smaller firms into geographic markets. As such the considerations are often different from the literature on bigger retailers, for example with less emphasis on chain effects. Mazzeo (2002) studies entry of roadside hotels, focusing in particular on the role of the quality of the hotel on the competitive effects on other (possibly different quality) hotels in the same market. Seim (2006) studies the geographic location of entry decisions in the video store industry. Essentially the same basic model can also be used to study network effects, viewing joining the network as "entering." Ackerberg and Gowrisankaran (2006) study adoption decisions of the automated clearinghouse payment system, and the corresponding network effects.

Finally and more recently, Ciliberto, Murry, and Tamer (2021) extended the above discrete choice based models to include two stage games where whether a firm decides to enter or not depends on prices it expects to charge after entry. This combines both a discrete (entry) and a continuous (pricing) component to the model and uses a two stage game with demand conditional on entry estimated in the second stage. They apply this framework to studying pricing and welfare in airline markets. There, the decision to enter and the payoff an airline receives from entry is a function of prices (and quantities) that the firm charges in equilibrium. So, the payoffs in the simple entry game now depends on prices that are "endogenous."

Also, similar models can be used to study empirical applications outside of empirical Industrial Organization, perhaps most notably social interactions in the broad field of applied microeconomics. For example, Soetevent and Kooreman (2007) and Card and Giuliano (2013) use related models to study interactions in the behaviors of teenagers. Bjorn and Vuong (1984) study interactions in the labor force participation decisions of married couples.

Another important set of empirical applications, but with empirical work that is more along the lines of the methods of Section 3, concerns product repositioning, the idea that firms respond to changes in the environment by changing the characteristics of the products they offer. This can be viewed as entry of products, rather than entry of firms. Eizenberg (2014) and Wollmann (2018) propose a two-stage model, where firms decide on product choices in the first stage and prices in the second stage. Revealed preference inequalities similar to those discussed in Section 3 are used to recover partial identification of the fixed costs of offering each product.

4.3.2. *Assumptions on information.* The econometrician must make an assumption about the information structure of the game, which concerns what the players know about the other players. The discussion so far has focused on the case of complete information, where all players in the game commonly know the utility functions of all players in the game. Essentially, this means that the players commonly know X, ϵ, and the parameters of the utility functions. However, from the perspective of the econometrician, X is observed but ϵ is unobserved and the parameters of the utility function are also unknown.

It is possible to vary the assumption about what the players know. Another assumption available to the econometrician is the assumption that the realizations of ϵ are the private information of each player. This is an empirically relevant instance of incomplete information, where players have private information about their utility functions.

By definition, incomplete information is a useful model for applications where players do not know everything about the profit functions of the other players. In particular, incomplete information can be relevant for one-shot interactions against previously unknown players. On the other hand, complete information can be a useful model for applications where players know everything about the profit functions of the other players. Consequently, complete information can be a useful model for the static long-run equilibrium that arises from a repeated interaction.

With incomplete information, players form beliefs about what they do not know given (conditional on) what they do know. When X is common knowledge among the players, it follows that player i gets utility 0 from not entering, as with complete information, and gets expected utility

$$x_i\beta_i + \Delta_i\Pi_i(y_{-i} = 1|X = x, \epsilon_i) + \epsilon_i$$

from entering, where $\Pi_i(y_{-i} = 1|X = x, \epsilon_i)$ is the beliefs that player i holds about the probability that the other player enters the market, given the information available to player i. Given the players use a threshold-crossing strategy,

$$y_i = 1 \Leftrightarrow x_i\beta_i + \Delta_i\Pi_i(y_{-i} = 1|X = x, \epsilon_i) + \epsilon_i \geq 0.$$

Consequently, with the additional assumption that the private information unobservables ϵ are independent across players conditional on the common knowledge X,

$$\Pi_i(y_{-i} = 1|X = x, \epsilon_i) = \Pi_i(\epsilon_{-i} \geq -x_{-i}\beta_{-i} - \Delta_{-i}\Pi_{-i}(y_i = 1|X = x, \epsilon_{-i})|X = x, \epsilon_i)$$

does not depend on ϵ_i, so beliefs can be written $\Pi_i(y_{-i} = 1|X = x) = \Pi_i(y_{-i} = 1|X = x, \epsilon_i)$.

Then, a Bayesian Nash equilibrium is a specification of $\Pi_2(y_1 = 1|X = x)$ and $\Pi_1(y_2 = 1|X = x)$ that is a solution to the system of equations

$$\Pi_2(y_1 = 1|X = x) = P(\epsilon_1 \geq -x_1\beta_1 - \Delta_1\Pi_1(y_2 = 1|X = x)|X = x)$$

$$\Pi_1(y_2 = 1|X = x) = P(\epsilon_2 \geq -x_2\beta_2 - \Delta_2\Pi_2(y_1 = 1|X = x)|X = x).$$

In general, there can be multiple solutions to this system of equations, corresponding to the existence of multiple Bayesian Nash equilibria. Often, it can be assumed that there is a finite number of Bayesian Nash equilibria.

In this literature, it is standard to assume either that there is a unique equilibrium or that a single equilibrium is used in the data generating process (e.g., Bajari, Benkard, and Levin (2007), Aradillas-López (2010) and Bajari, Hong, Krainer, and Nekipelov (2010)), without assuming which equilibrium is used in cases of the existence of multiple equilibria. This assumption can be true even if there are multiple equilibria in the underlying economic theory model. Under this assumption, it follows that $\Pi_{-i}(y_i = 1|X = x) = P(Y_i = 1|X = x)$, so the model can be written

$$P(Y_1 = 1|X = x) = P(\epsilon_1 \geq -x_1\beta_1 - \Delta_1 P(Y_2 = 1|X = x)|X = x)$$
$$P(Y_2 = 1|X = x) = P(\epsilon_2 \geq -x_2\beta_2 - \Delta_2 P(Y_1 = 1|X = x)|X = x).$$

Essentially, these equations reflect two binary choice models, with $P(Y_{-i} = 1|X = x)$ included as an explanatory variable in the equation for the entry decision of player i. Because $P(Y_{-i} = 1|X = x)$ is point identified directly from the observable data, the (point) identification of the parameters (β, Δ) follows from identification strategies for standard binary outcomes models.

Without the assumption that a single equilibrium is used in the data generating process, $P(Y_{-i} = 1|X = x)$ is a mixture of the probabilities that $y_i = 1$ over the multiple equilibria that are compatible with $X = x$. In particular, in general $P(Y_{-i} = 1|X = x)$ cannot be equated to $\Pi_i(y_{-i} = 1|X = x)$ for any particular Bayesian Nash equilibrium. Therefore, if it is allowed that multiple equilibria are used in the data generating process, the previous identification strategy is precluded: it is not possible to apply (point) identification strategies for standard binary outcomes models. However, as described in Aradillas-López (2020), a partial identification strategy that bounds the beliefs in the Bayesian Nash equilbria, and therefore bounds the parameters of the model, is possible.

For any specification of the unknown model parameters θ – which may include the finite-dimensional parameters and the distribution of the unobservables, and for any specification $X = x$, let $\Pi_{i,L,\theta}(y_{-i} = 1|X = x)$ be the smallest belief that player i has that $y_{-i} = 1$. Smallest means among beliefs that arise as part of a Bayesian Nash equilibrium. And let $\Pi_{i,U,\theta}(y_{-i} = 1|X = x)$ be the greatest belief that player i has that $y_{-i} = 1$ that arises as part of a Bayesian Nash equilibrium. These quantities are known by the econometrician since they arise from the underlying economic theory model, for any given θ. Using the threshold-crossing representation of the strategy, and assuming that $\Delta_i \leq 0$, $x_i\beta_i + \Delta_i\Pi_{i,U,\theta}(y_{-i} = 1|X = x) + \epsilon_i \geq 0 \Rightarrow y_i = 1$ because if it is profitable for player i to enter the market even given player i's most pessimistic beliefs that arise in a Bayesian Nash equilibrium, player i will enter the market in any Bayesian Nash equilibrium. Similarly, $y_i = 1 \Rightarrow x_i\beta_i + \Delta_i\Pi_{i,L,\theta}(y_{-i} = 1|X = x) + \epsilon_i \geq 0$ because if player i enters the market in any Bayesian Nash equilibrium, it must be profitable to enter the market given player i's most optimistic beliefs that arise in a Bayesian Nash equilibrium. Similar inequalities could be derived assuming $\Delta_i \geq 0$. Therefore,

$$
\begin{aligned}
P(\epsilon_i \geq -x_i\beta_i - \Delta_i\Pi_{i,U,\theta}(y_{-i} = 1|X = x)|X = x) &\leq P(Y_i = 1|X = x) \\
&\leq P(\epsilon_i \geq -x_i\beta_i - \Delta_i\Pi_{i,L,\theta}(y_{-i} = 1|X = x)|X = x)
\end{aligned}
$$

These inequalities are restrictions on the parameters, based on the distribution of the observed data, and therefore a source of identification. In general, multiple specifications of the parameters are compatible with these restrictions, resulting in partial identification.

Grieco (2014) shows that it is not necessary to choose between a model with complete information and incomplete information, proposing a model with a flexible information structure that nests complete information and incomplete information, and establishes partial identification results. More recently, Magnolfi and Roncoroni (2017) study the question of inference on discrete games with weak assumption on information. This approach is motivated by recent theory work on robust mechanism design (see Bergemann and Morris (2005)) and tries to explore the identified feature of an entry game using correlated equilibria.

4.4. **Models of auctions.** Auctions are instances of incomplete information games that are of particular importance to empirical Industrial Organization. Often, auctions are used to allocate a valuable object (or objects) when there is either heterogeneity and/or uncertainty about the potential buyers' valuations of the object. Correspondingly, identification of auction models is a important literature reviewed for example in Athey and Haile (2007) and Hendricks and Porter (2007).

In a typical auction model with private values, with a single object being auctioned, each of N participants in the auction has a privately known valuation θ_i for the object being auctioned. Valuations for all N bidders, $\theta = (\theta_1, \theta_2, \ldots, \theta_N)$, are drawn from the joint distribution $F(\theta)$. Based on their valuation, each bidder places a bid $b_i(\theta_i)$. Generally, the auction is characterized by functions $x_i(b_i, b_{-i}) \in [0, 1]$ and $t_i(b_i, b_{-i})$. Generally, $x_i(b_i, b_{-i}) \in \{0, 1\}$ indicates whether or not bidder i is allocated the object, given the bids of all bidders, but $x_i(b_i, b_{-i}) \in (0, 1)$ can be allowed to indicate the probability that bidder i is allocated the object, particularly in cases of random allocation when there is a tie for high bid. The $t_i(b_i, b_{-i})$ object indicates the (expected) transfer paid by bidder i, given the bids of all bidders. In most auction types, the high bidder wins and pays some transfer. The utility of bidder i given an allocation $x_i \in \{0, 1\}$ of the object and a transfer payment t_i is $\theta_i x_i - t_i$, so the expected utility of bidder i given its valuation θ_i and based on placing bid b_i is $\theta_i E(x_i(b_i, b_{-i})|\theta_i) - E(t_i(b_i, b_{-i})|\theta_i)$. Bidders place the bid that maximize this expected utility. There are multiple reasons that partial identification can arise in the empirical analysis of auctions.

In an idealized ascending auction, otherwise known as an English auction, based specifically on the button auction model of Milgrom and Weber (1982), and with private values, there is a unique dominant strategy equilibrium: each bidder bids that biddder's valuation. Consequently, the corresponding identification problem is straightforward, since bids directly reveal valuations. However, the button auction may not reflect important features of real-world ascending auctions. In the button auction the price continuously increases, with bidders dropping out of the auction until only one bidder remains. In real world ascending auctions, for example there can be "jump bids" which result is discontinuous increases in the price, and other deviations from the button auction model.

If the real world auction deviates from the button auction then it need not be the case that bids are equal to valuations, resulting in a more complicated identification problem. Moreover, it can be difficult to theoretically model real world auctions that deviate from the button auction. With this motivation, Haile and Tamer (2003) propose two assumptions that relate observed bids to valuations that are credible in ascending auctions even without the button auction model. The first assumption is that bids are weakly less than the corresponding valuation, so $b_i(\theta_i) \leq \theta_i$. It would not make sense for bidder i to bid more than θ_i, because that would open up the possibility that bidder i wins the auction and pays more than its valuation. The second assumption is that a bidder that loses the auction has a valuation that makes it un-profitable to beat the winning bid, so $\theta_i \leq b^* + \Delta$ where b^* is the winning bid and $\Delta \geq 0$ is the minimum bid increment. It would not make sense for bidder i to let another bidder win, when bidder i could win and make a profit. These assumptions are satisfied in the button auction model, but do not require the button auction model, and indeed do not require any particular model of the auction. The first assumption generates an upper bound on the distribution of valuations, and the second assumption generates a lower bound on the distribution of valuations. If the true data generating process is the button auction model, the lower bound equals the upper bound, so the identification result turns into a point identification result. In an empirical application to U.S. Forest Service auctions, Haile and Tamer (2003) find that the identified set for the distribution of valuations is tight, while finding somewhat wider bounds on the optimal reserve price. Chesher and Rosen (2017) use generalized instrumental variables to derive the sharp identified set.

