THE EMPIRICAL CONTENT OF GAMES WITH BOUNDED REGRESSORS

BRENDAN KLINE

UNIVERSITY OF TEXAS AT AUSTIN

Abstract. This paper develops a strategy for identification and estimation of complete information games that does not require a regressor that has large support, nor a parametric specification for the distribution of the unobservables. The identification result is a consequence of the fact that the complete information game framework has substantial empirical content for all values of the explanatory variables. The identification and estimation of the interaction effect parameter uses a non-standard but plausible condition on the unobservables: the assumption that the mode of the joint distribution of the unobservables of all agents is zero. A three-step semiparametric estimator is proposed that is based on this identification result. Under mild regularity conditions, the estimator is consistent and asymptotically normally distributed. The estimator is non-standard in the sense that the estimator of the interaction effect parameter converges at slower than the parametric rate. An intermediate result of this paper, potentially of independent interest, concerns identification and estimation of the direction of the interaction effect.

JEL codes: C14, C31, C72, L13

Keywords: identification, non-standard estimation, strategic interaction, entry game

1. Introduction

1.1. **Overview.** It is difficult to establish conditions under which the parameters of a complete information game are point identified, for many reasons. First, the outcomes of the model appear also as "explanatory variables," since the utility function of a particular player depends on the choices of the other players, resulting in a sort of simultaneity/endogeneity problem roughly analogous to standard simultaneous equations models. Second, and more unique to the context of games, the model likelihood is not uniquely determined by the parameters of the utility functions, because there

Date: October 4, 2013. Some of the results in this paper were previously distributed in an earlier version of "Identification of complete information games." Many thanks go to Joel Horowitz, Charles Manski, and Elie Tamer for many helpful comments and discussions. Any errors are mine. Email: brendan.kline@austin.utexas.edu.

is not necessarily a unique solution to the model due to multiple equilibria, and therefore there can be multiple potential outcomes of the game for a fixed specification of the utility functions. Nevertheless, despite these and other challenges, Tamer (2003), Bajari, Hong, and Ryan (2010), Dunker, Hoderlein, and Kaido (2013), Fox and Lazzati (2013), and Kline (2013), provide various conditions under which parameters of interest in complete information game models are point identified.

Those prior identification results differ in many important ways. Nevertheless, all those results are based on identification strategies that require a regressor that has large support, although the exact "way" that such a regressor is used may arguably vary across papers.[1] As shown by those papers, and as shown by the literature on single-agent models involving "choice," exploiting the existence of large support regressors is an attractive identification strategy. Indeed, in some single-agent models involving "choice," the model might not be point identified without a large support regressor, if parametric distributional assumptions are not maintained about the unobservables (e.g., French and Taber (2011)). (This paper shows point identification without a large support regressor, and without parametric distributional assumptions.)

There are a few reasons for studying identification (and estimation) in complete information games without a large support regressor. First, and perhaps most practically, in some empirical applications it may not be the case that there is a regressor satisfying the relevant large support condition. If there are not identification results that apply without a large support regressor, then point identification of the models used in those empirical applications may be in question. Second, and perhaps more generally, in order to understand the empirical content of complete information games, it is useful to know whether or not the point identification of complete information games requires the existence of a large support regressor. Or, equivalently, it is useful to know if (in some sense) either: the *only* information about the parameters of a complete information game "comes from" the extreme values of the explanatory

[1] Bajari, Hong, and Ryan (2010) give another identification result using only exclusions restrictions, assuming a known distribution of the unobservables, and show that a certain necessary condition for identification is satisfied under this approach. Similarly, the seminal earlier work of Bresnahan and Reiss (1990, 1991a,b) does not contain any formal identification results, but is suggestive (based on for example the empirical exercises) of an identification strategy again based on parametric distributional assumptions for the unobservables. Aradillas-Lopez and Rosen (2013) study an ordered response game, primarily from the partial identification perspective, but do show that *some* model parameters are point identified without a large support regressor, assuming either that the unobservables have a known distribution or a certain parametric distribution (the Farlie-Gumbel-Morgenstern distribution). The current paper avoids such distributional assumptions.

variables that exist for a large support regressor, or if such extreme values of the explanatory variables are *necessary* conditions for point identification (yet with perhaps the "non-extreme" values of the explanatory variables still providing some information about the parameters). And third, a side effect of the constructive identification results in this paper is a new estimation strategy.

The existence of sources of information that do not require a large support regressor is suggested by the fact that Nash equilibrium implies some (possibly set-valued, but generally non-trivial) restrictions on the outcome of the game, as a mapping from the utility functions to the outcomes, for *any* instance of the game. This paper shows that such restrictions indeed imply the existence of sources of point identification that do not require the existence of a large support regressor.

This paper exploits just one such source of identification: the existence of "unique potential outcomes." A "unique potential outcome" is an outcome of the game that occurs exclusively as a *unique* Nash equilibrium of the game, for any *allowed* specification of the utility functions. Importantly, this "uniqueness" is subject to restrictions on the space of utility functions that are allowed, related to whether the game involves strategic complements or strategic substitutes. Therefore, this paper also is concerned with determining whether the game involves strategic complements or strategic substitutes, in a "first step" before introducing the notion of "unique potential outcomes."

This definition implies that if the econometrician uses only the observations associated with the "unique potential outcomes," then the econometrician does not need to be concerned with the existence of multiple equilibria, and therefore avoids many of the problems encountered by other identification strategies. (In some sense, these problems are simply "shifted" to the problem of testing for whether the game involves strategic complements or strategic substitutes, a problem that is also addressed in this paper.) Of course, this does not imply that identification (or estimation) is trivial, and indeed the identification results show that identification is not trivial. (And the resulting estimator is also non-standard.) Indeed, this strategy might seem to involve "throwing away" information, and might suggest even the loss of point identification. (Of course, identification results based on "large support" regressors might also "throw away" information.) Despite this reasonable concern, this paper shows that under certain conditions an identification strategy based on "unique potential outcomes" does indeed result in point identification of the parameters of the utility functions, even without a regressor with large support. Some broader implications are discussed in the conclusions, in particular highlighting the tradeoff between support assumptions on the explanatory variables and assumptions on the solution concept.

Of course, identification arguments are always complicated by the fact that the relationship between the explanatory variables and the outcomes depends also on the unobservables. This paper uses two main assumptions on the unobservables. First, it is assumed that the unobservables are independent of the exogenous explanatory variables. This is a standard assumption in the literature on identification of complete information games, and indeed in the literature on identification of single-agent discrete choice, although of course there are important exceptions. The second assumption is the location assumption, which is non-standard but plausible: the assumption that the mode of the joint distribution of the unobservables of all agents is zero. Note that this paper does not use parametric distributional assumptions on the unobservables, nor does this paper require that the unobservables are independent across agents within a market. The latter allows, for example, a market fixed effect in the context of an entry game. The reason for making the assumption on the mode rather than another location assumption is discussed in the identification analysis; it is a somewhat inevitable feature of an identification strategy in the absence of a large support regressor.

Finally, a three-step semiparametric estimator based on these identification results is proposed, and the asymptotic properties are derived. The three steps of the estimator are: estimate the direction of the interaction effect (i.e., the sign of the interaction effect parameter "sgn (Δ)"), estimate the slope parameters (i.e., the coefficients on the exogenous explanatory variables "β"), and then estimate the "intercept" parameters (i.e., the interaction effect parameters, and the usual intercepts "Δ and α"). The interaction effect parameter is an "intercept" parameter because it appears in the model likelihood as an "intercept" or "shift" parameter, not as a slope coefficient on some exogenous explanatory variable.[2]

The first step is to estimate the direction of the interaction effect, or equivalently to estimate the sign of the interaction effect parameter. This is related to testing whether the game involves strategic complements or strategic substitutes. If the direction of

[2] For example, $(1, 1)$ can be a Nash equilibrium outcome in pure strategies if and only if $\alpha_1 + x_1\beta_1 + \Delta_1 + \epsilon_1 \geq 0$ and $\alpha_2 + x_2\beta_2 + \Delta_2 + \epsilon_2 \geq 0$. Similarly, $(1, 0)$ can be a Nash equilibrium outcome in pure strategies if and only if $\alpha_1 + x_1\beta_1 + \epsilon_1 \geq 0$ and $\alpha_2 + x_2\beta_2 + \Delta_2 + \epsilon_2 \leq 0$. Similar statements are true for each outcome. (And of course because of multiple equilibria, the statement "can be a Nash equilibrium" is not equivalent to the statement "is the outcome.") Thus, the interaction effect parameter, Δ, essentially acts like a "shift" or "intercept" parameter.

the interaction effect is known *a priori* by the econometrician, then this step can be skipped. For example, in an entry game, economic theory implies that the interaction effect is non-positive, but implies perhaps nothing about the magnitude. Otherwise, since the subsequent steps of the estimator depend on the direction of the interaction effect, the sign of the interaction effect parameter must be identified and estimated. The estimator of the sign of the interaction effect parameter converges arbitrarily fast, so has no asymptotic effect on the subsequent steps.

The results on identification and estimation of the direction of the interaction effect may be of independent interest, because it provides a complete information alternative to the results of de Paula and Tang (2012) that were shown in the context of incomplete information games. It is worth noting that the identification strategy used in this paper is substantially different from the identification strategy used in de Paula and Tang (2012). In particular, this paper exploits the same exclusion restriction needed to point identify the rest of the parameters, whereas de Paula and Tang (2012) exploits the assumption that the unobservables (i.e., the signals in the incomplete information game) are independent across agents.

Then, the second step is to estimate the slope parameters, and the third step is to estimate the "intercept" parameters. The estimator for the slope parameters is related to density-weighted average derivate estimation, and is $\sqrt{M}$-consistent and asymptotically normally distributed, where M is the number of markets in the data (i.e., the number of "games played" in the data). The estimator for the "intercept" parameters involves maximizing the derivatives of an unknown "regression function" that is estimated by non-parametric methods, and is asymptotically normally distributed but converges at slower than the $\sqrt{M}$-rate. The rate of convergence depends on the assumed smoothness of the density of the unobservables, with more smoothness resulting in faster rates of convergence. In principle, the rate of convergence can approach the parametric rate under strong smoothness assumptions, but under more realistic assumptions is $M^{\frac{1}{4}}$ (or faster under more smoothness). The rate of convergence does not depend on the number of explanatory variables in the model, due to a dimension reduction strategy. The rate of convergence of the estimator of the intercept parameters is discussed in remark 5.1, where it is argued that slower than parametric rate of convergence is not surprising.