An important object of interest in auction models is the reserve price. Optimal auction theory can be used to determine the optimal reserve price that results in the maximum revenue to the auctioneer, when the distribution of valuations is known. If the distribution of valuations is known only to be within a set of possible values, as with partial identification, Haile and Tamer (2003) show the optimal reserve price can correspondingly be partially identified to be within a set of possible values. Given that the optimal reserve price is only partially identified, Aryal and Kim (2013) propose a decision-theoretic rule for choosing the reserve price to use. In a related setup, Song (2014) propose a decision-theoretic rule for point decisions from an interval identified set, applying in particular to the choice of the reserve price. A closely related object of interest is the auction revenue under different auction formats and different reserve prices. Tang (2011) establishes partial identification of the revenue from alternative auctions formats with optimal reserve prices, when the observed data comes from a first-price auction, allowing for both private values and common values based on an affiliated signals and valuations model.

Unobserved heterogeneity is often a concern in empirical analysis of auctions, see e.g. Haile and Kitamura (2019). With unobserved heterogeneity, bidder valuations are drawn from a distribution conditional on the unobserved heterogeneity that is known to the bidders but not known to the econometrician. Hence, with unobserved heterogeneity, different auctions have valuations drawn from different distributions. Under suitable assumptions, sometimes more than directly predicted by economic theory, it is possible to point identify objects of interest even with unobserved heterogeneity, as summarized in Haile and Kitamura (2019). Under weaker assumptions, Armstrong (2013) shows that it is possible to partially identify objects of interest including the expected value of bidder valuations, the expected value of the highest valuation, and expected profits of a bidder, in a sealed bid first price auction with unobserved heterogeneity.

Endogenous participation of bidders, otherwise known as entry into auctions, is also often a concern in empirical analysis of auctions. Based on a model of private values, where each of the potential bidders observes a signal of their value before choosing whether to pay an entry cost, Gentry and Li (2014) establish partial identification of the distribution of valuations based on variation in entry. With enough entry variation, the result is point identification. Focusing on the general issue of entry, Gentry and Li (2014) works for many standard auction formats.

The (lack of) observability of all bids is yet another important consideration in the empirical analysis of auctions. For example, in general it can be difficult/impossible to observe the highest "bid" in an ascending auction in situations where that "bid" is effectively never placed given that the auction concludes when all of the other bidders have dropped out of the auction. In second price and ascending auctions, Komarova (2013) considers partial identification of the distribution of valuations under different conditions on the observed data, without assumptions that valuations are independent across bidders. In ascending auctions with correlated values, Aradillas-López, Gandhi, and Quint (2013) considers partial identification directly of the objects of interest, specifically bidder and seller profits, sidestepping the identification of the distribution of bidder valuations, using variation in the number of bidders in different auctions, assuming only the transaction price and number of bidders are observed. Coey, Larsen, Sweeney, and Waisman (2017) shows the same bounds are valid when the distribution of valuations are allowed to be asymmetric when bidder identities are unobserved, and propose related, tighter bounds when bidder identities are observed.

Auctions involving the allocation of multiple units are yet another important consideration in the empirical analysis of auctions, see e.g. Hortaçsu and McAdams (2018). In these auctions, bidders are observed to place the same bid for multiple units, either as an equilibrium outcome or as a restriction of the auction format, even if they have different valuations for the multiple units. McAdams (2008), Hortaçsu and McAdams (2010), and Kastl (2011) show how this can result in partial identification results.

Similar to the use of weaker assumptions on the solution concept in models of games with discrete action spaces, it is possible to use weaker assumptions on the solution concept in auction models. Based on bidders participating in multiple independent auctions, Gillen (2009) and An (2017) establishes partial identification and point identification results in a first-price auction where beliefs follow from "level-k" thinking. See also Aradillas-López and Tamer (2008). An (2017) shows this model is observationally equivalent to an asymmetric distribution of valuations.

Finally, there is a developing literature on identification with weak assumption on information motivated by the work of Bergemann and Morris (2005). See also Magnolfi and Roncoroni (2017) in the context of discrete games. Information, or what players know and the signals they receive, plays a key role in strategic setups where in a Bayesian Nash equilibrium this information is used to derive the equilibrium mapping. In empirical work, we rarely have access to the information players use and so it is natural to ask what can be learned about the fundamentals (such as utilities and payoffs) under weak assumptions on information. In an auction setup, Syrgkanis, Tamer, and Ziani (2017) study this question and use the results between Bayesian correlated equilibria and Bayesian Nash equilibrium. The statistical setup is a high dimensional linear program where it is possible to use the distribution of observed bids to recover features of the distribution of valuations. In addition, the same linear programming structure can be used to then simulate policy questions. The linear program is a moment inequality problem where the linear structure can be used in a computational simple way.

4.5. **Alternative assumptions.** Finally, there have been attempts to use other frameworks, and this subsection refers to some relevant parts of that literature. Perhaps the best known is the program evaluation literature. In settings with multiple decision makers, the standard assumptions and approaches of the program evaluation literature generally do not apply. This is because the outcome of any given decision maker generally depends on the decisions (or other characteristics) of the other decision makers. In short, the representation of a counterfactual outcome as depending only on the same decision maker's treatment is generally not valid in models with multiple decision makers.[37] Indeed the interaction between decision makers can generate multiple responses from a given set of exogenous treatments, so there can be multiple potential outcomes in response to a specified set of exogenous treatments, depending on which equilibrium is selected (for a discussion in nonparametric versions of simple games, see Kline and Tamer (2020)). Also, in cases with mixed strategies (or where

[37]More formally, such setups generally violate the Stable Unit Treatment Value Assumption as summarized by Imbens and Rubin (2015, Section 1.6).

equilibria are in terms of distributions over actions), the link between the observed data and the underlying outcomes is complicated.

It is possible to augment the standard treatment response models to allow for interactions, as in Manski (2013) and Lazzati (2015). That is, the response function for i can be augmented to depend on the treatments assigned to $j \neq i$, and further structure could be imposed. If that structure mimicked the relevant market and a maximizing assumption, or some other model of how behavior responds to incentives, one would come back to a framework resembling the one we propose or one of the extensions discussed elsewhere.

An alternative that is closer to the Industrial Organization theory that much of empirical Industrial Organization relies on, are solution concepts that replace the description of agents' behavior with weaker restrictions. This includes the use of the concepts of rationalizability (Bernheim (1984) and Pearce (1984)) and of iteratively dominated strategies (see the discussion in Tan and da Costa Werlang (1988)). Using weaker assumptions about the solution concept necessarily results in less ability to learn about objects of interest. On the other hand it may allow us to get a better approximation to agent's behavior. This is an avenue worth pursuing in the context of actual applied problems.[38]

Similarly weakening the parametric assumptions on the return function could help us to better approximate behavior and/or focus on a key parameter that determines market outcomes. For example Kline and Tamer (2012) shows that it is sometimes possible to partially identify the best response functions without assuming a particular form of the utility function, nor distributional assumptions on the unobservables. The identification comes exclusively from assumptions from economic theory and the data, rather than from functional form assumptions. Relatedly de Paula and Tang (2012) study whether it is possible to identify the sign of the interaction effects between firm responses in an incomplete information game. The sign is particularly important because it underlies the determination of whether policies

[38]The experimental literature has also proposed many candidates, including models of limited strategic reasoning including level-k thinking models and cognitive hierarchy models (e.g., Costa-Gomes, Crawford, and Broseta (2001), Camerer, Ho, and Chong (2004), Costa-Gomes and Crawford (2006), Crawford and Iriberri (2007)). Much of this appears in recent books and reviews, including Camerer (2011) and Crawford, Costa-Gomes, and Iriberri (2013)

are strategic substitutes or strategic complements, which can be of fundamental importance in evaluating possible policy changes. In other applications, like models of social interactions, the sign of interaction effects determines whether there are positive peer effects (a peer taking the behavior increases the utility of taking the behavior) or negative peer effects (a peer taking the behavior decreases the utility of taking the behavior). A key feature of their identification strategy is the fact negative interaction effects result in negatively correlated responses within markets, and positive interaction effects results in positively correlated responses within markets.

5. Estimation and inference

Standard approaches to estimation and inference are based on the assumption of point identification. Estimation and inference in partially identified models is fundamentally different, because the objects of interest are set-valued rather than singleton-valued. This issue was recognized even before partial identification became popular, in particular in Phillips (1989). As discussed in the previous sections, identification in models used in empirical Industrial Organization presents some unique issues, compared to identification in models used in other fields of economics. However, estimation and inference can basically follow the general literature on partially identified models, which has been summarized most recently, for example, in Canay and Shaikh (2017) and Molinari (2020). Following the literature, the discussion focuses on frequentist approaches in Sections 5.1-5.5, with a separate discussion of Bayesian approaches in Section 5.6.

Estimation and inference in partially identified models requires a measure of distance between two sets, namely the true identified set and an estimate of the identified set. The literature has focused on the Hausdorff distance. The Hausdorff distance between sets A and B is $d_H(A, B) = \max\{\sup_{a\in A} \inf_{b\in B} d(a, b), \sup_{b\in B} \inf_{a\in A} d(a, b)\}$, where $d(\cdot, \cdot)$ is a distance defined on the elements of the parameter space. The Hausdorff distance $d_H(A, B)$ is small when two conditions hold: every element $a \in A$ is "close" to at least one element of B

and every element $b \in B$ is "close" to at least one element of A. In some instances, the representation $d_H(A,B) = \sup_{c \in A \cup B} |\inf_{a \in A} d(c,a) - \inf_{b \in B} d(c,b)|$ is convenient.

5.1. **Estimation.** As a typical starting point, estimation of partially identified models involves replacing population quantities in the representation of the identified set with corresponding sample quantities. This is the analogy principle (e.g., Goldberger (1968), Manski (1988a)) familiar in point identified models, as applied to partially identified models. The first question is whether such an estimator is consistent. In general, consistency of an estimator concerns the distance between the estimator and true value. Based on the Hausdorff distance, an estimator $\hat{\Theta}_{I,N}$ of the identified set is consistent if $d_H(\hat{\Theta}_{I,N}, \Theta_I) \to^p 0$, which means the Hausdorff distance between the estimate and the identified set converges in probability to 0 as sample size increases. Although it is tempting to think otherwise, consistency of the sample analogue estimators of the population quantities in the representation of the identified set does not necessarily imply consistency of the sample analogue estimator $\hat{\Theta}_{I,N}$.

For an illustration of some of the main technical ideas surrounding estimation in partially identified models, and for simplicity of exposition, consider the problem of estimating a scalar component of a partially identified parameter. If Θ_I is convex (or connected), then the identified set for any scalar component δ of θ is an interval. The identified set for δ can be an interval even if Θ_I is not convex. Suppose further that the identified set for δ can be represented in terms of population expected values as in $\Delta_I = [E(Y_L), E(Y_U)]$, where Y_L and Y_U are (known functions of) random variables in the observed data set. Unless the model is misspecified, $E(Y_L) \leq E(Y_U)$, because otherwise $\Delta_I = \emptyset$, implying there would be no value of δ that is compatible with the assumptions and the observed data. In fact, checking whether the identified set is empty is one proposed method for checking for correct specification in partially identified models.

Then $\hat{\Delta}_{I,N} = [E_N(Y_L), E_N(Y_U)]$ is the analogy principle estimator of Δ_I that replaces population quantities in the representation of Δ_I with the corresponding sample quantities.

In particular, $E_N(\cdot)$ is the sample average from a sample of size N. Suppose that $E_N(Y_L) \to^p E(Y_L)$ and $E_N(Y_U) \to^p E(Y_U)$. Using the second representation of the Hausdorff distance,

$$d_H(\Delta_I, \hat{\Delta}_{I,N}) = \begin{cases} \max\{|E(Y_L) - E_N(Y_L)|, |E(Y_U) - E_N(Y_U)|\} & \text{if } E_N(Y_L) \leq E_N(Y_U) \\ \text{undefined} & \text{if } E_N(Y_L) > E_N(Y_U) \end{cases}.$$

$d_H(\Delta_I, \hat{\Delta}_{I,N})$ is undefined when $E_N(Y_L) > E_N(Y_U)$ because in that case $\hat{\Delta}_{I,N} = \emptyset$. Some sources would define $d_H(A, \emptyset) = \infty$ when $A \neq \emptyset$, but that difference in definition does not impact the analysis.

With this setup, it is possible to study whether $\hat{\Delta}_{I,N}$ is a consistent estimator. There are two cases to consider: $E(Y_L) < E(Y_U)$ and $E(Y_L) = E(Y_U)$. If $E(Y_L) < E(Y_U)$, then $P(E_N(Y_L) \leq E_N(Y_U)) \to 1$, so the first line of the expression for $d_H(\Delta_I, \hat{\Delta}_{I,N})$ is the relevant case for asymptotic analysis, and therefore $d_H(\Delta_I, \hat{\Delta}_{I,N}) \to^p 0$ because $E_N(Y_L) \to^p E(Y_L)$ and $E_N(Y_U) \to^p E(Y_U)$. If $E(Y_L) = E(Y_U)$, then in general[39] $P(E_N(Y_L) \leq E_N(Y_U)) \not\to 1$, because $P(E_N(Y_L) \leq E_N(Y_U)) \equiv P(\sqrt{N}(E_N(Y_L) - E_N(Y_U)) \leq 0) \approx \Phi(0) = \frac{1}{2}$ when a central limit theorem applies with *positive* asymptotic variance,[40] in which case even in large samples there is non-vanishing probability that $\hat{\Delta}_{I,N} = \emptyset$ and therefore $d_H(\Delta_I, \hat{\Delta}_{I,N}) \not\to^p 0$ since there is non-vanishing probability that $d_H(\Delta_I, \hat{\Delta}_{I,N})$ is undefined.