Despite the theoretically slower than parametric rate of convergence, the estimator seems to perform well in a Monte Carlo experiment in section 6.

1.2. **Outline of the paper.** Section 2 introduces the model and sets up the identification problem. Section 3 discusses identification under the assumption that the direction of the interaction effect is known *a priori* by the econometrician, or has been identified in a "first step" using the results of section 4. Section 4, which can be skipped if the direction of the interaction effect is indeed known *a priori* by the econometrician, discusses identification of the direction of the interaction effect. This can then be "plugged into" the results from section 3. Alternatively, section 4 may be of independent interest if identifying the direction of the interaction effect is the final goal of the analysis. Section 5 discusses estimation. Section 6 reports the results of a Monte Carlo experiment of the small sample performance of the estimator. Section 7 concludes, including a discussion of some additional identification results on the distribution of the unobservables and selection mechanism that are not in the main body of the paper.

2. Model

The complete information game is described in normal form in table 1. The row player is agent 1 and the column player is agent 2. The actions available to each agent are $S = \{0, 1\}$; for example, if this is a model of an entry game, the actions available to each agent are to enter the market (action 1) or to not enter the market (action 0). Table 1 gives the utility functions for the agents in market m, where a "market" is the unit of observation for the purposes of the identification and estimation arguments, as standard in the literature. The subscripting notation is: subscripted $1m$ refers to agent 1 in market m and subscripted $2m$ refers to agent 2 in market m, while a subscripted m refers to market m. The first payoff is the row payoff, and the second payoff is the column payoff.

	0	1
0	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ \alpha_2 + x_{2m}\beta_{2x} + w_m\beta_{2w} + \epsilon_{2m} \end{pmatrix}$
1	$\begin{pmatrix} \alpha_1 + x_{1m}\beta_{1x} + w_m\beta_{1w} + \epsilon_{1m} \\ 0 \end{pmatrix}$	$\begin{pmatrix} \alpha_1 + x_{1m}\beta_{1x} + w_m\beta_{1w} + \Delta_1 + \epsilon_{1m} \\ \alpha_2 + x_{2m}\beta_{2x} + w_m\beta_{2w} + \Delta_2 + \epsilon_{2m} \end{pmatrix}$

TABLE 1. Normal form

Equivalently, the utility functions are

$$u_{im}(0, y_{-im}) = 0 \text{ and } u_{im}(1, y_{-im}) = \alpha_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i y_{-im} + \epsilon_{im}$$

in market m. The normalization that $u_{im}(0, y_{-im}) = 0$ (or equivalent) is necessary for point identification, because best responses are invariant to additions to the utility functions of functions that depend only on the actions of the other agents. The solution concept is standard: pure strategy Nash equilibrium play.

The explanatory variables x_{im} are variables that are specific to agent im, and the explanatory variables w_m are variables that are shared among the agents in market m (for example, market characteristics). The notion of an "agent-specific" variable is somewhat vague at this point in the discussion: it is formalized by assumption 3.4 that requires (x_{1m}, x_{2m}, w_m) to have a density. It is allowed that there are no shared variables, but there must be at least one agent-specific variable per agent (i.e., neither of x_{1m} and x_{2m} can be void), because the agent-specific variables represent an exclusion restriction. This exclusion restriction is used to resolve the first difficulty in identification of complete information games that was described at the beginning of the paper: the fact that there is a simultaneity/endogeneity problem due to the fact that "outcomes" are also "explanatory variables." There are K_i agent-specific explanatory variables for agent i, and L shared explanatory variables.

The parameters have the usual interpretation. $\beta = (\beta_1, \beta_2)$, where $\beta_1 = (\beta_{1x}, \beta_{1w})$ and $\beta_2 = (\beta_{2x}, \beta_{2w})$, are the slope parameters, or equivalently the coefficients on the exogenous explanatory variables. $\alpha = (\alpha_1, \alpha_2)$ are the intercept parameters. Finally, $\Delta = (\Delta_1, \Delta_2)$ are the interaction effect parameters, which characterizes how utility depends on the action of the other agent.

As standard, it is assumed that the sign of the Δ parameters are weakly equal, in the sense that: either $\Delta_1 \leq 0$ and $\Delta_2 \leq 0$ (e.g., strategic substitutes, as in an entry game), or $\Delta_1 \geq 0$ and $\Delta_2 \geq 0$ (e.g., strategic complements, as in social interactions). The assumption that the sign of the Δ parameters are weakly equal is consistent with economic theory: it is equivalent to assuming that the game is either a game of strategic substitutes or a game of strategic complements. The assumption also guarantees the existence of a pure strategy Nash equilibrium. However, the direction of the interaction effect need not necessarily be known *a priori* by the econometrician, as the direction of the interaction effect is point identified in section 4.

In each market the econometrician observes the outcomes $y = (y_1, y_2)$ and the exogenous explanatory variables $z = (x_1, x_2, w)$, but does not observe $\epsilon = (\epsilon_1, \epsilon_2)$. The identification problem concerns uniquely recovering the parameters of the utility functions, $\theta = (\alpha_1, \alpha_2, \beta_{1x}, \beta_{1w}, \beta_{2x}, \beta_{2w}, \Delta_1, \Delta_2)$, from the population distribution of independent observations of the game, $P(y_1, y_2, x_1, x_2, w)$.

3. Identification

This identification strategy assumes that the direction of the interaction effect (i.e., the sign of Δ_1 and Δ_2) is known *a priori* by the econometrician, or has been identified in a "first step" using the results of section 4. If it is not known *a priori*, then the results of section 4 can be used to identify the direction of the interaction effect under almost exactly the same assumptions, and then the results of this section can be used. (One additional assumption is used to identify the direction of the interaction effect, implying that identification of the direction of the interaction effect cannot be proved as an intermediate step of the theorem in this section.)

3.1. **Sketch of identification strategy.**

The following sketches the identification strategy, which is formalized in section 3.2. This discussion is for the case where the econometrician knows (or has previously identified) that the interaction effect is non-positive, as in an entry game. (The case of a non-negative interaction effect is similar, but uses a different set of "unique potential outcomes.") It is important to note that some important technical details are necessarily omitted from this sketch, in order to more simply describe the basic identification strategy.

In the above game, when the interaction effect parameters are non-positive, $(0,0)$ is a "unique potential outcome" as $(0,0)$ is the Nash equilibrium if and only if $\epsilon_1 \leq -\alpha_1 - x_1\beta_{1x} - w\beta_{1w}$ and $\epsilon_2 \leq -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}$. Therefore, observing that the outcome is $(0,0)$ is *equivalent* to that condition on the unobservables. Therefore,

$$\begin{aligned} P(y = (0,0)|z) &= P(\epsilon_1 \leq -\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, \epsilon_2 \leq -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}) \\ &= P_0(-\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}), \end{aligned}$$

where $P_0(t_1, t_2) = P(\epsilon_1 \leq t_1, \epsilon_2 \leq t_2)$. Similarly,

$$\begin{aligned} P(y = (1,1)|z) &= P(\epsilon_1 \geq -\alpha_1 - x_1\beta_{1x} - w\beta_{1w} - \Delta_1, \epsilon_2 \geq -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2) \\ &= P_1(-\alpha_1 - x_1\beta_{1x} - w\beta_{1w} - \Delta_1, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2), \end{aligned}$$

where $P_1(t_1, t_2) = P(\epsilon_1 \geq t_1, \epsilon_2 \geq t_2)$. The above uses the assumption that the unobservables are independent of the explanatory variables, to justify dropping the conditioning of the distribution of ϵ by z, where as above: $z = (x_1, x_2, w)$ is the set of all exogenous explanatory variables.

Therefore, for explanatory variable k of agent i, $\frac{\partial P(y=(0,0)|z)}{\partial x_{ik}} = P_0^{(i)}(-\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w})(-\beta_{ixk})$, where $P_0^{(i)}(\cdot)$ indicates the evaluation of the derivative of $P_0(\cdot)$ with respect to its ith argument. And so,

$$\frac{\frac{\partial P(y=(0,0)|z)}{\partial x_{1k}}}{\frac{\partial P(y=(0,0)|z)}{\partial x_{ik'}}} = \frac{\beta_{ixk}}{\beta_{ixk'}},$$

so β_{ix} is point identified up to scale. (Note that this expression is also true taking expectations with respect to z, a fact that is used for estimation.) Also, a somewhat more technical, but ultimately similar, identification strategy shows that β_{1w} and β_{2w} are point identified up to the same scale normalization. The identification of β_{1w} and β_{2w} is more complicated because shared explanatory variables affect both players' utilities, but only the total effect of the explanatory variables is "identified" in this strategy, so some additional work is necessary to point identify the effect on each agent separately.

This identification strategy shows that β is point identified up to scale. However, the "intercept" parameters cannot be point identified using this strategy, because the "intercept" parameters are not the coefficients on explanatory variables. Identification of the "intercept" parameters is perhaps the most important part of the identification analysis, as the interaction effect parameter Δ is an "intercept" parameter. Identification (and/or estimation) for "intercept" parameters in econometric models is often much more difficult than for slope coefficients, and sometimes such parameters are "absorbed" into other parts of the model (see for example Andrews and Schafgans (1998), remark 5.1, and the below discussion of "absorbing" the intercept terms in this game model.) In contrast, it is important for this identification analysis to separately identify all "intercept" parameters.

Define $c_i = -x_i\beta_{ix} - w\beta_{iw}$, which is point identified given that β is point identified. Then, $P(y = (0,0)|c_1, c_2) = P_0(-\alpha_1 + c_1, -\alpha_2 + c_2)$ and $P(y = (1,1)|c_1, c_2) = P_1(-\alpha_1 - \Delta_1 + c_1, -\alpha_2 - \Delta_2 + c_2)$.

Then, notice that

$$\left.\frac{\partial^2 P(y = (0,0)|c_1, c_2)}{\partial c_1 \partial c_2}\right|_{(a_1,a_2)} = P_0^{(12)}(-\alpha_1 + a_1, -\alpha_2 + a_2),$$

where $P_0^{(12)}$ is the second cross partial derivative of P_0. The left hand side is observed, (a_1, a_2) is "known" as it is freely specified by the econometrician, and the right hand side is an unknown function of the parameters of interest, α_1 and α_2. Nevertheless, in general, it is impossible to use such an equation to identify the intercept parameters, since the unknown $P_0^{(12)}$ can "absorb" the intercept terms. That is because it could be that $P_0^{(12)}(-\alpha_1 + a_1, -\alpha_2 + a_2) = Q_0^{(12)}(-\alpha_1' + a_1, -\alpha_2' + a_2)$ for all (a_1, a_2) if $Q_0^{(12)}(t_1, t_2) \equiv P_0^{(12)}(t_1 - \alpha_1 + \alpha_1', t_2 - \alpha_2 + \alpha_2')$, implying that α_1 and α_2 would not be point identified, since $\{P_0, \alpha_1, \alpha_2\}$ would be observationally equivalent to $\{Q_0, \alpha_1', \alpha_2'\}$. However, under a certain location normalization on the unobservables, which translates to restrictions on $P_0^{(12)}$ that imply that it cannot "absorb" the intercept terms, α_1 and α_2 are point identified.

Specifically, notice that $P_0^{(12)}(t_1, t_2) = f_\epsilon(t_1, t_2)$, where f_ϵ is the density of the unobservables. Therefore, $\frac{\partial^2 P(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2}\Big|_{(a_1,a_2)}$ is the density of the unobservables evaluated at $(-\alpha_1 + a_1, -\alpha_2 + a_2)$. Still, since the density of the unobservables is unknown, additional assumptions are needed to point identify α_1 and α_2.

Suppose that the *mode* of the unobservable is $(0, 0)$. Then, $\frac{\partial^2 P(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2}\Big|_{(a_1,a_2)}$ is maximized as a function of (a_1, a_2) when $-\alpha_1 + a_1 = 0$ and $-\alpha_2 + a_2 = 0$, or equivalently when $a_1 = \alpha_1$ and $a_2 = \alpha_2$. Since $\frac{\partial^2 P(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2}\Big|_{(a_1,a_2)}$ is observed, this implies that indeed (α_1, α_2) is point identified. Similarly, $\frac{\partial^2 P(y=(1,1)|c_1,c_2)}{\partial c_1 \partial c_2}\Big|_{(b_1,b_2)}$ is maximized as a function of (b_1, b_2) when $b_1 = \alpha_1 + \Delta_1$ and $b_2 = \alpha_2 + \Delta_2$, so (Δ_1, Δ_2) is point identified. Notice that the corresponding estimator involves maximizing the derivatives of a regression function, estimated on "generated regressors," which is non-standard. See the estimation results in section 5.

3.2. **Formal identification results.** The following assumptions are sufficient for point identification of the parameters of the utility function.

The first assumption is a scale normalization. As in single-agent discrete choice models, the scale of the utility functions has no observable implications. See remark 3.5 on the fact that this assumption also is a sign assumption. (In short, even if the sign is not known *a priori* by the econometrician, the sign is point identified by the same identification strategy, and so to avoid distracting accounting details related to keeping track of the sign, it is assumed the signs are positive. If not, then the signs of x_{11} and/or x_{21} can be "flipped" by multiplying them by -1.)

Assumption 3.1 (Scale normalization)**.** $\beta_{1x1} = 1 = \beta_{2x1}$

The first substantiative assumptions primarily concern the unobservables.

Assumption 3.2 (Independence of the unobservables from the explanatory variables)**.** $\epsilon \perp z$

Recall that $z_m = (x_{1m}, x_{2m}, w_m)$ are the explanatory variables in market m. Assumption 3.2 allows that ϵ_{1m} is correlated with ϵ_{2m}. The assumption that the unobservables are independent of the explanatory variables is standard in the literature on identification of complete information games, and indeed in the literature on identification of single-agent discrete choice, although of course there are important exceptions.

Assumption 3.3 (Unobservables have mode at zero). *The distribution of ϵ admits an ordinary density $f_{\epsilon_1,\epsilon_2}(\cdot)$ with respect to Lebesgue measure, such that $f_{\epsilon_1,\epsilon_2}(\cdot)$ is continuous and positive everywhere. The unique mode of ϵ is $(0,0)$, defined by $(0,0) = \arg\max_{t_1,t_2} f_{\epsilon_1,\epsilon_2}(t_1,t_2)$.*

Particularly the part of assumption 3.3 concerned with the mode of ϵ deals with the fact that a necessary condition for point identification of α and Δ is an assumption on the location of ϵ. As with other location assumptions, the condition that the mode is at $(0,0)$ versus at some other point is a normalization, as the "true" mode is "absorbed" by α. Rather, the substantive condition is that the density of ϵ is maximized at a unique point, a condition satisfied by almost every distribution.

The location assumption is somewhat non-standard, as typically the location assumption concerns either the mean or median of the unobservables. However, $E(\epsilon) = 0$ and $\epsilon \perp z$ is not sufficient for point identification because there is no regressor with large support.[3] An assumption on the median, in the sense of $Med(\epsilon) = \{Med(\epsilon_i)\}_i$, was shown to be compatible with point identification in Kline (2013) with a large support regressor (and in fact allowing ϵ to not be independent from z), but without a large support regressor it seems difficult (although not necessarily impossible) to

[3] Manski (1988) has shown that mean independence is not sufficient for point identification in a single-agent discrete choice model, but does not consider the additional assumption that $\epsilon \perp z$. Indeed, Lewbel (2000) shows in single-agent discrete choice models that if there is a regressor with large support, then the addition of stochastic independence assumptions is sufficient for point identification. However, if there is not a regressor with large support, then the intercept is not point identified even if $E(\epsilon) = 0$ and $\epsilon \perp z$. This is discussed by Magnac and Maurin (2007) and Khan and Tamer (2010) for a single-agent discrete choice model, but the arguments extend to the case of a complete information game. Essentially, for any outcome y of the game, there is a set of ϵ, called $\mathcal{E}_y$, such that for any realization of $\epsilon \in \mathcal{E}_y$ and any realization of z, the outcome of the game is y. For example, for $y = (1,1)$ the set of $\mathcal{E}_y$ are ϵ such that ϵ_1 and ϵ_2 are both very large. By rearranging probability mass of ϵ within this region of ϵ-space it is possible to construct observationally equivalent models that have different locations parameters α, because the mean of the constructed distribution of ϵ can take arbitrarily large or small values since the mean functional is infinitely sensitive to sufficiently large outliers even if they have only very small probability.

achieve point identification using this assumption.[4] Therefore, the location assumption is made on the mode of the unobservables.

The identifying power of an assumption on the mode of ϵ compared to an assumption on the mean or median is due to the fact that the condition that $(0,0)$ is the mode of ϵ is a "local" property. Solving the problem of $\arg\max_{t_1,t_2} f_{\epsilon_1+c_1,\epsilon_2+c_2}(t_1,t_2)$, which is the problem of finding the mode of $(\epsilon_1+c_1,\epsilon_2+c_2)$ in order to identify (c_1,c_2), requires only that $f_{\epsilon_1+c_1,\epsilon_2+c_2}(\cdot)$ is known in a neighborhood of the mode. For example, suppose that the density of $(\epsilon_1+\alpha_1,\epsilon_2+\alpha_2)$ is point identified on some set $\mathcal{E}$. This is established in the identification result; however, it will generically be that $\mathcal{E} \subsetneq \mathbb{R}^2$ because there is no regressor with "large support," and the density of $(\epsilon_1+\alpha_1,\epsilon_2+\alpha_2)$ is point identified relative to the support of exogenous explanatory variables. Nevertheless, as long as $\mathcal{E}$ contains (α_1,α_2), the density of $(\epsilon_1+\alpha_1,\epsilon_2+\alpha_2)$ is point identified at its mode, so the mode of $(\epsilon_1+\alpha_1,\epsilon_2+\alpha_2)$ can be identified as the point in $\mathcal{E}$ with highest density. The mode is located at (α_1,α_2), which implies that α_1 and α_2 are point identified. The interaction effect parameters Δ_1 and Δ_2 are point identified similarly.

This identification strategy requires the following assumptions on the variation in the exogenous explanatory variables.

Assumption 3.4 (Continuous explanatory variables)**.** *The distribution of $z=(x_1,x_2,w)$ admits an ordinary density with respect to Lebesgue measure that is positive on some convex set with non-empty interior.*

This assumption is understood to apply to the variables that *actually exist* when there is not, for example, any shared variables. The assumption that z has an ordinary density implicitly reflects an exclusion restriction, as it implies that there is an element of x_1 that does not appear in x_2 or w, and similarly that there is an element of x_2 that does not appear in x_1 or w. The requirement of continuous explanatory variables (with a density on a convex set) arises because the identification strategy uses derivatives with respect to the explanatory variables.

[4]The difficulty is that identification of the joint cumulative distribution function of (ϵ_1,ϵ_2) on a set of points $\mathcal{E}$ does not necessarily imply identification of the marginal cumulative distribution function of ϵ_1 or ϵ_2 at any point, and assumptions about the median concern the marginal distributions while the identification strategy shows identification of the joint cumulative distribution function on a certain set of points.

Assumption 3.5 (Sufficient variation of the explanatory variables)**.** *The density of* $(-x_1\beta_{1x} - w\beta_{1w}, -x_2\beta_{2x} - w\beta_{2w})$ *exists and is strictly positive on a convex open set that contains* (α_1, α_2) *and* $(\alpha_1 + \Delta_1, \alpha_2 + \Delta_2)$.

This assumption requires that the densities of $\epsilon + \alpha$ and $\epsilon + \alpha + \Delta$ are "observed" at their modes, where "observed" is meant in the sense of the identification strategy above, which results in identifying α and $\alpha + \Delta$ since the mode of ϵ is zero. This assumption can have straightforward observable implications (under an additional assumption on the unobservables), as discussed in remark 3.4. Qualitatively similar assumptions have been used before in other contexts. For example Horowitz (2009, Corollary 4.1) uses a similar assumption to identify a binary choice model with median restrictions.[5]

Assumption 3.6 (Full rank marginal effects of the excluded agent-specific observable)**.** *If there is a shared explanatory variable (i.e., w is not void), either:*

(1) $\beta_{1w} = \beta_w = \beta_{2w}$, *or*
(2) $E(P_x(z))$ *exists and has full rank, where* $P_x(z)$ *is the* 2×2 *matrix with* $P_{x,ij}(z) = \frac{\partial P(y=(0,0)|z)}{\partial x_{i1}} \frac{\partial P(y=(0,0)|z)}{\partial x_{j1}}$.

If there is not a shared explanatory variable, then there is no assumption.