In words, if Δ_I is a *non-degenerate* interval, then $\hat{\Delta}_{I,N}$ is consistent. However, if Δ_I is a singleton, equivalent to δ being point identified, then $\hat{\Delta}_{I,N}$ is not consistent in general. It may be surprising that $\hat{\Delta}_{I,N} = [E_N(Y_L), E_N(Y_U)]$ is not a consistent estimator of $\Delta_I = [E(Y_L), E(Y_U)]$ when $E(Y_L) = E(Y_U)$, considering the endpoints of $\hat{\Delta}_{I,N}$ are consistent estimators of the endpoints of Δ_I when $E_N(Y_L) \to^p E(Y_L)$ and $E_N(Y_U) \to^p E(Y_U)$. This does not imply consistency of $\hat{\Delta}_{I,N}$ because $[a, b] = \emptyset$ when $a > b$, so even if $E_N(Y_L) \approx E(Y_L) = E(Y_U) \approx E_N(Y_U)$ it may be that $E_N(Y_L) > E_N(Y_U)$ and therefore $\hat{\Delta}_{I,N} = \emptyset$ while

[39]Note that even if $E(Y_L) = E(Y_U)$, it is still possible that $P(E_N(Y_L) \leq E_N(Y_U)) \to 1$, specifically when $E_N(Y_L) - E_N(Y_U)$ has *zero* asymptotic variance, for example because the underlying realizations of Y_L and Y_U are equal. This does arise in some relevant cases, for example with interval censored data, with the $Y_L = Y_U$ case being the case where each "interval" is actually a singleton.

[40]Note that the standard "scaling" by the asymptotic variance would not impact the result.

$\Delta_I = \{E(Y_L)\} = \{E(Y_U)\}$. Although illustrated in this particular case of an interval identified set, it is important to keep in mind that in general similar issues arise when working with partially identified models, where the "geometry" of the identified set plays a role in the properties of the estimator. For similar reasons, inference in partially identified models is complicated in particular in these cases.

Obviously, the case that Δ_I is a singleton is an important special case. Practically, this analysis suggests that $\hat{\Delta}_{I,N}$ may perform poorly in finite samples when $E(Y_L) \approx E(Y_U)$, such that Δ_I is a "small" interval, since in that case there is non-negligible finite sample probability that $E_N(Y_L) > E_N(Y_U)$ and hence $\hat{\Delta}_{I,N} = \emptyset$.

Based on such considerations, an alternative estimator of the identified set is $\hat{\Delta}^{\epsilon}_{I,N} = [E_N(Y_L) - \epsilon_N, E_N(Y_U) + \epsilon_N]$ where $\epsilon_N > 0$ is a sequence that satisfies $\epsilon_N \to 0$. Compared to $\hat{\Delta}_{I,N}$, $\hat{\Delta}^{\epsilon}_{I,N}$ is consistent under more general conditions. However, $\hat{\Delta}^{\epsilon}_{I,N}$ introduces the tuning parameter ϵ_N that directly impacts the estimation results. For $\hat{\Delta}^{\epsilon}_{I,N}$,

$$d_H(\Delta_I, \hat{\Delta}^{\epsilon}_{I,N}) = \begin{cases} \max\{|E(Y_L) - E_N(Y_L) + \epsilon_N|, |E(Y_U) - E_N(Y_U) - \epsilon_N|\} & \text{if } E_N(Y_L) \leq E_N(Y_U) + 2\epsilon_N \\ \text{undefined} & \text{if } E_N(Y_L) > E_N(Y_U) + 2\epsilon_N \end{cases}.$$

If $\epsilon_N\sqrt{N} \to \infty$, then $P(E_N(Y_L) > E_N(Y_U) + 2\epsilon_N) \to 0.$[41] Therefore, as long as $\epsilon_N \to 0$ sufficiently slowly, $d_H(\Delta_I, \hat{\Delta}^{\epsilon}_{I,N}) \to^p 0$ even when Δ_I is a singleton.

Similar ideas of "expanding" the estimate of the identified set apply more generally, with more complicated analyses used in other settings. In this particular setup, the rate of ϵ_N can be directly tied to the rate of convergence of the sample averages. In settings with "large" identified sets, which in this setup means $E(Y_L)$ is not "close" to $E(Y_U)$ and often is the case with partially identified models, these considerations about "expanding" the estimate of the identified set are less relevant. The need to "expand" the identified set can also arise in misspecified models in which the true identified set is empty, and the econometrician aims

[41] If $E(Y_L) - E(Y_U) = a < 0$, then $P(E_N(Y_L) - E_N(Y_U) \leq 0) \geq P(|E_N(Y_L) - E_N(Y_U) - a| < -a) \to 1$, implying $P(E_N(Y_L) - E_N(Y_U) > 2\epsilon_N) \to 0$. If $E(Y_L) = E(Y_U)$ and a central limit theorem applies to $\sqrt{N}(E_N(Y_L) - E_N(Y_U))$, then $P(\sqrt{N}(E_N(Y_L) - E_N(Y_U)) \geq 2\epsilon_N\sqrt{N}) \to 0$.

to report a "pseudo-true" identified set.[42] Another possibility notes that $E_N(Y_L) > E_N(Y_U)$ happens with non-negligible probability in large samples only when the parameter is point identified with $\delta = E(Y_L) = E(Y_U)$, which implies corresponding estimators of δ. Unlike the previous estimators which can be the empty set, this results in an estimate of the identified set that is non-empty regardless of the degree of misspecification of the model, which may or may not be a desirable property.

5.2. **Overview of inference.** Inference in partially identified models tends to present additional challenges. First, there is a difference between inference on the identified set and inference on the *elements* of the identified set. Such a consideration does not arise in point identified models, when the identified set is a singleton.

One inference target is a confidence set for the identified set. A confidence set $\mathcal{C}_{N,\alpha}$ for the identified set Θ_I is defined to be a set-valued function of the data, that at least satisfies the condition that $\liminf_{N\to\infty} P(\Theta_I \subseteq \mathcal{C}_{N,\alpha}) \geq 1-\alpha$ for specified $\alpha \in (0,1)$. With this definition, $\mathcal{C}_{N,\alpha}$ contains the identified set Θ_I with at least repeated sampling probability $1-\alpha$ in large samples. This definition allows the confidence set to be conservative, since the confidence set is allowed to contain the identified set with probability greater than the nominal level $1-\alpha$. Sometimes it is possible to establish a confidence set satisfies the condition that $\lim_{N\to\infty} P(\Theta_I \subseteq \mathcal{C}_{N,\alpha}) = 1-\alpha$. With this condition, $\mathcal{C}_{N,\alpha}$ is an asymptotically valid confidence set for the identified set, and not conservative.

Another inference target is a confidence set for the *elements* of the identified set. A confidence set $\mathcal{C}_{N,\alpha}$ for the *elements* of the identified set Θ_I at least satisfies the condition that $\liminf_{N\to\infty} P(\theta' \in \mathcal{C}_{N,\alpha}) \geq 1-\alpha$ for any $\theta' \in \Theta_I$. With this definition, $\mathcal{C}_{N,\alpha}$ contains any given element of the identified set Θ_I with at least repeated sampling probability $1-\alpha$ in large samples. This definition allows the confidence set to be conservative. Sometimes it is possible

[42]Empirical work can result in empty identified sets, particularly when there are many inequalities or restrictions on the parameters, because models are generally mis-specified. Still, policy decisions are going to be made, and the question is whether the audience can trust a particular estimate more than the alternative estimates that are available.

to establish a confidence set satisfies the further condition that $\lim_{N\to\infty} P(\theta' \in \mathcal{C}_{N,\alpha}) = 1-\alpha$ at least for some θ'. With this condition, $\mathcal{C}_{N,\alpha}$ is an asymptotically valid confidence set for the elements of the identified set, and not conservative for those θ'.

A confidence set for Θ_I treats Θ_I per se as the object of interest. However, there can be situations where the "true value" of the parameter θ_0 is the object of interest. This suggests constructing a confidence set for θ_0 rather than Θ_I. Of course, the true value necessarily satisfies $\theta_0 \in \Theta_I$, so a confidence set for the elements of the identified set is guaranteed to contain the "true value" θ_0 with at least probability $1-\alpha$ in repeated large samples. There are also cases where a confidence set for the identified set is identical to a confidence set for the true parameter.[43] However and more generally, a confidence set for the elements of the identified set is not guaranteed to contain *all* values of the parameter that are compatible with the assumptions and the observed data. Essentially, a confidence set for the elements of the identified set corresponds to "testing" individual specifications of θ for compatibility with the assumptions and the observed data. By definition, a confidence set for the identified set is also a confidence set for the elements of the identified set, but a confidence set for the elements of the identified set is not necessarily a confidence set for the identified set. Therefore, a confidence set for the identified set tends to be larger than a confidence set for the elements of the identified set. Henry and Onatski (2012) provides a robust control argument for preferring inference for the identified set.

In partially identified models, uniform validity of inference is important. The previous conditions that characterized confidence sets concerned pointwise validity, where "pointwise" means (implicitly) assuming a single fixed data generating process. Uniform validity of confidence sets requires conditions that hold uniformly across a set of data generating processes. A confidence set for the identified set is uniformly valid over $\mathcal{P}$ if it satisfies the condition that $\liminf_{N\to\infty} \inf_{P\in\mathcal{P}} P(\Theta_I \subseteq \mathcal{C}_{N,\alpha}) \geq 1-\alpha$ where $\mathcal{P}$ is a space of possible data generating processes. This means that, for sufficiently large sample size, and then

[43] In likelihood settings with density p_θ, any θ that belongs to the identified set is defined as one where $p_\theta = p_0$, the true data density. And so, for any such θ the LR statistic just depends on p_0 (and of course an estimator $\hat{p}$ of p_0).

for any data generating process in $\mathcal{P}$, the confidence set contains the identified set with repeated probability approximately at least $1-\alpha$. Pointwise validity can be written with quantifiers reversed: $\inf_{P \in \mathcal{P}} \liminf_{N \to \infty} P(\Theta_I \subseteq \mathcal{C}_{N,\alpha}) \geq 1-\alpha$. With pointwise validity, the minimum sample size N_ϵ such that $P(\Theta_I \subseteq \mathcal{C}_{N,\alpha}) \geq 1-\alpha-\epsilon$ for $N \geq N_\epsilon$ can depend on the data generating process. With uniform validity, the minimum sample size N_ϵ such that $P(\Theta_I \subseteq \mathcal{C}_{N,\alpha}) \geq 1-\alpha-\epsilon$ for $N \geq N_\epsilon$ *cannot* depend on the data generating process. In words, pointwise validity says that for any given data generating process, there is a large enough sample size such that the confidence set contains the identified set with at least probability $1-\alpha$. Possibly the sample size depends on the data generating process. A stronger condition is that the same sample size is large enough such the confidence set contains the identified set with at least probability $1-\alpha$, for any data generating process in $\mathcal{P}$.

If a confidence set is not uniformly valid over $\mathcal{P}$, then for any sample size, there is a data generating process in $\mathcal{P}$ such that the confidence set does not contain the identified set with at least probability approximately $1-\alpha$. Such a confidence set could still be pointwise valid if, given any data generating process in $\mathcal{P}$, there is a large enough sample size such that the confidence set contains the identified set with at least probability $1-\alpha$. This distinction is de-emphasized in point identified models, because in many point identified models the distinction between pointwise validity and uniform validity is negligible. See also Canay and Shaikh (2017) for more on this point. Essentially, uniformity tends to fail, even if the confidence set is pointwise valid, when for any sample size, it is possible to find a data generating process in $\mathcal{P}$ such that the asymptotic approximation is not a good approximation for that sample size. In particular, in moment inequality condition models, the number of moment inequality conditions that hold as equalities is relevant for the asymptotic distribution in many approaches to inference, and this depends discontinuously on the data generating process, resulting in potential failures of uniformly valid inference. Uniform validity of a confidence set depends on the specific construction of the confidence set, in addition to the model in which the confidence set is used.

The chapter discusses some approaches to inference in partially identified models. We will start with the moment inequality approach, the criterion function approach, and the random set approach. In many cases, these different approaches are just different perspectives on the same underlying statistical problem. For example, the moment inequality approach is a special case of the criterion function approach. The chapter also discusses the Bayesian alternative to frequentist inference.