If the shared explanatory variables affect the utility of each agent equally, then this assumption is immediately satisfied. In particular, this would be implied by the assumption of "exchangeable" agents, with $\beta_1 = \beta = \beta_2$, which often arises in models where the "labeling" of the players has no real economic content. Otherwise, in order to point identify the different effects on each agent of the shared explanatory variable(s), the identification strategy requires that the marginal effects of one unit increases in the utility of each agent, on average, have full rank effects on the outcomes. Due to the normalization that $\beta_{ix1} = 1$ for all agents i, this is equivalent to requiring that $E(P_x(z))$ has full rank. This assumption is used because a shared explanatory variable necessarily has an effect on the utility of all agents, and only the total effect on outcomes is "observed" in the data, so some additional work is necessary to point identify the effect on each agent separately. Note that this is an assumption on the observables, so has observable implications.

[5]Essentially the assumption in Horowitz (2009) is that $\alpha + x\beta$ conditional on x_{-1} (all but the first component of x) has positive density on some interval $[-\delta, \delta]$, and that is equivalent to $-x\beta$ conditional on x_{-1} having positive density on $[\alpha - \delta, \alpha + \delta]$, or equivalently $-x\beta$ conditional on x_{-1} having density in a neighborhood of the "intercept" term α.

The following theorem gives an identification result for a non-positive interaction effect, as in an entry game. The case of a non-negative interaction effect can be addressed similarly, essentially studying the outcomes $(0,1)$ and $(1,0)$, and adjusting the statement of the theorem appropriately. Recall again that section 4 shows how to point identify the sign of the interaction effect parameter if it is not known *a priori* by the econometrician.

Also, the following theorem allows a certain "weight function" $\pi(\cdot)$ in the expectations that characterize the parameters in terms of population quantities; for estimation purposes in section 5, $\pi(\cdot)$ is taken to be the density of z.

Theorem 3.1 (Identification)**.** *Suppose that $\Delta_1 \leq 0$ and $\Delta_2 \leq 0$. Under assumptions 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6, $\theta = (\alpha_1, \alpha_2, \beta_{1x}, \beta_{1w}, \beta_{2x}, \beta_{2w}, \Delta_1, \Delta_2)$ is point identified, and can be expressed in terms of observables as follows.*

Let $\pi(\cdot)$ be any non-negative function of z, that is strictly positive on a set of z that has positive probability under the data generating process for z. Then, assuming that the expectations in these expressions exist (see remark 3.1): for any i and $k \in \{1, 2, \ldots, K_i\}$,

$$\beta_{ixk} = \frac{E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial x_{ik}}\right)}{E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial x_{i1}}\right)}.$$

If $\beta_{1w} = \beta_{2w}$, then for any $l \in \{1, 2, \ldots, L\}$,

$$\beta_{wl} = \frac{E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial w_l}\right)}{\sum_i^2 E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial x_{i1}}\right)}.$$

Alternatively, if $E(P_x(z))$ has full rank, then for any $l \in \{1, 2, \ldots, L\}$,

$$\begin{aligned}\beta_{wl} = (\beta_{1wl}, \beta_{2wl}) &= \left(E\left(P_x(z)\right)\right)^{-1} E\left(P^{[1]}(z)'\frac{dP(y=(0,0)|z)}{dw_l}\right)\\ &= \left(E\left(P^{[1]}(z)'P^{[1]}(z)\right)\right)^{-1} E\left(P^{[1]}(z)'\frac{dP(y=(0,0)|z)}{dw_l}\right)\end{aligned}$$

where $P^{[1]}(z)$ is the 1×2 matrix whose ith entry is $\frac{\partial P(y=(0,0)|z)}{\partial x_{i1}}$.

Further, set $c_i \equiv -x_i\beta_{ix} - w\beta_{iw}$, which is point identified by the above. Then

$$(\alpha_1, \alpha_2) = \arg\max_{a_1,a_2} \left.\frac{\partial^2 P(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2}\right|_{a_1,a_2}$$

and

$$(\Delta_1, \Delta_2) = \arg\max_{b_1,b_2} \left.\frac{\partial^2 P(y=(1,1)|c_1,c_2)}{\partial c_1 \partial c_2}\right|_{b_1,b_2} - \arg\max_{a_1,a_2} \left.\frac{\partial^2 P(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2}\right|_{a_1,a_2},$$

where the maximization is over the support of (c_1, c_2)*.*

Remark 3.1 (Existence of the expectations)**.** Theorem 3.1 requires that the expectations $E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial x_{ik}}\right)$, $E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial w_l}\right)$, $E(P_x(z))$, and $E\left(P^{[1]}(z)'\frac{dP(y=(0,0)|z)}{dw_l}\right)$ exist for all i, k, and l. This essentially reflects a mild regularity condition on the data generating process. A sufficient condition is that the densities of ϵ_1 and ϵ_2 are bounded above (which is implicitly implied by the mode assumption), and that $\pi(\cdot)$ is integrable with respect to the data generating process for z. See lemma 9.1, stated after the proof of theorem 3.1, for the details.

Remark 3.2 (Equilibrium existence)**.** Equilibrium existence is addressed (and shown to always exist under these assumptions) in the context of a more general model in remark 8.3.

Remark 3.3 (Extensions)**.** It is possible to extend this result in many directions. For example, it can be extended to cases of "unique potential outcomes" with $N > 2$. The notation becomes more cumbersome, but the basic idea from section 3.1 is unchanged. Specifically, the unique potential outcomes become $(0, \ldots, 0)$ and $(1, \ldots, 1)$. Also, for the purposes of identifying β, analogous results apply to the outcome $(1, 1)$ in place of the outcome $(0, 0)$, increasing efficiency of estimation. And, as noted previously, it can be extended to cases with a non-negative interaction effect by studying different outcomes. Finally, the expression for β_{wl} is also valid with another weight function $\pi(\cdot)$ inside all of the expectations, as long $E(\pi(z)P_x(z))$ has full rank. (The details of this claim are obvious from the proof.)

Remark 3.4 (Conditions on the support of the explanatory variables)**.** Assumption 3.5 has observable implications under the maintained assumption that the density of ϵ is suitably "unimodal." (The definition of a "unimodal" density is not standardized, as sometimes it means achieving the maximal density at a unique point, but allowing local maxima, while often it means having only one local (and therefore global) maxima.) By "unimodal," this remark means that the density of ϵ has an everywhere negative definite Hessian matrix, and has a unique global maximum and no other local maxima. Therefore, implicitly it is assumed in this remark that the density of ϵ is twice continuously differentiable.

The identification result in theorem 3.1 (together with the proof) shows that $\alpha = (\alpha_1, \alpha_2)$ and $\alpha + \Delta = (\alpha_1 + \Delta_1, \alpha_2, \Delta_2)$ are point identified as follows:

$$(\alpha_1, \alpha_2) = \arg\max_{a_1, a_2} \left. \frac{\partial^2 P(y = (0,0)|c_1, c_2)}{\partial c_1 \partial c_2} \right|_{a_1, a_2} = \arg\max_{a_1, a_2} f_\epsilon(-\alpha_1 + a_1, -\alpha_2 + a_2).$$

$$(\alpha_1 + \Delta_1, \alpha_2 + \Delta_2) = \arg\max_{b_1, b_2} \left. \frac{\partial^2 P(y = (1,1)|c_1, c_2)}{\partial c_1 \partial c_2} \right|_{b_1, b_2} = \arg\max_{b_1, b_2} f_\epsilon(-\alpha_1 - \Delta_1 + b_1, -\alpha_2 - \Delta_2 + b_2)$$

So, by definition, under the assumption that the density of ϵ is "unimodal," the first order conditions (e.g., $\left.\frac{\partial^3 P(y=(0,0)|c_1,c_2)}{\partial^2 c_1 \partial c_2}\right|_{a_1,a_2} = 0$) to these maximization problems will be *uniquely* satisfied when evaluated at α and $\alpha + \Delta$ respectively. In contrast, the first order conditions when evaluated at any other parameter specification will not be satisfied. (Any other place where the first order conditions would be satisfied would be a local maximum, since the Hessian is negative definite everywhere, but there is only a unique global maximum.)

Further, note that if assumption 3.5 is satisfied, then the parameters defined by the above maximization problems over the support of (c_1, c_2) will indeed equal α and $\alpha + \Delta$. In contrast, if assumption 3.5 is not satisfied, then since the maximization is by construction over the support of (c_1, c_2), the parameters defined by the above maximization cannot equal α and/or $\alpha + \Delta$, depending on what part of the assumption fails. So, together with the above discussion of first order conditions, this means that assumption 3.5 has the observable implication that the parameters defined by the above maximization problems uniquely satisfy the first order conditions of the maximization problem.

Note that similar arguments would apply if a known, increasing, transformation of the density has an everywhere negative definite Hessian matrix (e.g., log-"strictly" concave densities).

Remark 3.5 (Scale normalization)**.** Assumption 3.1 on scale normalization also implies a sign assumption, in order to avoid distracting accounting details related to the sign. Note that the sign can be identified by the same identification strategy, because the proof of identification shows that $\frac{\partial P(y=(0,0)|z)}{\partial x_{i1}} = F_\epsilon^{[i,1]}(-\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w})(-\beta_{ix1})$, so the sign of β_{ix1} is point identified, since $F_\epsilon^{[i,1]} > 0$. So by "flipping" the sign of x_{11} and x_{21} appropriately, by multiplying by -1, the sign assumption is without loss of generality.

4. Identification of the direction of the interaction effect

This section provides an identification strategy that point identifies the direction of the interaction effect. Although the results are stated in the context of the model studied in the rest of this paper, in fact the identification strategy is valid in more general specifications, as will be obvious.

4.1. **Sketch of the identification strategy.** The following sketches the identification strategy, which is formalized in section 4.2.

Consider two specifications of the observables $z = (x_1, x_2, w) = (x_{11}, x_{1(-1)}, x_2, w)$. In the first specification: $z' = (x'_{11}, x^*_{1(-1)}, x^*_2, w^*)$. In the second specification: $z'' = (x''_{11}, x^*_{1(-1)}, x^*_2, w^*)$, where $x'_{11} < x''_{11}$. So, z' and z'' are equal, except that z'' has a larger value for x_{11}, the first agent-specific explanatory variable of agent 1.

Since $\beta_{11} = 1 > 0$, for any given realization of the unobservables, agent 1 has greater utility from taking action 1 when the observables are z'' compared to when the observables are z'. Intuitively, this should imply that the probability that agent 1 takes action 1 is greater when the observables are z'' compared to when the observables are z'. This would unambiguously be true in a single-agent model, but because of the interaction it is not necessarily true in a game without certain (reasonable) assumptions on equilibrium selection. (See below for that condition.)