5.3. **Moment inequality approach.** Some empirical models imply that the true value of the parameter θ_0 satisfies restrictions of the form $E(m(W,\theta_0)) \geq 0$, where $m(\cdot) = (m_1(\cdot), m_2(\cdot), \ldots, m_J(\cdot))$ is a vector-valued function that is known by the econometrician. Such restrictions are known as unconditional moment inequality conditions. See for example Section 3. Correspondingly, the identified set is $\Theta_I = \{\theta : E(m(W,\theta)) \geq 0\}$. Other empirical models imply that θ_0 satisfies restrictions of the form $E(m(W,\theta_0)|X = x) \geq 0$ for all x. Such restrictions are known as conditional moment inequality conditions. Correspondingly, the identified set is $\Theta_I = \{\theta : E(m(W,\theta)|X = x) \geq 0 \text{ for all } x\}$. Any conditional moment inequality condition implies the infinite set of unconditional moment inequality conditions of the form $E(m(W,\theta_0)g(X)) \geq 0$ for any non-negative function $g(\cdot)$. A similar approach is taken, for example, in the textbook analysis of a linear model with point identification when converting the conditional moment conditions of the form $E(Y - X\beta_0|X = x) = 0$ that arises from the assumption that $Y = X\beta_0 + \epsilon$ and $E(\epsilon|X = x) = 0$ into unconditional moment conditions like $E(X'(Y - X\beta_0)) = 0$. Sometimes the $g(\cdot)$ is known as an instrumental function. Much of the econometric theory literature on models with conditional moment inequalities essentially converts the model into a model with unconditional moment inequalities. Although moment inequality conditions are sometimes viewed as equivalent to partial identification, there can be partially identification results that do not result in moment inequality conditions in a natural way.

The combination of two moment inequality conditions $E(m_e(W,\theta_0)) \geq 0$ and $E(-m_e(W,\theta_0)) \geq 0$ results in the moment equality condition $E(m_e(W,\theta_0)) = 0$. Therefore, moment equality conditions are a special case of moment inequality conditions. However, including moment equality conditions as the combination of two moment inequality conditions often requires particular care in the theoretical analysis, because it means that two elements of the vector $m(W, \theta)$ are perfectly negatively correlated. Some inference approaches explicitly allow for moment inequality conditions and moment equality conditions, treated separately. If moment equality conditions are to be used alongside moment inequality conditions, it is important to verify that the inference approach adequately accommodates this.

Generally, estimation of a model based on moment inequality conditions is straightforward. The estimate of the identified set is the set of θ that satisfy the sample analogue of the moment inequality conditions, possibly with the "expansion" discussed above. However, in many cases such an estimator is biased in small samples towards finding a too-small identified set, so it may be desirable to bias-correct the sample-analogue estimator, as discussed for example in Haile and Tamer (2003), Kreider and Pepper (2007), Andrews and Shi (2013), and Chernozhukov, Lee, and Rosen (2013). Kaido and Santos (2014) study efficiency bounds when the moment inequalities and thus identified set is convex, finds the plug-in estimator is consistent, and proposes a valid bootstrap. Bootstrap validity has been explored more generally in moment inequality models, as discussed below.

Inference in moment inequality conditions is not straightforward. In unconditional moment inequality condition models, the hypothesis that $\theta' \in \Theta_I$ for a particular candidate value of the parameter θ' is equivalent to the hypothesis that $E(m(W, \theta')) \geq 0$. Essentially, this amounts to testing non-negativity of a particular finite-dimensional population mean. In conditional moment inequality models, the hypothesis that $\theta' \in \Theta_I$ for a particular candidate value of the parameter θ' is equivalent to the hypothesis that $E(m(W, \theta')g(X)) \geq 0$ for any non-negative function $g(\cdot)$. Essentially, this amounts to testing non-negativity of a particular infinite-dimensional population mean. Consequently, inference in moment inequality models involves many of the problems that arise in testing for non-negativity of a population mean. Finally, in some instances, the moment inequalities can only be based on zero correlation between an unobservable and a covariate, in which case by construction there is a finite set of unconditional moment inequalities.

One of the central difficulties is that the asymptotic distributions of the proposed test statistics tend to be not pivotal, meaning that they depend on unknown features of the data generating process. Before turning to the econometric theory issues, one particular implication relevant for empirical research is that the critical value must be determined for each candidate value of the parameter, a potentially computationally expensive step. Generally, this problem becomes more serious with increasing numbers of moment inequality conditions. Related computational problems tend to arise when the dimension of Θ is large and when using a two-step estimator, with a first step that does something like estimating Nash prices that go into the calculation of the profitability of different actions.

The asymptotic distributions of the proposed test statistics tend to depend on which of the moment inequality conditions hold as equalities in the population, in the sense that $E(m_j(W, \theta')) = 0$ for certain $j \in \{1, 2, \ldots, J\}$. In particular, this means that the asymptotic distributions depend *discontinuously* on this unknown feature of the data generating process. The moment inequality conditions that hold as strict inequalities tend to not influence the behavior of the test statistic, essentially because the test statistics measure the violation of the restrictions that $E(m(W, \theta')) \geq 0$, and if a particular moment inequality condition indeed holds strictly as $E(m_j(W, \theta')) > 0$, then even with only modest sample sizes it will be evident with near certainty that moment inequality condition is not violated. Of course, the econometrician does not *ex ante* know whether $E(m_j(W, \theta')) > 0$ or $E(m_j(W, \theta')) = 0$ for any particular θ' that satisfies all of the moment inequality conditions, and therefore the asymptotic distributions of the proposed test statistics depend on unknown features of the data generating process. This contrasts with textbook moment equality condition models á la Hansen (1982), where the econometrician presumes that all moment equality conditions hold as equalities by construction for the true value of the parameter.

One possible response to this problem is to find the "least favorable" asymptotic distribution which results in the largest critical value across all relevant values of the unknown parameters, an approach taken for instance in Rosen (2008). Generally, the "least favorable" asymptotic distribution corresponds to assuming, for purposes of calculating the critical value, that all of the moment inequality conditions hold as equalities. This approach can result in critical values that are "too large" resulting in conservative inference such that the resulting confidence sets are "too large." However, an important advantage is that this approach does not require determining a different critical value for each value of the parameter. Another possible response is to try to "estimate" which of the moment inequality conditions hold as equalities, and impose that on the determination of the critical value, as in the generalized moment selection approach of Andrews and Soares (2010). Andrews and Barwick (2012) compare many of the proposed methods from the literature, and propose a recommended approach. By taking a particular "two-step" approach that determines which moment inequality conditions hold as strict inequalities based on a particular confidence interval for the moments relevant for the moment inequality conditions, Romano, Shaikh, and Wolf (2014) are able to achieve computational speed gains. Menzel (2014) and Chernozhukov, Chetverikov, and Kato (2019) consider the problem of many moment inequality conditions, a common situation in empirical practice where the model can imply more moment inequality conditions than there are observations in the observed data to estimate the moment inequality conditions, including possibly arising from conversion from unconditional moment inequality conditions.

The earlier results applied to models with unconditional moment inequality conditions, but some models involve conditional moment inequality conditions. A set of conditional moment inequalities can be converted into an infinite number of unconditional moment inequalities, as done in Andrews and Shi (2013). The methods of Andrews and Shi (2013) have been implemented in Stata as the `cmi_test` command in Andrews, Kim, and Shi (2017). Andrews and Shi (2014) further allow for the possibility of a nonparametric or semiparametric parameter, allowed to be different for different values of a conditioning variable. The choice of test statistic is particularly important in moment inequality condition models. Armstrong (2014, 2015) show some advantages of a particular Kolmogorov-Smirnov statistic. Aradillas-López, Gandhi, and Quint (2016) and Lee, Song, and Whang (2018) are inference methods based on a one-sided L_p test. Chernozhukov, Lee, and Rosen (2013) study "intersection bounds" which includes conditional moment inequalities, using a precision-corrected sup/inf estimator of the bounds. The methods of Chernozhukov, Lee, and Rosen (2013) have been implemented in Stata as the `cmi_test` command in Chernozhukov, Kim, Lee, and Rosen (2015).

Uniform validity of inference is an important problem in models with moment inequality conditions, following Imbens and Manski (2004). Uniform validity of inference concerns whether tests reject true null hypotheses at (no more than) the nominal rate (in large samples) even for the least favorable data generating process. The failure of uniformly valid inference would imply that, for any sample size, there is a data generating process such that the test rejects true null hypotheses at above the nominal rate, resulting in confidence sets that are "too small." Practically, failure of uniformly valid inference in models with moment inequality conditions tends to happen when the parameter is actually point identified or has smaller identified set such that the parameter is "almost" point identified. In particular, for any fixed sample size, Imbens and Manski (2004) show a confidence interval with nominal 95% coverage that indeed has (approximately) 95% coverage when the parameter is partially identified but only 90% coverage when the parameter is point identified. This is a failure of uniformity since the coverage is below the nominal rate when there is point identification. Therefore, much of the more recent econometric theory literature takes special care to establish uniformly valid inference.

Although it is common empirical practice to use a bootstrap or other resampling method for inference, particularly in settings where other inference methods are not easily available to the empirical researcher, it is important to note that standard bootstrap methods are not valid in moment inequality methods. This is related to the invalidity of the bootstrap in other settings with inequality constraints, as discussed in Andrews (2000) and Horowitz (2019). However, various modified bootstrap methods are known to be valid, as in Bugni (2010) and Canay (2010).

Inference on functions of the parameters, or components of the parameters, in partially identified models introduces additional challenges. In point identified models with an asymptotically normal estimator, this can be accomplished based on using the delta method, or just the corresponding component of the asymptotic distribution, respectively. Without the possibility of such an "asymptotically normal" estimator in a partially identified model, that approach is not possible in a partially identified model. Nevertheless, with a confidence set for (the elements of) the identified set, it is possible to project that confidence set to any function of the parameter, or component of the parameter. However, such a confidence set can have coverage rates substantially above the nominal rate, particularly when the parameter has more than just a few dimensions. Correspondingly, the power of the corresponding test would be low, making it difficult to reject false hypotheses about the function/component of the parameter. Bugni, Canay, and Shi (2017) and Kaido, Molinari, and Stoye (2019) have proposed methods for avoiding conservativeness of confidence sets for functions/components of the parameter in models with moment inequality conditions. With different additional "linearity" assumptions on the structure of the moment inequalities, Cho and Russell (2018), Andrews, Roth, and Pakes (2019), Flynn (2019), and Gafarov (2019) establish potentially computationally more appealing confidence sets for objects of interest, among other contributions. It is worth emphasizing that in a large class of models, it is possible to characterize the identified set using solutions to linear programs. This linear program characterization of the identified is helpful as it makes the computational problem (used to conduct inference for example) much easier. See for instance Syrgkanis, Tamer, and Ziani (2017).

Misspecification in partially identified models can have effects on the results that can be unexpected. Ponomareva and Tamer (2011) show that estimates of an assumed linear model with partially identified parameters due to interval-censoring of the outcome do not necessarily "approximate" a non-linear model, as would be the case in textbook ordinary least squares analysis of a linear model. Furthermore, if the true data generating process is non-linear, but the assumed model is linear, the result can be tight bounds on the parameters of interest due to very few "linear functions" fitting in between the "non-linear bounds." This suggests further reasons for only using credible assumptions, already a main theme of the partial identification literature and a main driver for using moment inequality approaches.

5.4. **Criterion function approach.** Some empirical models can be characterized by a non-negative objective function $Q(\theta) \geq 0$ such that the identified set is $\Theta_I = \{\theta : Q(\theta) = 0\}$. For example, moment inequality conditions can be characterized by a criterion function with those properties. If the true value of the parameter θ_0 satisfies moment inequality conditions $E(m(W, \theta_0)) \geq 0$, then $Q(\theta) = ||\min(E(m(W, \theta)), 0)||^2$, where $\min(E(m(W, \theta)), 0)$ is the element-wise minimum of the elements of the vector $E(m(W, \theta))$ and 0. If all elements of the vector $E(m(W, \theta))$ are non-negative, then $Q(\theta) = 0$. If any element of the vector $E(m(W, \theta))$ is negative, then $Q(\theta) > 0$. Therefore, $E(m(W, \theta)) \geq 0$ if and only if $Q(\theta) = 0$.

If there is a sample objective function $Q_N(\theta)$ that estimates $Q(\theta)$, then Chernozhukov, Hong, and Tamer (2007) shows that $\Theta_{I,N} = \{\theta : Q_N(\theta) \leq \epsilon_N\}$ estimates Θ_I in the Hausdorff distance when either $\epsilon_N = 0$ or $\epsilon_N > 0$ with $\epsilon_N \to 0$ at an appropriate rate, depending on the properties of the objective function. For example, with moment inequality conditions, $Q_N(\theta) = ||\min(E_N(m(W, \theta)), 0)||^2$. Roughly, connecting the criterion function approach to the previous discussion of estimating the identified set Δ_I when Δ_I is an interval, the $\epsilon_N = 0$ case corresponds to $\hat{\Delta}_{I,N}$, and $\epsilon_N \to 0$ case corresponds to $\hat{\Delta}^{\epsilon}_{I,N}$.

Based on determining a critical value $c_{N,\alpha}$ that accounts for the sampling variability of $\sup_{\theta \in \Theta_I} Q_N(\cdot)$, Chernozhukov, Hong, and Tamer (2007) show that $\mathcal{C}_{N,\alpha} = \{\theta : Q_N(\theta) \leq c_{N,\alpha}\}$ is a valid pointwise confidence set for the identified set in the sense that $\liminf_{N \to \infty} P(\Theta_I \subseteq \mathcal{C}_{N,\alpha}) \geq 1 - \alpha$. The critical value can be determined either by subsampling or based on asymptotic approximations for particular forms of the objective function. Similarly, based on determining a critical value $c_{N,\alpha}$ that accounts for the sampling variability of $Q_N(\theta)$, Chernozhukov, Hong, and Tamer (2007) show that $\mathcal{C}_{N,\alpha} = \{\theta : Q_N(\theta) \leq c_{N,\alpha}\}$ is a valid pointwise confidence set for the elements of the identified set in the sense that $\liminf_{N \to \infty} P(\theta' \in \mathcal{C}_{N,\alpha}) \geq 1 - \alpha$ for any $\theta' \in \Theta_I$. Romano and Shaikh (2008) and Romano and Shaikh (2010) also consider inference for partially identified models with an objective function representation, respectively considering inference for the elements of the identified set and for the identified set. Their approach is proven to be uniformly valid, avoids the need for estimating the identified set when determining the critical value, and can be smaller while maintaining coverage rates. More recently and similar to the above, Chen, Christensen, and Tamer (2018) consider inference on the identified set that is defined as the maximizer of an objective function (this covers GMM and likelihood for example). However, the construction of the confidence set is based on simulation draws from a quasi posterior that is constructed by combining this objective function with a prior. In some instances, this approach leads to easy to compute confidence sets.

5.5. **Random set approach.** A key insight and distinguishing feature of the random set approach to partial identification is the recognition that many models exhibiting partial identification can be represented as involving a *set* of random variables that are compatible with the assumptions and the observed data. Essentially, a set of random variables is a random set. The random set approach to partial identification has been summarized by Beresteanu, Molchanov, and Molinari (2012), and Molchanov and Molinari (2014, 2018). Early and influential work in this approach includes Beresteanu and Molinari (2008), Beresteanu, Molchanov, and Molinari (2011), and Beresteanu, Molchanov, and Molinari (2012).

It is possible to give enough basic background relevant for the purposes of this chapter without getting too much into the technical details of random set theory. See Molchanov (2017) or the above cited works for the technical details. Essentially, a "random closed set" X is a closed set that informally is "random" (i.e., a mapping from the elements of an underlying probability space to the space of closed sets) and, more formally, is "measurable" in the sense that $\{X \cap K \neq \emptyset\}$ is a measurable event with respect to the underlying probability space for any given compact set K. In other words, the probability that X intersects any given compact set K is defined. The probability that X intersects K is known as the "hitting" probability. Then, a "selection" of a random set X is a random quantity x, also a mapping from the elements of the underlying probability space, whose realization is contained in

X for every realization of the underlying probability space. Informally, the random set contains the random selections. More formally, Artstein's inequality says that the set of distributions of the set of selections of a random set are exactly those distributions F such that $F(K) \leq P(\{X \cap K \neq \emptyset\})$ for all compact sets K. Essentially, this characterizes the possible distributions of the "selection" associated with a random set. Finally, the Aumann expectation of a random set X is the closure of the set of expected values of the integrable selections of X. There exist laws of large numbers and central limit theorems for random sets.

Random sets relate to partial identification because often in partially identified models it is relevant to work with sets of random variables, hence random sets. For example, suppose an unobserved random variable Y is known to satisfy the inequality conditions $Y_L \leq Y \leq Y_U$ with probability 1, where (Y_L, Y_U) are observed random variables. Then, Y is a selection from the random set $[Y_L, Y_U]$.

Molinari (2020) is a review of the partial identification literature largely from the perspective of random set theory. Beresteanu and Molinari (2008) prove validity of proposed inference methods for partially identified models in which the identified set can be given a suitable random set representation. Beresteanu, Molchanov, and Molinari (2011) characterize the sharp identified set for a class of models with convex moment predictions in terms of random set theory, which in particular holds for many models based on game theory. Galichon and Henry (2011) use optimal transport theory, related to random set theory, to establish a characterization of the sharp identified set in discrete games with pure strategy Nash equilibria with distributional assumptions on the unobservables. As detailed in that paper, the introduction of core determining classes reduces the computational burden. Henry, Méango, and Queyranne (2015) prove a different characterization of the identified set, allowing for combinatorial optimization methods, which further reduces computational burden. Bontemps, Magnac, and Maurin (2012) consider linear models with some emphasis on instrumental variables.

5.6. **Bayesian approach.** In point identified models, particularly parametric or semi-parametric models, frequentist confidence sets and Bayesian credible sets are essentially the same quantity. That is, for a parameter in a point identified model, a $(1-\alpha)\%$ confidence set is essentially the same as a $(1-\alpha)\%$ credible set. One implication is that the prior used in the Bayesian analysis has negligible impact on the posterior in large samples. Essentially, the data "overwhelms" the prior in large samples. Another implication is that reported frequentist confidence sets can be used as a Bayesian credible set, and vice versa.

In a partially identified model, the prior can influence the posterior even in large samples, in particular by shaping the posterior distribution on the identified set. This can happen because the data only is informative about the location of the identified set, but not the location of the true value of the parameter within the identified set, and so the prior can still play a role in shaping the posterior on the identified set. In some extreme cases when the data reveals nothing about the parameter, the prior and posterior can be exactly the same, as in Poirier (1998). This means, in particular, that the Bayesian analysis of a partially identified model can appear to result in information about or even "rule out" certain parameter values because of prior information against those parameter values, rather than information coming from the data. In particular, Moon and Schorfheide (2012) proved that a $(1-\alpha)\%$ credible set tends to be "too small" to be a $(1-\alpha)\%$ confidence set. Liao and Jiang (2010) study in particular the posterior for a parameter that is partially identified by moment inequality conditions, requiring a limited information likelihood approach to deal with the fact that such a model does not specify a distribution of the data, and come to the same conclusion. It is also possible to test the moment inequality conditions, without a prior over the parameter, as in Kline (2011).

Therefore, Bayesian analysis of partially identified models comes with certain tradeoffs.

First, partial identification does not necessitate any change to the Bayesian approach to inference. Partial identification does necessitate changes to frequentist approaches to inference. Simply as a practical matter, this is a substantial advantage of the Bayesian approach. The posterior can still be derived as usual in partially identified model. The posterior can have potentially undesirable properties, such as depending on the prior even in large samples, but it is nevertheless the posterior. There is no definitional reason why a posterior must not depend on the prior in large samples, or indeed have any other particular property, other than following from the appropriate Bayesian updating logic. To this point, Lindley (1972, page 46) famously wrote that "unidentifiability causes no real difficulties in the Bayesian approach." Of course, as discussed also below, if the prior is "wrong" in some sense then the posterior will also be "wrong." On the other hand, applying frequentist inference methods for point identified models to partially identified models is necessarily incorrect.

Second, if the prior information really is believed by the econometrician, at least to the same degree of belief as the econometrician has in other parts of the model like the likelihood, it can be useful to include the prior information in the analysis. Indeed, not including credible prior information can result in unnecessarily learning less about the parameter than it is possible to learn. The chapter returns to this point later, as a discussion about assumptions in general. Of course, if the prior information is not credible, then it should not be used, and Bayesian analysis based on such prior information can be viewed with skepticism particularly in partially identified models where the data do not dominate the prior leading to poor results. It may be relevant to take special care to be clear that the results depend on the prior beliefs, but that would be true also for Bayesian analysis of point identified models in small samples where the data does not dominate the prior. It may be relevant to report both an estimate of the identified set and also the posterior, as suggested by Moon and Schorfheide (2012). Another possibility is to report Bayesian inference about the identified set, rather than about the parameter, as in Kline and Tamer (2016).

Third, many of the complications with frequentist inference stem from the fact that the asymptotic distribution of the proposed test statistics depends discontinuously on unknown properties of the data generating process. Put simply, this concerns repeating sampling behavior, and so this is not important for the Bayesian analysis. One particular advantage is that the Bayesian approach does not involve the determination of a critical value, which can be computationally burdensome in some frequentist approaches to inference, especially when a different critical value must be determined for each value of the parameter because of the discontinuities in the asymptotic distribution. A related advantage of the Bayesian approach is that it faces essentially no additional difficulties when doing inference on functions of the parameters or components of the parameters.

As with frequentist inference in partially identified models, there is a choice between Bayesian inference for the identified set and Bayesian inference for the elements of the identified set. Standard implementations of Bayesian inference will result in the latter, in that they result in a posterior for the parameter. However, it is also possible to consider Bayesian inference for the identified set. If so, a few different approaches are possible. Kline and Tamer (2016) is based on the existence of a representation of the identified set that is a mapping from a point identified parameter (the reduced form) to the identified set for the parameter of interest, a feature of many partially identified models. This approach has potential computational advantages compared to other (frequentist) inference methods, as it simply requires draws from the posterior of the point identified model and computation of the identified set as a function of the point identified model. In that setup, the prior is placed on the point identified parameter, which in many cases are summary statistics of the data, like the probabilities of different market outcomes in an entry game model. By using an appropriate prior/likelihood for discrete outcomes, like entry decisions, the model automatically respects the fact that the entry probabilities must be valid probabilities. It is straightforward also in that setup to use “weak priors” or priors based on minimal distributional assumptions, for the point identified parameter, like the Bayesian bootstrap.

In addition, in models based on likelihoods or objective functions such as GMM (and moment inequalities), and in cases where an explicit reduced form mapping may not be available, Chen, Christensen, and Tamer (2018) provides a quasi-Bayesian approach that allows the econometrician to draw via Monte Carlo simulations from a proposed quasi-likelihood and then use these draws to construct confidence regions for the identified set. These confidence regions have attractive large sample (frequentist) coverage properties and work well in practice.

Liao and Simoni (2019) is based on a convex identified set, a feature of many partially identified models, and is based on the support function of the identified set. This approach also can have computational advantages in particular due to the convexity. Giacomini and Kitagawa (2021) take a multiple-prior robust Bayesian approach, focusing on convex identified sets or the convex hull of non-convex identified sets.

6. Implementation of partial identification

6.1. **Computational considerations.** In point identified models, the computational problem associated with estimation of the parameter often boils down to finding the solution to an equation of the form $Q_N(\theta) = 0$ or finding the solution to a maximization problem of the form $\max_{\theta \in \Theta} Q_N(\theta)$. Although this can be a computationally intensive problem, there is a vast literature in computer science providing algorithms for solving these problems. As a practical matter, this means that an empirical researcher can in many cases simply apply existing implementations of the algorithms to the relevant problem.

In partially identified models, the computational problem often boils down to finding *all* of the solutions to an equation of the form $Q_N(\theta) = 0$, or perhaps $Q_N(\theta) \leq \epsilon_N$, or finding *all* of the (approximate) solutions to a maximization problem of the form $\max_{\theta \in \Theta} Q_N(\theta)$. This is the characterization of a partially identified model from the criterion function approach from Section 5.4, which nests other setups including moment inequality models from Section 5.3. As a consequence, it is not possible anymore to simply apply existing implementations of solver algorithms or optimization algorithms, when such algorithms return only a single solution. The computational approaches in partially identified models depends on the specifics of the model.

One possibility is a grid search, or guess-and-verify approach, that essentially amounts to checking, for many candidate values θ' of the parameter, whether $Q_N(\theta') \approx 0$ or $Q_N(\theta') \approx \max Q_N(\theta)$ as appropriate for the model. This is slow, depending in part on the computational cost of evaluating $Q_N(\cdot)$ and in part on the dimension of θ which influences the number of values of θ' that need to be checked. Inference with this approach is even slower, in particular when the inference method requires determining a different critical value for each candidate value of the parameter. Inference methods that do not have this feature are therefore of particular relevance when it is computationally costly to evaluate $Q_N(\cdot)$ and/or when θ has more than just a few elements. This grid search can focus on, or at least begin with, regions of the parameter space that are reasonable based on previous studies. In some cases, a non-sharp identified set can be easier to compute (for example because it is based on fewer moment inequality conditions, or has a simple closed form representation), which is another starting point for the grid search for the sharp identified set. Also, the grid search is embarrassingly parallel, and thus computational speedup is almost directly proportionate to the number of processing units in a parallel computing environment.

Another possibility is available when it is possible to determine a representation of the identified set that is less computationally costly to evaluate. One such case is when the identified set can be written directly in terms of population means of the observable data, for example $\Theta_I = \{\theta : E(Y_L) \leq \theta \leq E(Y_U)\}$. In such a case, the computational burden is reduced essentially to zero, at least for estimation. Inference can still be slow for the reasons outlined above. Another such case is when the identified set can be written in terms of a less costly computational problem, for example a linear programming problem rather than a non-linear programming problem.

In instances where computation is overly burdensome, it is possible to take other approaches. In particular, it is possible to work with non-sharp identified sets that are easier to compute, at the cost of learning less about the parameter. For example, in moment inequality models, computational cost is generally increasing in the number of moment inequality conditions used. If the econometrician uses fewer moment inequality conditions than are actually implied by the model, there can be a computational speedup at the cost of learning less than would be learned if all moment inequality conditions were used. Based on understanding the model, it may be possible to drop the moment inequality conditions that provide less information about the parameter. Indeed, in some models, some moment inequality conditions are redundant, in that they are implied by other moment inequality conditions, and therefore they provide no additional information about the parameter but do impose a computational cost when used.

6.2. **Simulation based approaches.** In some instances, models can be estimated using simulation based methods that are shown to be theoretically attractive and computationally simple. We highlight two approaches that can be used: Kline and Tamer (2016) and Chen, Christensen, and Tamer (2018).