If that does happen, if there is a negative interaction effect, agent 2 will tend to get lower utility from taking action 1 when the observables are z'' compared to when the observables are z', because agent 1 is more likely to take action 1 when the observables are z'' compared to when the observables are z', which decreases the utility agents 2 gets from taking action 1. And so, agent 2 should be less likely to take action 1 when the observables are z'' compared to when the observables are z'. Similarly, if there is a positive interaction effect, agent 2 should be more likely to take action 1 when the observables are z'' compared to when the observables are z'. Thus, the marginal effect of x_{11} on the probability that $y_2 = 1$ should (under reasonable conditions) be equal to the sign of Δ_2, thereby point identifying the sign of Δ_2. And similarly, the marginal effect of x_{21} on the probability that $y_1 = 1$ should (under reasonable conditions) be equal to the sign of Δ_1.

The identification strategy therefore concerns finding conditions under which indeed $\frac{\partial P(y_2=1|z)}{\partial x_{11}}$ has the same sign as Δ_2 (and similarly for $\frac{\partial P(y_1=1|z)}{\partial x_{21}}$ and Δ_1). It turns out that a sufficient condition is assumption 4.1, which intuitively says (from the perspective of identifying Δ_2): conditional on z and the unobservables being such that there are multiple equilibria, the probability that an equilibrium with $y_1 = 1$ is selected is indeed weakly increasing with x_{11}, exactly as suggested in the sketch of the identification strategy. Essentially, as long as there is a positive marginal effect of x_{11} on the probability that $y_1 = 1$, the sign of Δ_2 is equal to the sign of the marginal effect of x_{11} on the probability that $y_2 = 1$.

4.2. **Formal identification results.** Let $\mathcal{R}^-(z,\theta) = \{(\epsilon_1, \epsilon_2) : -\alpha_1 - x_{1m}\beta_{1x} - w_m\beta_{1w} \leq \epsilon_{1m} \leq -\alpha_1 - x_{1m}\beta_{1x} - w_m\beta_{1w} - \Delta_1, -\alpha_2 - x_{2m}\beta_{2x} - w_m\beta_{2w} \leq \epsilon_{2m} \leq -\alpha_2 - x_{2m}\beta_{2x} - w_m\beta_{2w} - \Delta_2\}$ be the set of ϵ such that, for that specification of z and θ, the game with a non-positive interaction effect has multiple equilibria. Similarly, let $\mathcal{R}^+(z,\theta) = \{(\epsilon_1, \epsilon_2) : -\alpha_1 - x_{1m}\beta_{1x} - w_m\beta_{1w} - \Delta_1 \leq \epsilon_{1m} \leq -\alpha_1 - x_{1m}\beta_{1x} - w_m\beta_{1w}, -\alpha_2 - x_{2m}\beta_{2x} - w_m\beta_{2w} - \Delta_2 \leq \epsilon_{2m} \leq -\alpha_2 - x_{2m}\beta_{2x} - w_m\beta_{2w}\}$ be the set of ϵ such that, for that specification of z and θ, the game with a non-negative interaction effect has multiple equilibria. (These are the well known "boxes" of multiple equilibria.) Then, let

$$\mathcal{R}(z,\theta) = \begin{cases} \mathcal{R}^-(z,\theta) & \text{if } \Delta_1 \leq 0 \text{ and } \Delta_2 \leq 0 \\ \mathcal{R}^+(z,\theta) & \text{if } \Delta_1 \geq 0 \text{ and } \Delta_2 \geq 0 \end{cases}$$

be the overall set of ϵ such that there are multiple equilibria.

The following assumption is added to those in the previous section in order to point identify the direction of the interaction effect. It is an assumption on the properties of the selection mechanism: either $P(y_1 = 1|\epsilon \in \mathcal{R}^+(z,\theta), z)$ in the case of a non-negative interaction effect or $P(y_1 = 1|\epsilon \in \mathcal{R}^-(z,\theta), z)$ in the case of a non-positive interaction effect. (So, the assumption can be stated on $P(y_1 = 1|\epsilon \in \mathcal{R}(z,\theta), z)$.) These give the probability that an equilibrium with agent 1 taking action 1 is selected, conditional on a certain value of z and the unobservables being such that there are multiple equilibria; therefore, these are properties of the equilibrium selection mechanisms. Less formally, these concern the outcomes "inside" the "boxes" of multiple equilibria.

Assumption 4.1 (Monotonic selection mechanism)**.** *The selection mechanisms are weakly increasing and differentiable, in the sense that $P(y_1 = 1|\epsilon \in \mathcal{R}(z,\theta), z)$ and $P(y_2 = 1|\epsilon \in \mathcal{R}(z,\theta), z)$ are differentiable functions of z, with the former weakly increasing with x_{11} and the latter weakly increasing with x_{21}.*

In particular, this assumption is satisfied if the selection mechanism randomizes over the multiple equilibria according to a fixed probability: if $P(y_i = 1|\epsilon \in$

$\mathcal{R}(z,\theta), z) = \lambda_i$ (i.e., whenever there are multiple equilibria, there is λ_i probability that the outcome is $y_i = 1$), then obviously the assumption is satisfied. Further, the assumption is satisfied even if the selection mechanism depends on z but not x_{11} and x_{21}: if $P(y_i = 1|\epsilon \in \mathcal{R}(z,\theta), z) = \lambda_i(z)$ where $\lambda_i(\cdot)$ does not depend on x_{11} or x_{21}, then obviously the assumption is satisfied. Thus, exclusions restrictions are sufficient to imply assumption 4.1.

But even without an exclusion restriction, it is enough simply that the selection mechanism is monotone, in the intuitive sense described above: the probability that $y_i = 1$ should suitably increase with x_{i1}. Considering that the utility of agent 1 increases with x_{11}, assumption 4.1 formalizes the intuition that the marginal effect of x_{11} on the selection mechanism should be that the selection mechanism chooses the equilibrium with $y_1 = 1$ with weakly increasing probability as a function of x_{11}. (And similarly for agent 2.) An alternative but qualitatively similar assumption is given in remark 4.2 for a slightly different identification strategy. That version of the assumption says that if a given outcome y^* of the game is more likely *to satisfy the conditions for being a Nash equilibrium* when the observables are z'' compared to when the observables are z', then also that outcome y^* of the game is also more likely *to actually be the outcome of the game* when the observables are z'' compared to when the observables are z'. So, again, the selection mechanism satisfies a monotonicity condition.

Essentially the only "counterexample" to assumption 4.1 is a selection mechanism of the following sort: either $P(y_1 = 1|\epsilon \in \mathcal{R}(z,\theta), z) = \lambda_1(z)$ where $\lambda_1(z)$ is a *decreasing* function of x_{11}, or $P(y_2 = 1|\epsilon \in \mathcal{R}(z,\theta), z) = \lambda_2(z)$ where $\lambda_2(z)$ is a *decreasing* function of x_{21}. These sorts of selection mechanisms are somewhat pathological, for the following reason (focusing on agent 1, although similar arguments apply to agent 2). Note that a marginal increase in x_{11} results in an increased utility from taking action 1 for agent 1, for any value of ϵ and any action of the opponent. Therefore, these "counterexample" selection mechanisms are selection mechanisms that have the property that: even though the utility agent 1 gets from taking action 1 increases with x_{11} (and there is no direct effect on agent 2), nevertheless the selection mechanism is such that *in the region of multiple equilibria*, the equilibrium with $y_1 = 1$ is less likely after a marginal increase in x_{11} (and therefore less likely after a marginal increase in the utility that agent 1 gets from taking action 1). Although Nash equilibrium alone does not rule out such a response to a marginal increase in x_{11} (since Nash equilibrium is a "pointwise" concept, as it separately describes the Nash equilibrium outcomes for each given specification of the utility functions), it seems a somewhat pathological response when comparing equilibrium outcomes across specifications of the utility functions.

In the following theorem, let sgn $(\cdot)$ be the function that returns 1 if the argument is positive, 0 if the argument is zero, and -1 if the argument is negative.

Theorem 4.1 (Identification of the sign of the interaction effect)**.** *Under assumptions 3.1, 3.2, 3.3, 3.4, and 4.1, the sign of the interaction effect parameter is point identified:*

$$\text{sgn } (\Delta_1) = \text{sgn } E\left(\pi(z)\frac{\partial P(y_1 = 1|z)}{\partial x_{21}}\right)$$

and

$$sign(\Delta_2) = \text{sgn } E\left(\pi(z)\frac{\partial P(y_2 = 1|z)}{\partial x_{11}}\right)$$

for any weight function $\pi(\cdot)$ of z satisfying the same property as in theorem 3.1, also with the same remark about the existence of the expectations.

Remark 4.1 (Comparison to de Paula and Tang (2012))**.** Note that the identification result in de Paula and Tang (2012) for the direction of the interaction effect in incomplete information games relies on the assumption that the unobservables (i.e., the signals in the incomplete information game) are independent across agents, which, although completely standard in the literature on incomplete information games, has a different meaning in complete information games that would evidently rule out unobservable market fixed effects in complete information games. The identification result in theorem 4.1 does not use such an independence assumption, and instead relies on the exclusion restrictions entailed by the existence of agent-specific explanatory variables. Nevertheless, the appendix in section 8 shows how the de Paula and Tang (2012) identification strategy does carry over into the complete information game framework. An interesting feature of that result is that it shows that their "test statistic" is valid for both complete and incomplete information games.

Remark 4.2 (Extensions)**.** The same strategy works for $N > 2$, but with a slightly different assumption in place of assumption 4.1. This not only provides a result for $N > 2$, but also provides a slightly different perspective on the "meaning" of assumption 4.1. First, extend the notation and assumptions other than assumption 4.1 in the obvious way, with the utility functions being

$$u_i(0, y_{-i}) = 0 \text{ and } u_i(1, y_{-i}) = \alpha_i + x_i\beta_{ix} + w\beta_{iw} + \Delta_i(y_{-i}) + \epsilon_i,$$

where $\Delta_i(y_{-i})$ is a function of the actions of the agents other than i, that gives the effect of those actions on the utility of agent i.