In Kline and Tamer (2016), there is an explicit mapping from a reduced form parameter (often summary statistics of the data) to the identified set and Bayesian approaches are then used to map the posterior of these reduced form parameters to the identified set. For instance, consider moment inequalities of the form

$$P(Y|X=x) \leq m(x;\theta)$$

where data on (Y, X) are available, $m(\cdot)$ is a known (up to θ) mapping and the parameter of interest is θ. For instance, $P(Y|X=x)$ can be the observed probabilities of the entry decisions in an entry game. The identified set is $\Theta_I = \{\theta : P(Y|X=x) \leq m(x;\theta) \quad \forall x \in \mathcal{S}_X\}$ where $\mathcal{S}_X$ is the support of X. Suppose also that $\mathcal{S}_X$ is a finite set. The above maps $P(Y|X)$ to Θ_I via the above inequalities. The basic idea of Kline and Tamer (2016) is to first construct posterior distributions on the finite dimensional vector $P(Y|X)$ and use draws from this posterior to construct a posterior distribution for Θ_I via the identified set mapping.

In Chen, Christensen, and Tamer (2018) on the other hand, the starting point is an optimal objective function $L(\theta)$. This can be a likelihood, an optimally weighted GMM, or optimal moment inequality objective function. The identified set of interest here is

$$\Theta_I = \left\{\theta \in \Theta : L(\theta) = \sup_{\vartheta\in\Theta} L(\vartheta)\right\}.$$

This Θ_I may be a proper (nonsingleton) set or a singleton. Similar to Chernozhukov and Hong (2003), Chen, Christensen, and Tamer (2018) uses draws from a quasi posterior that uses the sample analog $L_N(\theta)$ of $L(\theta)$. These draws are based on the QLR objective function

$Q_N(\theta) = 2N[L_N(\hat{\theta}) - L_N(\theta)]$ where $L_N(\hat{\theta}) = \sup_{\theta\in\Theta} L_N(\theta) + o_p(\frac{1}{N})$. So, to construct a confidence set for Θ_I for instance, Chen, Christensen, and Tamer (2018) recommends one to obtain draws $\{\theta^1, \ldots, \theta^B\}$ from the quasi-posterior Π_N:

$$\Pi_n(A \,|\, data) = \frac{\int_A e^{NL_N(\theta)} d\Pi(\theta)}{\int_\Theta e^{NL_N(\theta)} d\Pi(\theta)}, \quad A \in B(\Theta)$$

for some prior Π on Θ, the parameter space. Then, a $100\alpha\%$ MC confidence set for Θ_I is:

$$\hat{\Theta}_\alpha = \{\theta \in \Theta : L_N(\theta) \geq \zeta^{mc}_{N,\alpha}\}$$

with $\zeta^{mc}_{N,\alpha}$ being the $(1-\alpha)$ quantile of $\{L_N(\theta^1), \ldots, L_N(\theta^B)\}$.

For instance, for the two player entry game in Section 4 above, the likelihood of the i-th observation based on normal errors (similar to Equation 8 above):

$$\begin{aligned}
\kappa_{00}(\theta) \quad &= \quad P(\epsilon_1 \leq -\beta_1, \epsilon_2 \leq -\beta_2) \\
\kappa_{11}(\theta) \quad &= \quad P(\epsilon_1 \geq -\beta_1 - \Delta_1,\, \epsilon_2 \geq -\beta_2 - \Delta_2) \\
\kappa_{10}(\theta) \quad &= \quad s \times P(-\beta_1 \leq \epsilon_1 \leq -\beta_1 - \Delta_1,\, -\beta_2 \leq \epsilon_2 \leq -\beta_2 - \Delta_2) \\
&+ \quad P(\epsilon_1 \geq -\beta_1, \epsilon_2 \leq -\beta_2) \\
&+ \quad P(\epsilon_1 \geq -\beta_1 - \Delta_1, -\beta_2 \leq \epsilon_2 \leq -\beta_2 - \Delta_2)
\end{aligned}$$

where s denotes the equilibrium selection probability which is here treated as a parameter to be estimated. It is clear that the above choice probabilities do not point identify the parameter vector $\theta = (\beta_1, \beta_2, \Delta_1, \Delta_2, \rho, s)$ where ρ is the correlation of the normal errors (normalized to have variance 1). The likelihood of the i-th observation $(D_{00,i}, D_{10,i}, D_{11,i}, D_{01,i}) = (d_{00}, d_{10}, d_{11}, 1 - d_{00} - d_{10} - d_{11})$ is:

$$l(\theta) = [\kappa_{00}(\theta)]^{d_{00}}[\kappa_{10}(\theta)]^{d_{10}}[\kappa_{11}(\theta)]^{d_{11}}[1 - \kappa_{00}(\theta) - \kappa_{10}(\theta) - \kappa_{11}(\theta)]^{1-d_{00}-d_{10}-d_{11}}$$

The objective function $L(\theta)$ that is used in the simulation procedure above would then be the likelihood of the data. Simulation and empirical results for this model are provided in

Chen, Christensen, and Tamer (2018) where sequential Monte Carlo algorithms are used to obtain the sequence of draws.

More generally, as long as one obtains an optimal objective function such as a likelihood, one can then use the above quasi posterior to construct valid frequentist confidence sets[44] for Θ_I.

6.3. **Reporting empirical results from a partially identified model.**

6.3.1. *The overall identified set, marginal identified sets, or the identified sets for objects of interest.* In partially identified models, as with point identified models, it is common in empirical practice to report results for individual components of the parameter. If $\theta = (\theta_1, \theta_2, \ldots, \theta_K)$ is the parameter of the model, the identified set for θ is Θ_I and the identified set for any individual component θ_k is $\Theta_{I,k}$. It is common empirical practice to report the identified sets $\Theta_{I,k}$, for example the identified sets of the coefficients of a linear function that parameterizes some important part of the model, like the utility function.

The collection of the identified sets $\Theta_{I,k}$ for all individual components of the parameter contains less information about the parameters than does the identified set Θ_I for the overall parameter. When Θ_I is known by the econometrician to be the Cartesian product of $\Theta_{I,k}$'s, the information contained in the $\Theta_{I,k}$'s is the same as the information contained in Θ_I. Otherwise, in general, Θ_I contains information about restrictions on the parameter across components of the parameter: certain specifications of some components of the parameter can imply restrictions on other components of the parameter.

In such cases, it can be useful for empirical research to report Θ_I, or at least the identified set for relevant combinations of the components of the parameter. For example, if θ_1 and θ_2 are of particular interest, it can be useful to report the identified set for (θ_1, θ_2), rather than the identified set for θ_1 and the identified set for θ_2. If the audience of the empirical research has prior beliefs about θ_1, reporting the identified set for (θ_1, θ_2) allows the reader to

[44]Chen, Christensen, and Tamer (2018) contains procedures to construct confidence sets for subsets of θ. Also, Chen, Christensen, and Tamer (2018) contains a simple grid search based procedure (no simulation draws required) that allows one to obtain confidence intervals on scalar subvectors of θ (such as Δ_1) by comparing a profiled version of the objective function to a χ^2 cutoff.

draw conclusions about θ_2. In particular, if the audience has information that restricts θ_1 to be within a proper subset of $\Theta_{I,1}$, which in particular could happen if some other empirical research reports a narrower identified set for θ_1, the audience can use that combined with the identified set for (θ_1, θ_2) to come up with an identified set for θ_2 that is a proper subset of $\Theta_{I,2}$. For example, in the context of the model of market entry from Section 4, reporting the identified set for $(\beta_1, \Delta_1, \beta_2, \Delta_2)$ makes it possible for the audience to use information about β_1 (the effect of observable characteristics of the firms and the market on profits) to draw conclusions about Δ_1 (the competitive effect of entry on profits), for example.

If θ has one, two, or three real-valued components, then it is possible to visually report the identified set for θ. If θ has more than three real-valued components, or if θ contains components that are not real-valued (e.g., a distribution of an unobservable), then it can be much more difficult to visually report the identified set for θ. One possibility is to partition θ into $\theta = (\theta^{(1)}, \theta^{(2)})$, where $\theta^{(1)}$ contains at most three real-valued components. Then, it is possible to report a collection of identified sets for $\theta^{(1)}$ at various specified values for $\theta^{(2)}$. This can demonstrate to the audience how the identification region of $\theta^{(1)}$ depends on $\theta^{(2)}$.

This consideration is unique to partially identified models. In point identified models, the collection of estimates of all individual components θ_k is exactly as much information as an estimate of θ. This is a trivial observation, but contrasts with the situation in partially identified models.

In some cases, it is relevant to report the identified set for some other object of interest. One such case is the identified set for counterfactual outcomes, perhaps under alternative policy interventions. Counterfactuals in models with incompleteness, for example multiple equilibria, present particular questions even setting aside partial identification, as discussed in Section 6.3.2. Other such cases include other functions of the parameters, for example marginal effects in non-linear models. In all such cases, inference methods that accommodate working with functions of the parameter, rather than just the parameter itself, are necessary, and were discussed in Section 5.

6.3.2. *Counterfactuals in models with incompleteness and/or multiple equilibria.* Often, partial identification arises in models with incompleteness and/or with multiple equilibria. Regardless of the point identification or partial identification of the parameters of the model, counterfactual predictions in these models require that the econometrician state how the counterfactual outcomes are determined in cases of incompleteness and/or multiple equilibria. Obtaining predictive distributions for the purposes of evaluating policy counterfactuals is an important component of any applied work in economics and naturally the ability to provide model anchored counterfactuals using data is the natural next step for any inference approach.

One possibility in the current setup is to report counterfactual predictions that make no assumption on the selection mechanism. A necessary implication of this is that the counterfactual outcomes can only be predicted to be within some bounds, corresponding to the multiple equilibrium outcomes. In general, this approach sticks most closely to the conclusions from the theoretical model, but it may be desirable to consider approaches that result in tighter counterfactual predictions. In addition, enumerating the set of equilibria may be computationally intensive and can be totally impractical in some models. This problem is not generic to partially identified or moment inequality models and are more a function of prediction in models with multiple equilibria. Generally in these models, counterfactual predictions will be partially identified unless the selection mechanism is point identified (and assumed to also apply to counterfactuals).

Another possibility is to use the identified set for the selection mechanism to refine the counterfactual predictions. In some models, this might entail using point identification of the selection mechanism. However, caution is warranted with this approach. Relative to the utility functions, it is comparatively less clear whether the selection mechanism is policy-invariant (e.g., Lucas (1976)), and as such it may be a concern that the selection mechanism in the observed data is not reflective of the selection mechanism in counterfactual worlds. Fundamentally the problem is we don't know what causes the observed selection, so we don't know if the counterfactual will change the selection.

With the aim of tighter counterfactual predictions, another possibility is to report counterfactual predictions using the assumption of a known, specific selection mechanism. If the model parameters are partially identified, the counterfactual predictions generally will also be partially identified, even with the assumption of a specific selection mechanism. In some cases, counterfactual outcomes allowing for any selection mechanism can be bounded by the counterfactual outcome based on a specific selection mechanism. For example, the counterfactual outcome of a particular firm is necessarily less than the counterfactual outcome in the case when the selection mechanism selects the equilibrium that maximizes that outcome. In that circumstance, using counterfactual predictions based on a specific selection mechanism can simplify the interpretation or computation of the results, and does not change what information the econometrician is reporting. In some cases, the assumption of a known selection mechanism can be used in the identification strategy for model parameters; in other cases, the assumption of a known selection mechanism can be used only for the purposes of counterfactual predictions. See for example the discussion in Section 4 for assumptions on the selection mechanism.

Yet another possibility is to use some more theoretically motivated way of selecting among the multiple equilibria. Lee and Pakes (2009) suggest using learning models, specifically either a best-response dynamics model of learning or a fictitious play model of learning. By computing which equilibria these learning models converge to, given the primitives of the underlying model, it is possible to come up with counterfactual predictions in response to a policy intervention, particularly when the learning process "begins" at a reasonable "pre-intervention" point. Note though that in these models it is likely that this learning dynamics converges to a distribution of equilibria, assigning probabilities to the different ones. So, one may be able to present this distribution over equilibria or moments of it. See also Wollmann (2018). Jun and Pinkse (2020) consider multiple approaches for point prediction of counterfactual outcomes in a complete information game, along the lines of those considered in Section 4, ultimately recommending a maximum entropy approach. Essentially, the maximum entropy approach makes predictions based on the specification that the unknown selection mechanism is as close to the uniform distribution as possible given the assumptions and observable data, since the uniform distribution maximizes entropy when there are not constraints.

7. Conclusions

Compared to point identification results, partial identification necessarily results in the econometrician learning less about the object of interest. In some applications, this can imply that there are equivocal answers to major qualitative questions like whether a particular policy intervention will have a positive effect or negative effect on targeted outcomes, with the answer being that both a positive effect and a negative effect are compatible with the assumptions and the data. In other applications, this can imply that there are definitive answers to major qualitative questions like whether a particular policy intervention will have a positive effect or negative effect on target outcomes, yet less equivocal answers to quantitative questions like the magnitude of the effect of the policy intervention.

Manski (2003, page 1) and Manski (2009, page 3) has referred to the tradeoff between the strength of the assumptions and the credibility of the results as the Law of Decreasing Credibility. Further, almost by definition, strictly weaker assumptions are simultaneously weakly more credible and result in learning weakly less about the objects of interest.[45] Generally, empirical research tries to achieve the twin goals of being credible and coming to definitive conclusions. Unfortunately, as above, there is generally some tradeoff between these goals. Prioritizing point identification entails a very strong position on the tradeoff between the credibility of the assumptions and the definitiveness of the empirical findings. Although it can be good for researchers to find the strongest assumptions that are credible in a given empirical setting, it is not clear that should necessarily lead to learning a lot about the

[45]However, there are important instances where strictly weaker assumptions are strictly more credible and yet result in learning the same amount about the objects of interest. In those instances, essentially there is a free lunch. One familiar example of this phenomenon is when point identification can be established under both weaker and stronger assumptions, often with the point identification result based on stronger assumptions appearing in the literature before the point identification result based on weaker assumptions.

objects of interest. It can be an important result to show that under the strongest credible assumptions, not much can be learned about the objects of interest. Further, particularly for identification strategies that are constructive, it can be possible to report and focus on "why" not much can be learned about the objects of interest and, perhaps, correspondingly collect additional data. For instance, if some sort of "non-response" is an important factor, then the econometrician can consider whether it would be possible to incentivize higher response rates (e.g., Horowitz and Manski (1998)). Research that provides guidance for collection of more informative data can be an important contribution to the literature. Moreover, in some cases, the only known identification strategy is a partial identification strategy.

On the other hand, partial identification results are not *per se* necessarily better than point identification results. Again, there is a tradeoff between assumptions and conclusions. That tradeoff does not necessarily favor using weaker assumptions in all cases. Reporting identified sets based on unnecessarily weak assumptions can lead to unnecessarily pessimistic conclusions about the inability to learn much about the object of interest. Even if an assumption is at best "approximately" true, it can be worthwhile to use that assumption as an approximation, as compared to avoiding the use of that assumption entirely. It is not necessarily good research practice to avoid using credible assumptions, or reasonable assumptions, or assumptions that are motivated by economic theory simply for the purpose of not using assumptions. On the basis that assumptions and data are twin inputs to the empirical research finding, avoiding using credible or reasonable assumptions would be roughly the same as avoiding using imperfect but reasonable data.

It is also important to alert empirical researchers using moment inequalities to the problem of misspecification. Again, the motivation for the use of weak assumptions is the tradeoff between the strength of assumptions and credibility of the results. However, moment inequality models in general can still be misspecified. In that case, since the "true" identified set is empty (because the intersection of these moment inequalities is empty), so estimates from such (misspecified) models (especially ones that use many moment inequalities) appear to be "tight" and so researchers in these cases can get point-like estimates in a partially identified model. A note of caution here is in order. Empirical researchers using these models should try to probe as to whether these estimates are sensitive to the choice of the set of moment inequalities used for instance,[46] or other specification of the model. The theoretical literature[47] on misspecification in econometrics and especially in partially identified models and moment inequality models is not as developed and we view work there as important given its relevance to empirical work.

Another issue is that software implementations of the methods for estimation and inference in partially identified models are limited. Recent implementations in Stata include the work of Chernozhukov, Lee, and Rosen (2013) in `clrbound` and related commands in Chernozhukov, Kim, Lee, and Rosen (2015), the implementation of Manski and Pepper (2000) and related papers in `tebounds` in McCarthy, Millimet, and Roy (2015), and the implementation of Andrews and Shi (2013) in `cmi_test` in Andrews, Kim, and Shi (2017). More recently, computationally simpler inference approaches such as Chen, Christensen, and Tamer (2018), Chernozhukov, Chetverikov, and Kato (2019), Cox and Shi (2019), Kaido, Molinari, and Stoye (2019), and Kline and Tamer (2016) can be used for both full and subvector inference.

Overall, in this chapter, we highlighted approaches to identification and inference in models in Industrial Organization with partial identification and/or moment inequalities. The attractiveness of these methods from an applied econometrics perspective is that the econometrician is able to learn about the parameters in models of interest without using the strong assumptions required for these models to be point identified. Rather, the methods (moment inequalities or others) are derived directly from optimizing behavior, are built on assumptions that are theoretically motivated, and econometric methods are available to provide theoretically attractive inference results.

[46] For instance, one can examine whether the estimates change substantially when using a different set of moment inequalities. This usually is a sign that the model is misspecified and as with misspecified models in general, different moments estimate different parameters (or pseudo-true value). So, care should be taken when comparing estimates from different sets of (misspecified) moment inequalities as these may be estimating different parameters.

[47] For some work on this, see Ponomareva and Tamer (2011) and Andrews and Kwon (2019).

References

Abbring, J. H., J. J. Heckman, P.-A. Chiappori, and J. Pinquet (2003): "Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish?," *Journal of the European Economic Association*, 1(2-3), 512–521.

Ackerberg, D. A., and G. Gowrisankaran (2006): "Quantifying equilibrium network externalities in the ACH banking industry," *The RAND Journal of Economics*, 37(3), 738–761.

Afriat, S. N. (1967): "The construction of utility functions from expenditure data," *International Economic Review*, 8(1), 67–77.

Afriat, S. N. (1973): "On a system of inequalities in demand analysis: an extension of the classical method," *International Economic Review*, 14(2), 460–472.

Ahn, H., and J. L. Powell (1993): "Semiparametric estimation of censored selection models with a nonparametric selection mechanism," *Journal of Econometrics*, 58(1-2), 3–29.

An, Y. (2017): "Identification of first-price auctions with non-equilibrium beliefs: A measurement error approach," *Journal of Econometrics*, 200(2), 326–343.

Andrews, D. W. (2000): "Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space," *Econometrica*, 68(2), 399–405.

Andrews, D. W., and P. J. Barwick (2012): "Inference for parameters defined by moment inequalities: A recommended moment selection procedure," *Econometrica*, 80(6), 2805–2826.

Andrews, D. W., W. Kim, and X. Shi (2017): "Commands for testing conditional moment inequalities and equalities," *The Stata Journal*, 17(1), 56–72.

Andrews, D. W., and S. Kwon (2019): "Inference in moment inequality models that is robust to spurious precision under model misspecification."

Andrews, D. W., and X. Shi (2013): "Inference based on conditional moment inequalities," *Econometrica*, 81(2), 609–666.

Andrews, D. W., and X. Shi (2014): "Nonparametric inference based on conditional moment inequalities," *Journal of Econometrics*, 179(1), 31–45.

Andrews, D. W., and G. Soares (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78(1), 119–157.

ANDREWS, I., J. ROTH, AND A. PAKES (2019): "Inference for linear conditional moment inequalities."

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

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

ARADILLAS-LÓPEZ, A., A. GANDHI, AND D. QUINT (2013): "Identification and inference in ascending auctions with correlated private values," *Econometrica*, 81(2), 489–534.

ARADILLAS-LÓPEZ, A., A. GANDHI, AND D. QUINT (2016): "A simple test for moment inequality models with an application to English auctions," *Journal of Econometrics*, 194(1), 96–115.

ARADILLAS-LÓPEZ, A., AND A. M. ROSEN (2018): "Inference in ordered response games with complete information."

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

ARMSTRONG, T. B. (2013): "Bounds in auctions with unobserved heterogeneity," *Quantitative Economics*, 4(3), 377–415.

ARMSTRONG, T. B. (2014): "Weighted KS statistics for inference on conditional moment inequalities," *Journal of Econometrics*, 181(2), 92–116.

ARMSTRONG, T. B. (2015): "Asymptotically exact inference in conditional moment inequality models," *Journal of Econometrics*, 186(1), 51–65.

ARYAL, G., AND D.-H. KIM (2013): "A point decision for partially identified auction models," *Journal of Business & Economic Statistics*, 31(4), 384–397.

ASKER, J., C. FERSHTMAN, J. JEON, AND A. PAKES (2020): "A computational framework for analyzing dynamic auctions: The market impact of information sharing," *The RAND Journal of Economics*, 51(3), 805–839.

ATHEY, S., AND P. A. HAILE (2007): "Nonparametric approaches to auctions," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6A, chap. 60, pp. 3847–3965. Elsevier.

AUMANN, R., AND A. BRANDENBURGER (1995): "Epistemic conditions for Nash equilibrium," *Econometrica*, 63(5), 1161–1180.

Bajari, P., C. L. Benkard, and J. Levin (2007): "Estimating dynamic models of imperfect competition," *Econometrica*, 75(5), 1331–1370.

Bajari, P., J. Hahn, H. Hong, and G. Ridder (2011): "A note on semiparametric estimation of finite mixtures of discrete choice models with application to game theoretic models," *International Economic Review*, 52(3), 807–824.

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.

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 (2005): "Robust mechanism design," *Econometrica*, 73(6), 1771–1813.

Bernheim, B. D. (1984): "Rationalizable strategic behavior," *Econometrica*, 52(4), 1007–1028.

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

Berry, S. T., and G. Compiani (2020): "An instrumental variable approach to dynamic models."

Bjorn, P. A., and Q. H. Vuong (1984): "Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation."

Blundell, R., and R. J. Smith (1994): "Coherency and estimation in simultaneous models with censored or qualitative dependent variables," *Journal of Econometrics*, 64(1-2), 355–373.

Bontemps, C., T. Magnac, and E. Maurin (2012): "Set identified linear models," *Econometrica*, 80(3), 1129–1155.

Bresnahan, T. F., and P. C. Reiss (1990): "Entry in monopoly market," *The Review of Economic Studies*, 57(4), 531–553.

Bresnahan, T. F., and P. C. Reiss (1991a): "Empirical models of discrete games," *Journal of Econometrics*, 48(1-2), 57–81.

Bresnahan, T. F., and P. C. Reiss (1991b): "Entry and competition in concentrated markets," *Journal of Political Economy*, 99(5), 977–1009.

Bugni, F. A. (2010): "Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set," *Econometrica*, 78(2), 735–753.

Bugni, F. A., I. A. Canay, and X. Shi (2017): "Inference for subvectors and other functions of partially identified parameters in moment inequality models," *Quantitative Economics*, 8(1), 1–38.

Camerer, C. F. (2011): *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press.

Camerer, C. F., T.-H. Ho, and J.-K. Chong (2004): "A cognitive hierarchy model of games," *The Quarterly Journal of Economics*, 119(3), 861–898.

Canay, I. A. (2010): "EL inference for partially identified models: Large deviations optimality and bootstrap validity," *Journal of Econometrics*, 156(2), 408–425.

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

Card, D., and L. Giuliano (2013): "Peer effects and multiple equilibria in the risky behavior of friends," *Review of Economics and Statistics*, 95(4), 1130–1149.

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003): "2001 Lawrence R. Klein lecture: Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice," *International Economic Review*, 44(2), 361–422.

Chamberlain, G. (1980): "Analysis of covariance with qualitative data," *The Review of Economic Studies*, 47(1), 225–238.

Chambers, C. P., and F. Echenique (2016): *Revealed Preference Theory*, Econometric Society Monographs. Cambridge University Press.

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

Chen, X., E. T. Tamer, and A. Torgovitsky (2011): "Sensitivity analysis in semiparametric likelihood models."

Chernozhukov, V., D. Chetverikov, and K. Kato (2019): "Inference on causal and structural parameters using many moment inequalities," *The Review of Economic Studies*, 86(5), 1867–1900.

Chernozhukov, V., and H. Hong (2003): "An MCMC approach to classical estimation," *Journal of Econometrics*, 115(2), 293–346.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and confidence regions for parameter sets in econometric models 1," *Econometrica*, 75(5), 1243–1284.

Chernozhukov, V., W. Kim, S. Lee, and A. M. Rosen (2015): "Implementing intersection bounds in Stata," *The Stata Journal*, 15(1), 21–44.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection bounds: Estimation and inference," *Econometrica*, 81(2), 667–737.

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

Cho, J., and T. M. Russell (2018): "Simple inference on functionals of set-identified parameters defined by linear moments."

Ciliberto, F., C. Murry, and E. Tamer (2021): "Market structure and competition in airline markets."

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

Coey, D., B. Larsen, K. Sweeney, and C. Waisman (2017): "Ascending auctions with bidder asymmetries," *Quantitative Economics*, 8(1), 181–200.

Conley, T. G., and C. R. Udry (2010): "Learning about a new technology: Pineapple in Ghana," *American Economic Review*, 100(1), 35–69.

Costa-Gomes, M., V. P. Crawford, and B. Broseta (2001): "Cognition and behavior in normal-form games: An experimental study," *Econometrica*, 69(5), 1193–1235.

Costa-Gomes, M. A., and V. P. Crawford (2006): "Cognition and behavior in two-person guessing games: An experimental study," *American Economic Review*, 96(5), 1737–1768.

Cox, G., and X. Shi (2019): "Simple adaptive size-exact testing for full-vector and subvector inference in moment inequality models."

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

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013): "Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications," *Journal of Economic Literature*, 51(1), 5–62.

Crawford, V. P., and N. Iriberri (2007): "Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions?," *Econometrica*, 75(6), 1721–1770.

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

Deaton, A. (1986): "Demand analysis," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 3, chap. 30, pp. 1767–1839. Elsevier.

Dickstein, M. J., and E. Morales (2018): "What do exporters know?," *The Quarterly Journal of Economics*, 133(4), 1753–1801.