Then, let

$$\begin{aligned}
P_{(-1)}(z) &\equiv P(y_2 = 1, \ldots, y_N = 1 \text{ is part of an equilibrium}|z) \\
&= P(y_1 = 1, \ldots, y_N = 1 \text{ is an equilibrium}|z) \\
&+ P(y_1 = 0, y_2 = 1, \ldots, y_N = 1 \text{ is an equilibrium}|z) \\
&= P(\epsilon_i \geq -\alpha_1 - x_i\beta_{ix} - w\beta_{iw} - \Delta_i(1, \ldots, 1)(\forall i)) \\
&+ P(\epsilon_1 \leq -\alpha_1 - x_1\beta_{1x} - w\beta_{1w} - \Delta_1(1, \ldots, 1), \\
&\qquad \epsilon_i \geq -\alpha_i - x_i\beta_{ix} - w\beta_{iw} - \Delta_i(0, 1, \ldots, 1)(\forall i \neq 1)).
\end{aligned}$$

Note that these probabilities are not observed in the data, because they correspond to the probability that certain outcomes satisfy the conditions of being a Nash equilibrium, not the probability that these outcomes are actually realized in the data. The probability that an outcome is a Nash equilibrium is generally strictly greater than the probability that outcome occurs in the data, because of multiple equilibria.

Then, letting $c_i = -\alpha_i - x_i\beta_{ix} - w\beta_{iw}$,

$$\begin{aligned}
\frac{\partial P_{(-1)}(z)}{\partial x_{11}} = \beta_{1x1} \Bigg(&\int_{c_2 - \Delta_2(1,\ldots,1))}^{\infty} \cdots \int_{c_N - \Delta_N(1,\ldots,1)}^{\infty} f_{\epsilon_1,\ldots,\epsilon_N}(c_1 - \Delta_1(1, \ldots, 1), e_2, \ldots, e_N) de_N \cdots de_2 \\
&- \int_{c_2 - \Delta_2(0,1,\ldots,1)}^{\infty} \cdots \int_{c_N - \Delta_N(0,1,\ldots,1)}^{\infty} f_{\epsilon_1,\ldots,\epsilon_N}(c_1 - \Delta_1(1, \ldots, 1), e_2, \ldots, e_N) de_N \cdots de_2 \Bigg).
\end{aligned}$$

So, since $\beta_{1x1} = 1$, $\operatorname{sgn} \frac{\partial P_{(-1)}(z)}{\partial x_{11}}$ is equal to the sign of the difference of integrals. By inspection of the regions of integration, if there is a negative interaction effect, in the sense of implying that $\Delta_i(0, 1, \ldots, 1) > \Delta_i(1, \ldots, 1)$ for all i, then the difference is negative; conversely, if there is a positive interaction effect, then the difference is positive. If there is a "monotonic" selection mechanism that satisfies $\operatorname{sgn} \frac{\partial P(y_2=1,\ldots,y_N=1 \text{ is part of an equilibrium}|z)}{x_{11}} = \operatorname{sgn} \frac{\partial P(y_2=1,\ldots,y_N=1|z)}{x_{11}}$ (cf., discussion of assumption 4.1), then the sign of the interaction effect parameters is identified by $\operatorname{sgn} E\left(\pi(z) \frac{\partial P(y_2=1,\ldots,y_N=1|z)}{\partial x_{11}}\right)$, similarly to theorem 4.1.

5. Estimation

This section provides a three-step semiparametric estimator corresponding to the identification results. Apparently, these are the first semiparametric estimators available for complete information games. The focus is on the second- and third-step estimators corresponding to estimating β, α, and Δ according to theorem 3.1. The first-step estimator corresponding to estimating the sign of Δ according to theorem 4.1 can be easily constructed combining the main ideas of the estimators corresponding to the first part of theorem 3.1 (estimating β) with theorem 8.1, and so is omitted.[6] Since the first-step estimator converges arbitrarily fast (as it amounts to estimating the sign of a parameter, or similarly it amounts to consistent model selection), it has no asymptotic effect on subsequent steps of estimation.

The proposed estimator is essentially based on the "analogy principle," given the constructive identification results in theorem 3.1. The estimator for β is standard and uses existing estimation results for semiparametric models, as the identification result shows that β can be expressed as functions of certain weighted average derivatives. (But, of course, the fact that such an estimation strategy is even sensible is the contribution of the identification result in this paper.) However, the estimator for α and Δ, which arguably concerns the most important parameters of the model, is not standard, and does not entirely use existing estimation results: rather, it involves estimating unknown regression functions (with *generated* regressors that are based on the estimator for β), and then maximizing those estimated regression functions. As a result, it has slower than parametric rate of convergence.

The estimation strategy for α and Δ is unique to the identification strategy based on the mode assumption on the unobservables, and so it does not seem possible to directly use "existing estimation results" to estimate α and Δ. As noted in section 3.1, identification (and estimation) of the "intercept" parameters in many semiparametric models is difficult. Indeed, many strategies, including those based on "average derivatives" and other approaches based on index models, tend to "absorb" the intercept into parts of the model that are typically not proved to be identified (and/or are

[6] In short, the idea is that the estimator of $E\left(\pi(z)\frac{\partial P(y_2=1|z)}{\partial x_{11}}\right)$ in theorem 4.1 is similar to the estimator of β in theorem 3.1, and the sign of $E\left(\pi(z)\frac{\partial P(y_2=1|z)}{\partial x_{11}}\right)$ can then be estimated as in estimating the sign appearing in theorem 8.1, as appears in the estimation result in theorem 8.2. Basically, the estimator of the sign uses the fact that an estimator for the sign of θ that is estimated by $\hat{\theta}$ with $\hat{\theta}-\theta = O_p(M^{-\frac{1}{2}})$ is sgn $\left(1\left\{M^{\frac{1}{4}}|\hat{\theta}| \geq 1\right\}\hat{\theta}\right)$, which converges arbitrarily fast.

not estimated). For example, in the typical approach to index models, the "intercept" parameter is absorbed into the unknown link function. As noted in section 3.1, in a model of a game, it is important to separately identify and estimate the "intercept" parameters, particularly since the interaction effect parameter is an "intercept" parameter. Moreover, in a model of a game, there are actually multiple "intercept" parameters *per agent* (α_i and Δ_i for each agent i), further complicating estimation.

5.1. **Estimation of** β**.** The first part of theorem 3.1 suggests the following estimator for β based on density-weighted average derivative estimation à la Powell, Stock, and Stoker (1989), with the weight function equal to the density of the explanatory variables (i.e., $\pi(z) = p(z)$). This section assumes $\beta_{1w} \equiv \beta_w \equiv \beta_{2w}$ in order to use the first expression for the identification of β_w in terms of population quantities, but the properties of the estimator without that assumption would be derived similarly.[7]

The data consists of independent observations of the markets (i.e., "plays of the game") indexed by $m = 1, 2, \ldots, M$. Let $\hat{\hat{\delta}}_{M,ixk} = -\frac{2}{M(M-1)} \frac{1}{h_M^{d+1}} \sum_{m=1}^{M} \sum_{m' \neq m} 1[y_m = (0,0)] K^{[1,i\tilde{x}k]} \left(\frac{z_m - z_{m'}}{h_M} \right)$, where $K_1 = dim(x_1)$, $K_2 = dim(x_2)$, $L = dim(w)$, and $d = K_1 + K_2 + L$. The function $K(\cdot)$ is a kernel and h_M is a sequence of bandwidths. The notation $K^{[1,ixk]}(\cdot)$ means the derivative of $K(\cdot)$ with respect to the x_{ik} component of the explanatory variables z. Similarly, let $\hat{\hat{\delta}}_{M,wl} = -\frac{2}{M(M-1)} \frac{1}{h_M^{d+1}} \sum_{m=1}^{M} \sum_{m' \neq m} 1[y_m = (0,0)] K^{[1,wl]} \left(\frac{z_m - z_{m'}}{h_M} \right)$, where the notation $K^{[1,wl]}(\cdot)$ means the derivative of $K(\cdot)$ with respect to the w_l component of the explanatory variables z.

Let $\hat{\hat{\delta}}_M = (\hat{\hat{\delta}}_{M,1x1}, \hat{\hat{\delta}}_{M,1x2}, \ldots, \hat{\hat{\delta}}_{M,1xK_1}, \hat{\hat{\delta}}_{M,2x1}, \ldots, \hat{\hat{\delta}}_{M,2xK_2}, \hat{\hat{\delta}}_{M,w1}, \ldots, \hat{\hat{\delta}}_{M,wL})$. Let δ be the associated population quantities, where $\delta_{ixk} = E\left(p(z) \frac{\partial P(y=(0,0)|z)}{\partial x_{ik}} \right)$ and $\delta_{wl} = E\left(p(z) \frac{\partial P(y=(0,0)|z)}{\partial w_l} \right)$.

The econometrician may or may not want to use the assumption that $\beta_1 \equiv \beta \equiv \beta_2$, which would increase efficiency of estimation, and is required when the labeling of the agents has no economic content.

If so, let $\hat{\hat{\beta}}_{M,xk} = \frac{\hat{\hat{\delta}}_{M,1xk} + \hat{\hat{\delta}}_{M,2xk}}{\hat{\hat{\delta}}_{M,1x1} + \hat{\hat{\delta}}_{M,2x1}}$ for $k \geq 2$, and $\hat{\hat{\beta}}_{M,wl} = \frac{\hat{\hat{\delta}}_{M,wl}}{\hat{\hat{\delta}}_{M,1x1} + \hat{\hat{\delta}}_{M,2x1}}$. Let $\hat{\hat{\beta}}_{M,-1} = (\hat{\hat{\beta}}_{M,x2}, \ldots, \hat{\hat{\beta}}_{M,xK}, \hat{\hat{\beta}}_{M,w1}, \ldots, \hat{\hat{\beta}}_{M,wL})$. Let $\beta_{-1} = (\beta_{x2}, \ldots, \beta_{xK}, \beta_{w1}, \ldots, \beta_{wL})$. (In this case, $K_1 = K = K_2$.)

[7]The properties of the estimator of β_{wl} under the assumption that $E(P_x(z))$ has full rank, which depends on the expectation of certain *products* of derivatives, can be derived similarly to the results of Samarov (1993).