Diewert, W. E. (1973): "Afriat and revealed preference theory," *The Review of Economic Studies*, 40(3), 419–425.

Doraszelski, U., G. Lewis, and A. Pakes (2018): "Just starting out: Learning and equilibrium in a new market," *American Economic Review*, 108(3), 565–615.

Eizenberg, A. (2014): "Upstream innovation and product variety in the US home PC market," *Review of Economic Studies*, 81(3), 1003–1045.

Ellickson, P. B., S. Houghton, and C. Timmins (2013): "Estimating network economies in retail chains: a revealed preference approach," *The RAND Journal of Economics*, 44(2), 169–193.

Erev, I., and A. E. Roth (1998): "Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria," *American Economic Review*, 88(4), 848–881.

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

Feltovich, N. (2000): "Reinforcement-based vs. belief-based learning models in experimental asymmetric-information games," *Econometrica*, 68(3), 605–641.

Fershtman, C., and A. Pakes (2012): "Dynamic games with asymmetric information: A framework for empirical work," *The Quarterly Journal of Economics*, 127(4), 1611–1661.

Flavin, M. A. (1981): "The adjustment of consumption to changing expectations about future income," *Journal of Political Economy*, 89(5), 974–1009.

Flynn, Z. (2019): "Inference based on continuous linear inequalities via semi-infinite programming."

Gafarov, B. (2019): "Inference in high-dimensional set-identified affine models."

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

Gentry, M., and T. Li (2014): "Identification in auctions with selective entry," *Econometrica*, 82(1), 315–344.

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

Gillen, B. J. (2009): "Identification and estimation of level-k auctions."

Goldberger, A. S. (1968): *Topics in Regression Analysis*. Macmillan.

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.

Haile, P. A., and Y. Kitamura (2019): "Unobserved heterogeneity in auctions," *The Econometrics Journal*, 22(1), C1–C19.

Haile, P. A., and E. Tamer (2003): "Inference with an incomplete model of English auctions," *Journal of Political Economy*, 111(1), 1–51.

Hanoch, G., and M. Rothschild (1972): "Testing the assumptions of production theory: a nonparametric approach," *Journal of Political Economy*, 80(2), 256–275.

Hansen, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50(4), 1029–1054.

Hansen, L. P., and K. J. Singleton (1982): "Generalized instrumental variables estimation of nonlinear rational expectations models," *Econometrica*, 50(5), 1269–1286.

Heckman, J. J. (1978a): "Dummy endogenous variables in a simultaneous equation system," *Econometrica*, 46(4), 931–959.

Heckman, J. J. (1978b): "Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence," *Annales de l'inséé*, 30/31, 227–269.

Heckman, J. J. (1981): "Heterogeneity and State Dependence," in *Studies in Labor Markets*, ed. by S. Rosen, chap. 3, pp. 91–140. University of Chicago Press.

Hendricks, K., and R. H. Porter (2007): "An empirical perspective on auctions," in *Handbook of Industrial Organization*, ed. by M. Armstrong, and R. H. Porter, vol. 3, chap. 32, pp. 2073–2143. Elsevier.

Henry, M., R. Méango, and M. Queyranne (2015): "Combinatorial approach to inference in partially identified incomplete structural models," *Quantitative Economics*, 6(2), 499–529.

Henry, M., and A. Onatski (2012): "Set coverage and robust policy," *Economics Letters*, 115(2), 256–257.

Ho, K., and A. Pakes (2014): "Hospital choices, hospital prices, and financial incentives to physicians," *American Economic Review*, 104(12), 3841–84.

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

Holmes, T. J. (2011): "The diffusion of Wal-Mart and economies of density," *Econometrica*, 79(1), 253–302.

Horowitz, J. L. (2019): "Bootstrap methods in econometrics," *Annual Review of Economics*, 11, 193–224.

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

Horowitz, J. L., and C. F. Manski (1998): "Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations," *Journal of Econometrics*, 84(1), 37–58.

Horowitz, J. L., and C. F. Manski (2000): "Nonparametric analysis of randomized experiments with missing covariate and outcome data," *Journal of the American Statistical Association*, 95(449), 77–84.

Hortaçsu, A., and D. McAdams (2010): "Mechanism choice and strategic bidding in divisible good auctions: An empirical analysis of the turkish treasury auction market," *Journal of Political*

Economy, 118(5), 833–865.

Hortaçsu, A., and D. McAdams (2018): "Empirical work on auctions of multiple objects," *Journal of Economic Literature*, 56(1), 157–84.

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

Illanes, G. (2017): "Switching costs in pension plan choice."

Imbens, G. W., and C. F. Manski (2004): "Confidence intervals for partially identified parameters," *Econometrica*, 72(6), 1845–1857.

Imbens, G. W., and D. B. Rubin (2015): *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Ishii, J. (2004): "Compatibility, Competition, and Investment in Network Industries: ATM Networks in the Banking Industry," Ph.D. thesis, Harvard University.

Jia, P. (2008): "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry," *Econometrica*, 76(6), 1263–1316.

Jovanovic, B. (1989): "Observable implications of models with multiple equilibria," *Econometrica*, 57(6), 1431–1437.

Jun, S. J., and J. Pinkse (2020): "Counterfactual prediction in complete information games: point prediction under partial identification," *Journal of Econometrics*, 216(2), 394–429.

Kaido, H., F. Molinari, and J. Stoye (2019): "Confidence intervals for projections of partially identified parameters," *Econometrica*, 87(4), 1397–1432.

Kaido, H., and A. Santos (2014): "Asymptotically efficient estimation of models defined by convex moment inequalities," *Econometrica*, 82(1), 387–413.

Kastl, J. (2011): "Discrete bids and empirical inference in divisible good auctions," *The Review of Economic Studies*, 78(3), 974–1014.

Katz, M. (2007): "Supermarkets and Zoning Laws," Ph.D. thesis, Harvard University.

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. (2018): "An empirical model of non-equilibrium behavior in games," *Quantitative Economics*, 9(1), 141–181.

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. Elsevier.

Komarova, T. (2013): "Partial identification in asymmetric auctions in the absence of independence," *The Econometrics Journal*, 16(1), S60–S92.

Kooreman, P. (1994): "Estimation of econometric models of some discrete games," *Journal of Applied Econometrics*, 9(3), 255–268.

Kotz, S., N. Balakrishnan, and N. L. Johnson (2000): *Continuous Multivariate Distributions, Volume 1: Models and Applications.* John Wiley & Sons, 2 edn.

Kreider, B., and J. V. Pepper (2007): "Disability and employment: reevaluating the evidence in light of reporting errors," *Journal of the American Statistical Association*, 102(478), 432–441.

Lazzati, N. (2015): "Treatment response with social interactions: Partial identification via monotone comparative statics," *Quantitative Economics*, 6(1), 49–83.

Lee, R. S., and A. Pakes (2009): "Multiple equilibria and selection by learning in an applied setting," *Economics Letters*, 104(1), 13–16.

Lee, S., K. Song, and Y.-J. Whang (2018): "Testing for a general class of functional inequalities," *Econometric Theory*, 34(5), 1018–1064.

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.

Lindley, D. V. (1972): *Bayesian Statistics, A Review*, vol. 2 of *CBMS-NSF Regional Conference Series in Applied*. SIAM.

Lucas, R. E. (1976): "Econometric policy evaluation: A critique," in *Carnegie-Rochester Conference Series on Public Policy*, vol. 1, pp. 19–46.

Magnolfi, L., and C. Roncoroni (2017): "Estimation of discrete games with weak assumptions on information."

Manski, C. F. (1987): "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica*, 55(2), 357–362.

Manski, C. F. (1988a): *Analog Estimation Methods in Econometrics*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman and Hall.

Manski, C. F. (1988b): "Identification of binary response models," *Journal of the American statistical Association*, 83(403), 729–738.

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. (1996): "Learning about treatment effects from experiments with random assignment of treatments," *Journal of Human Resources*, 31(4), 709–733.

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. (2013): "Identification of treatment response with social interactions," *The Econometrics Journal*, 16(1), S1–S23.

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.

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.

Mazzeo, M. J. (2002): "Product choice and oligopoly market structure," *RAND Journal of Economics*, 33(2), 221–242.

McAdams, D. (2008): "Partial identification and testable restrictions in multi-unit auctions," *Journal of Econometrics*, 146(1), 74–85.

McCarthy, I., D. L. Millimet, and M. Roy (2015): "Bounding treatment effects: A command for the partial identification of the average treatment effect with endogenous and misreported treatment assignment," *The Stata Journal*, 15(2), 411–436.

McFadden, D. L. (2005): "Revealed stochastic preference: a synthesis," *Economic Theory*, 26(2), 245–264.

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

Milgrom, P. R., and R. J. Weber (1982): "A theory of auctions and competitive bidding," *Econometrica*, 50(5), 1089–1122.

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

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

Molchanov, I. S. (2017): *Theory of Random Sets*, Probability and Its Applications. Springer, 2 edn.

Molinari, F. (2008): "Partial identification of probability distributions with misclassified data," *Journal of Econometrics*, 144(1), 81–117.

Molinari, F. (2010): "Missing treatments," *Journal of Business & Economic Statistics*, 28(1), 82–95.

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

Molinari, F., and M. Peski (2006): "Generalization of a result on "regressions, short and long"," *Econometric Theory*, 22(1), 159–163.

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

Morales, E., G. Sheu, and A. Zahler (2019): "Extended gravity," *The Review of Economic Studies*, 86(6), 2668–2712.

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

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

Nishida, M. (2015): "Estimating a model of strategic network choice: The convenience-store industry in Okinawa," *Marketing Science*, 34(1), 20–38.

Nosko, C. (2010): "Competition and quality choice in the CPU market."

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. (1994): "Dynamic structural models, problems and prospects: mixed continuous discrete controls and market interaction," in *Advances in Econometrics, Sixth World Congress*, ed. by C. Sims, vol. 2, chap. 5, pp. 171–274. Cambridge University Press.

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

Pakes, A. (2014): "The 2013 Lawrence R. Klein Lecture: Behavioral and descriptive forms of choice models," *International Economic Review*, 55(3), 603–624.

Pakes, A., and P. McGuire (1994): "Computing Markov-Perfect Nash equilibria: Numerical implications of a dynamic differentiated product model," *RAND Journal of Economics*, 25(4), 555–589.

Pakes, A., and P. McGuire (2001): "Stochastic algorithms, symmetric Markov perfect equilibrium, and the 'curse' of dimensionality," *Econometrica*, 69(5), 1261–1281.

Pakes, A., and J. Porter (2019): "Moment inequalities for multinomial choice with fixed effects."

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: Eleventh World Congress*, ed. by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, vol. 1, chap. 8, pp. 268–323. Cambridge University Press Cambridge.

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

Pearce, D. G. (1984): "Rationalizable strategic behavior and the problem of perfection," *Econometrica*, 52(4), 1029–1050.

Phillips, P. C. (1989): "Partially identified econometric models," *Econometric Theory*, 5(2), 181–240.

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.

Powell, J. L. (1986): "Censored regression quantiles," *Journal of Econometrics*, 32(1), 143–155.

Reiss, P. C., and P. T. Spiller (1989): "Competition and entry in small airline markets," *The Journal of Law and Economics*, 32(2, Part 2), S179–S202.

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.

Romano, J. P., A. M. Shaikh, and M. Wolf (2014): "A practical two-step method for testing moment inequalities," *Econometrica*, 82(5), 1979–2002.

Rosen, A. M. (2008): "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146(1), 107–117.

Salmon, T. C. (2001): "An evaluation of econometric models of adaptive learning," *Econometrica*, 69(6), 1597–1628.

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

Santos, A. (2012): "Inference in nonparametric instrumental variables with partial identification," *Econometrica*, 80(1), 213–275.

Seim, K. (2006): "An empirical model of firm entry with endogenous product-type choices," *The RAND Journal of Economics*, 37(3), 619–640.

Shi, X., M. Shum, and W. Song (2018): "Estimating semi-parametric panel multinomial choice models using cyclic monotonicity," *Econometrica*, 86(2), 737–761.

Soetevent, A. R., and P. Kooreman (2007): "A discrete-choice model with social interactions: with an application to high school teen behavior," *Journal of Applied Econometrics*, 22(3), 599–624.

Song, K. (2014): "Point decisions for interval-identified parameters," *Econometric Theory*, 30(2), 334–356.

Syrgkanis, V., E. Tamer, and J. Ziani (2017): "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, 167–195.

Tan, T. C.-C., and S. R. da Costa Werlang (1988): "The Bayesian foundations of solution concepts of games," *Journal of Economic Theory*, 45(2), 370–391.

Tang, X. (2011): "Bounds on revenue distributions in counterfactual auctions with reserve prices," *The RAND Journal of Economics*, 42(1), 175–203.

Varian, H. R. (1982): "The nonparametric approach to demand analysis," *Econometrica*, 50(4), 945–973.

Varian, H. R. (1983): "Non-parametric tests of consumer behaviour," *The Review of Economic Studies*, 50(1), 99–110.

Varian, H. R. (1984): "The nonparametric approach to production analysis," *Econometrica*, 52(3), 579–597.

Wollmann, T. G. (2018): "Trucks without bailouts: Equilibrium product characteristics for commercial vehicles," *American Economic Review*, 108(6), 1364–1406.