If not, let $\hat{\beta}_{M,ixk} = \frac{\hat{\delta}_{M,ixk}}{\hat{\delta}_{M,ix1}}$ for each i and $k \geq 2$, and $\hat{\beta}_{M,wl} = \frac{\hat{\delta}_{M,wl}}{\hat{\delta}_{M,1x1}+\hat{\delta}_{M,2x1}}$. Let $\hat{\beta}_{M,-1} = (\hat{\beta}_{M,1x2}, \ldots, \hat{\beta}_{M,1xK_1}, \hat{\beta}_{M,2x2}, \ldots, \hat{\beta}_{M,2xK_2}, \hat{\beta}_{M,w1}, \ldots, \hat{\beta}_{M,wL})$. Let $\beta_{-1} = (\beta_{1x2}, \ldots, \beta_{1xK_1}, \beta_{2x2}, \ldots, \beta_{2xK_2}, \beta_{w1}, \ldots, \beta_{wL})$.

In either case, $\hat{\beta}_M$ is a $\sqrt{M}$-consistent and asymptotically normally distributed estimator of β_{-1}. The regularity conditions are the usual sorts of conditions, as in the results of Powell, Stock, and Stoker (1989) which are discussed in Horowitz (2009). Let $S = \frac{d+4}{2}$ if d is even and $S = \frac{d+3}{2}$ if d is odd.

Assumption 5.1 (Continuous explanatory variables, II)**.** *The probability density function $p(\cdot)$ of z has all mixed partial derivatives up to order $S+1$. In particular, $p(z) = 0$ on the boundary of the support of z.*

Assumption 5.2 (Smooth population quantities)**.** *The components of $\frac{\partial P(y=(0,0)|z)}{\partial z}$ and $\frac{\partial p(z)}{\partial z}(1[y = (0,0)], z')$ have finite second moments. $E(1[y = (0,0)]\partial^r p(z))$ exists for $0 < r \leq S+1$. There is a function $m(z)$ such that $E((1 + 1[y = (0,0)] + ||z||)m(z))^2 < \infty$, $||\frac{\partial p(z+\psi)}{\partial z} - \frac{\partial p(z)}{\partial z}|| < m(z)||\psi||$, and $||\frac{\partial p(z+\psi)P(y=(0,0)|z+\psi)}{\partial z} - \frac{\partial p(z)P(y=(0,0)|z)}{\partial z}|| < m(z)||\psi||$.*

Assumption 5.3 (Higher order kernel)**.** *$K(\cdot)$ is a kernel of order S that is symmetric about the origin, bounded, and differentiable.*

Assumption 5.4 (Bandwidth rate)**.** *It holds that $Mh_M^{2S} \to 0$ and $Mh_M^{d+2} \to \infty$.*

Finally, for the covariance matrix, in case it is assumed that $\beta_1 \equiv \beta \equiv \beta_2$, use the notation that $A(\delta) =$

$$\begin{pmatrix} \frac{-\sum_i \delta_{ix2}}{(\sum_i \delta_{ix1})^2} & \frac{1}{\sum_i \delta_{ix1}} & 0 & \cdots & 0 & \frac{-\sum_i \delta_{ix2}}{(\sum_i \delta_{ix1})^2} & \frac{1}{\sum_i \delta_{ix1}} & 0 & \cdots & 0 \\ \frac{-\sum_i \delta_{ix3}}{(\sum_i \delta_{ix1})^2} & 0 & \frac{1}{\sum_i \delta_{ix1}} & \cdots & 0 & \frac{-\sum_i \delta_{ix3}}{(\sum_i \delta_{ix1})^2} & 0 & \frac{1}{\sum_i \delta_{ix1}} & \cdots & 0 \\ & & \cdots & & & & & \cdots & & \\ \frac{-\sum_i \delta_{ixK}}{(\sum_i \delta_{ix1})^2} & 0 & & \cdots & \frac{1}{\sum_i \delta_{ix1}} & \frac{-\sum_i \delta_{ixK}}{(\sum_i \delta_{ix1})^2} & 0 & & \cdots & \frac{1}{\sum_i \delta_{ix1}} \end{pmatrix}$$

and $B(\delta) =$

$$\begin{pmatrix} \frac{-\delta_{w1}}{(\sum_i \delta_{ix1})^2} & \cdots & \cdots & \frac{-\delta_{w1}}{(\sum_i \delta_{ix1})^2} & \cdots & \cdots & \frac{1}{\sum_i \delta_{ix1}} & \cdots & & \\ \frac{-\delta_{w2}}{(\sum_i \delta_{ix1})^2} & \cdots & \cdots & \frac{-\delta_{w2}}{(\sum_i \delta_{ix1})^2} & \cdots & \cdots & & \frac{1}{\sum_i \delta_{ix1}} & \cdots & \\ & \cdots & \cdots & \cdots & & & & & & \\ \frac{-\delta_{wL}}{(\sum_i \delta_{ix1})^2} & \cdots & \cdots & \frac{-\delta_{wL}}{(\sum_i \delta_{ix1})^2} & \cdots & \cdots & & & & \frac{1}{\sum_i \delta_{ix1}} \end{pmatrix}.$$

The matrix $A(\delta)$ is $(K-1)\times 2K$ and $B(\delta)$ is $L\times(2K+L)$. The k-th row of $A(\delta)$ has non-zero entries in the columns: 1, $k+1$, $K+1$, and $K+k+1$. The l-th row of $B(\delta)$ has non-zero entries in the columns: 1, $K+1$, and $2K+l$. Finally, let $C(\delta)=\begin{pmatrix} A(\delta) & 0 \\ B(\delta) & \end{pmatrix}$. The matrix $C(\delta)$ is $((K-1)+L)\times(2K+L)$.

If it is not assumed that $\beta_1\equiv\beta\equiv\beta_2$, use the notation that $A(\delta)=$

$$\begin{pmatrix} \frac{-\delta_{1x2}}{(\delta_{1x1})^2} & \frac{1}{\delta_{1x1}} & 0 & \dots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \frac{-\delta_{1x3}}{(\delta_{1x1})^2} & 0 & \frac{1}{\delta_{1x1}} & \dots & 0 & 0 & 0 & 0 & \cdots & 0 \\ & & \dots & & & & & \dots & & \\ \frac{-\delta_{1xK_1}}{(\delta_{1x1})^2} & 0 & & \dots & \frac{1}{\delta_{1x1}} & 0 & 0 & & \cdots & 0 \\ 0 & 0 & 0 & \dots & 0 & \frac{-\delta_{2x2}}{(\delta_{2x1})^2} & \frac{1}{\delta_{2x1}} & 0 & \cdots & 0 \\ 0 & 0 & 0 & \dots & 0 & \frac{-\delta_{2x3}}{(\delta_{2x1})^2} & 0 & \frac{1}{\delta_{2x1}} & \cdots & 0 \\ & & \dots & & & & & \dots & & \\ 0 & 0 & & \dots & 0 & \frac{-\delta_{2xK_2}}{(\delta_{2x1})^2} & 0 & & \cdots & \frac{1}{\delta_{2x1}} \end{pmatrix}$$

and $B(\delta)$ be defined as above. Finally, let $C(\delta)=\begin{pmatrix} A(\delta) & 0 \\ B(\delta) & \end{pmatrix}$. The matrix $C(\delta)$ is $((K_1-1)+(K_2-1)+L)\times(K_1+K_2+L)$.

Again, this result assumes a non-positive interaction effect, but the case of a non-negative interaction effect is similar. Note that the dimension of the estimators depend on whether or not the econometrician assumes that $\beta_1\equiv\beta\equiv\beta_2$.

Theorem 5.1 (Estimation of β)**.** *Suppose that $\beta_{1w}\equiv\beta_w\equiv\beta_{2w}$ and $\Delta_1\le 0$ and $\Delta_2\le 0$. Under assumptions 3.1, 3.2, 3.3, 3.4, 5.1, 5.2, 5.3, and 5.4, it holds that*

$$\sqrt{M}(\hat\beta_{M,-1}-\beta_{-1})\to^d N(0,C(\delta)V(\delta)C(\delta)'),$$

where

$$V(\delta)=4E\left(\left(p(z)\frac{\partial P(y=(0,0)|z)}{\partial z}-(1[y=(0,0)]-P(y=(0,0)|z))\frac{\partial p(z)}{\partial z}\right)\times\right.$$
$$\left.\left(p(z)\frac{\partial P(y=(0,0)|z)}{\partial z}-(1[y=(0,0)]-P(y=(0,0)|z))\frac{\partial p(z)}{\partial z}\right)'\right)-4\delta\delta'.$$

5.2. **Estimation of α and Δ.** The second part of theorem 3.1 suggests the following estimator for α and Δ: first, estimate $\left.\frac{\partial^2 P(y=(0,0)|c_1,c_2)}{\partial c_1\partial c_2}\right|_{a_1,a_2}$ and $\left.\frac{\partial^2 P(y=(1,1)|c_1,c_2)}{\partial c_1\partial c_2}\right|_{b_1,b_2}$ by a non-parametric regression of $(0,0)$ and $(1,1)$ on the generated regressors $\hat c_m$ where $\hat c_{im}=-x_{im}\hat\beta_{ix}-w_m\hat\beta_{iw}$, where $\hat\beta$ is an estimate of β; second, estimate α and $\alpha+\Delta$ by maximizing the first stage estimators. Estimate Δ by differencing these two estimates. Although this paper does propose a particular estimator for β, the estimator for α and Δ does not depend critically on the details of the estimator for β, other than that it converges at the parametric rate.

So, let

$$\hat{Q}_M(\gamma) = \left(\left. \frac{\partial^2 P_M(y=(0,0)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1 \partial \hat{c}_2} \right|_{a_1,a_2} \quad \left. \frac{\partial^2 P_M(y=(1,1)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1 \partial \hat{c}_2} \right|_{b_1,b_2} \right)'$$

be a 2×1 vector sample objective function, where $\gamma = (a_1, a_2, b_1, b_2)$, and

$$\left. \frac{\partial^2 P_M(y = (0,0)|\hat{c}_1, \hat{c}_2)}{\partial \hat{c}_1 \partial \hat{c}_2} \right|_{a_1,a_2}$$

and

$$\left. \frac{\partial^2 P_M(y = (1,1)|\hat{c}_1, \hat{c}_2)}{\partial \hat{c}_1 \partial \hat{c}_2} \right|_{b_1,b_2}$$

are non-parametric estimators based on generated regressors $\hat{c}_m = (\hat{c}_{m1}, \hat{c}_{m2})$.

Let $\gamma_0 = (\alpha_1, \alpha_2, \alpha_1 + \Delta_1, \alpha_2 + \Delta_2)$.

The estimator $\hat{\gamma}_M$ maximizes the components of $\hat{Q}_M(\gamma)$.

Let

$$\begin{aligned} \widehat{\Delta Q}_M(\gamma) = & \left(\left. \frac{\partial^2 P_M(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2} \right|_{a_1,a_2} \quad \left. \frac{\partial^2 P_M(y=(1,1)|c_1,c_2)}{\partial c_1 \partial c_2} \right|_{b_1,b_2} \right)' \\ & - \left(\left. \frac{\partial^2 P_M(y=(0,0)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1 \partial \hat{c}_2} \right|_{a_1,a_2} \quad \left. \frac{\partial^2 P_M(y=(1,1)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1 \partial \hat{c}_2} \right|_{b_1,b_2} \right)', \end{aligned}$$

so that the "infeasible" objective function that uses the true c_{im} as the regressors is $Q_M(\gamma) = \hat{Q}_M(\gamma) + \widehat{\Delta Q}_M(\gamma)$. This is infeasible because $c_{im} = -x_{im}\beta_{ix} - w_m\beta_{iw}$ are not observed. The parameter space for γ is Γ.

Assumption 5.5 (Compact parameter space)**.** Γ *is compact.*

The following assumptions require that the non-parametric estimator is suitably well-behaved. They are high-level assumptions that admit a variety of possible approaches to the non-parametric estimation, including kernel regression.

Assumption 5.6 (Uniform convergence of regression estimate)**.** *The infeasible* $Q_M(\gamma)$ *converges in probability to* $Q(\gamma) = \left(\left. \frac{\partial^2 P(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2} \right|_{a_1,a_2} \quad \left. \frac{\partial^2 P(y=(1,1)|c_1,c_2)}{\partial c_1 \partial c_2} \right|_{b_1,b_2} \right)'$ *uniformly over the parameter space* Γ*.*

Assumption 5.6 requires the usual uniform convergence properties of a non-parametric regression estimator over a compact set, for the *infeasible* estimator that uses the "true" regressors $c = (c_1, c_2)$. Therefore, conditions under which it holds are understood, and can be found in the literature on non-parametric regression.

Assumption 5.7 (Lipschitz properties of feasible and infeasible estimators)**.** *There is a function of the data* $\tilde{R}_{M,1}(a_1, a_2, B)$ *with a continuous derivative with respect to B such that* $\tilde{R}_{M,1}(a_1, a_2, \beta) = \frac{\partial^2 P_M(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2}\Big|_{a_1,a_2}$ *and* $\tilde{R}_{M,1}(a_1, a_2, \hat{\beta}) = \frac{\partial^2 P_M(y=(0,0)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1 \partial \hat{c}_2}\Big|_{a_1,a_2}$. *There is a function of the data* $\tilde{R}_{M,2}(b_1, b_2, B)$ *with a continuous derivative with respect to B such that* $\tilde{R}_{M,2}(b_1, b_2, \beta) = \frac{\partial^2 P_M(y=(1,1)|c_1,c_2)}{\partial c_1 \partial c_2}\Big|_{b_1,b_2}$ *and* $\tilde{R}_{M,2}(b_1, b_2, \hat{\beta}) = \frac{\partial^2 P_M(y=(1,1)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1 \partial \hat{c}_2}\Big|_{b_1,b_2}$. *And, if* $\tilde{\beta} - \beta = O_p(M^{-\frac{1}{2}})$, *then* $\sup_{a_1,a_2} \left|\left| \frac{\partial \tilde{R}_{M,1}(a_1,a_2,B)}{\partial B}\Big|_{\tilde{\beta}} \right|\right| = O_p(1)$ *and* $\sup_{b_1,b_2} \left|\left| \frac{\partial \tilde{R}_{M,2}(b_1,b_2,B)}{\partial B}\Big|_{\tilde{\beta}} \right|\right| = O_p(1)$.

Assumption 5.7 is used to imply that the difference between the infeasible estimator $Q_M(\gamma)$ and the feasible estimator $\hat{Q}_M(\gamma)$ is asymptotically negligible, as long as $\hat{\beta}$ converges at the parametric rate. This follows from assumption 5.7 by a Taylor series approximation to $\hat{Q}_M(\gamma)$. (See the proof of theorem 5.2 for the details.)

The parameter "B" in assumption 5.7 reflects a particular specification of the β parameter (for example, an estimate $\hat{\beta}$). So, letting $c_{im}(B) = -x_{im}B_{ix} - w_m B_{iw}$ be the generated regressors at parameter "B", then $\tilde{R}_{M,1}(a_1, a_2, B)$ can be the second derivative of the non-parametric regression of $1[y = (0, 0)]$ on $c_m(B)$, and $\tilde{R}_{M,2}(a_1, a_2, B)$ can be the second derivative of the non-parametric regression of $1[y = (1, 1)]$ on $c_m(B)$. This establishes the first part of the assumption. The second part of assumption 5.7 can be established by application of a suitable uniform law of large numbers. In particular, in the case that the estimation is by non-parametric kernel regression, see the arguments of Horowitz (2009, Section 2.4).

The preceding assumptions are sufficient for consistency.

The following additional assumptions imply asymptotic normality, and establish the rate of convergence. Essentially, these assumptions require that: there is not a parameter on the boundary problem, the optimization problem is well-behaved in the sample and the population, and the estimates of the first and second derivatives of the objective function suitably converge to the corresponding population quantities at suitable rates.

Let $\frac{\partial Q(\gamma)}{\partial \gamma}$ be a 4×1 vector with components $\frac{\partial^3 P(y=(0,0)|c_1,c_2)}{\partial c_1^2 \partial c_2}\Big|_{a_1,a_2}$, $\frac{\partial^3 P(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2^2}\Big|_{a_1,a_2}$, $\frac{\partial^3 P(y=(1,1)|c_1,c_2)}{\partial c_1^2 \partial c_2}\Big|_{b_1,b_2}$, and $\frac{\partial^3 P(y=(1,1)|c_1,c_2)}{\partial c_1 \partial c_2^2}\Big|_{b_1,b_2}$. Also let $\frac{\partial^2 Q(\gamma)}{\partial \gamma \partial \gamma'}$ be the 4×4 matrix of derivatives with respect to γ' of $\frac{\partial Q(\gamma)}{\partial \gamma}$. Let $\frac{\partial Q_M(\gamma)}{\partial \gamma}$ and $\frac{\partial^2 Q_M(\gamma)}{\partial \gamma \partial \gamma'}$ respectively be the infeasible non-parametric estimators based on c, and let $\frac{\partial \hat{Q}_M(\gamma)}{\partial \gamma}$ and $\frac{\partial^2 \hat{Q}_M(\gamma)}{\partial \gamma \partial \gamma'}$ respectively be the feasible non-parametric estimators based on $\hat{c}$.

Assumption 5.8 (Parameter in the interior)**.** *γ_0 is in the interior of Γ.*

Assumption 5.9 (Asymptotic distribution of derivatives)**.** *$\left.\frac{\partial Q_M(\gamma)}{\partial \gamma}\right|_{\gamma_0}$ exists and $r_M \left.\frac{\partial Q_M(\gamma)}{\partial \gamma}\right|_{\gamma_0} \to^d N(0, \Omega_0)$ at the rate r_M with $r_M M^{-\frac{1}{2}} \to 0$.*

Assumption 5.10 (Uniform convergence of derivatives)**.** *$\frac{\partial^2 Q_M(\gamma)}{\partial \gamma \partial \gamma'}$ exists and is continuous on a neighborhood of γ_0 and converges in probability to $\frac{\partial^2 Q(\gamma)}{\partial \gamma \partial \gamma'}$ uniformly over a neighborhood of γ_0.*

Assumption 5.11 (Lipschitz properties of feasible and infeasible estimators)**.** *The conditions in assumption 5.7 hold for all components of $\frac{\partial Q_M(\gamma)}{\partial \gamma}$ and $\frac{\partial^2 Q_M(\gamma)}{\partial \gamma \partial \gamma'}$.*

Assumption 5.12 (Negative definite Hessian)**.** *The Hessian of $f_{\epsilon_1, \epsilon_2}(\cdot)$ is negative definite on a neighborhood of $(0, 0)$.*

Assumption 5.13 (Smooth feasible objective function)**.** *The components of $\hat{Q}_M(\gamma)$ have continuous second derivatives with respect to γ in a neighborhood of γ_0.*

In general, if the econometrician does not assume that $\alpha_1 \equiv \alpha \equiv \alpha_2$ and $\Delta_1 \equiv \Delta \equiv \Delta_2$, as below, the estimator is $\hat{\psi}_M = (\hat{\alpha}_{M1}, \hat{\alpha}_{M2}, \hat{\Delta}_{M1}, \hat{\Delta}_{M2})$, where $\hat{\alpha}_{M1} = \hat{\gamma}_{M1}$, $\hat{\alpha}_{M2} = \hat{\gamma}_{M2}$, $\hat{\Delta}_{M1} = \hat{\gamma}_{M3} - \hat{\gamma}_{M1}$, and $\hat{\Delta}_{M2} = \hat{\gamma}_{M4} - \hat{\gamma}_{M2}$. The corresponding true parameter is $\psi = (\alpha_1, \alpha_2, \Delta_1, \Delta_2)$. Further, $\alpha = (\alpha_1, \alpha_2)$ and $\Delta = (\Delta_1, \Delta_2)$.

Further, the estimation result allows the econometrician to assume that $\alpha_1 \equiv \alpha \equiv \alpha_2$ and $\Delta_1 \equiv \Delta \equiv \Delta_2$, by imposing that condition on the objective function, to increase efficiency of estimation. Note that this involves a different "definition" of α and Δ compared to above, where α and Δ were vectors. As in previous sections, this is required when the labeling of the agents has no economic content. The constrained objective function is $Q^r(\gamma^r) = Q(\gamma_1^r, \gamma_1^r, \gamma_2^r, \gamma_2^r)$, where $\gamma^r = (a, b)$. In that case, the estimator is $\hat{\psi}_M = (\hat{\alpha}_M, \hat{\Delta}_M)$, where $\hat{\alpha}_M = \hat{\gamma}_{M1}^r$, and $\hat{\Delta}_M = \hat{\gamma}_{M2}^r - \hat{\gamma}_{M1}^r$. The true parameter is $\psi = (\alpha, \Delta)$.

The following estimation result assumes a non-positive interaction effect, but the case of a non-negative interaction effect is similar. Note that the dimension of the estimators depend on whether or not the econometrician assumes that $\alpha_1 \equiv \alpha \equiv \alpha_2$ and $\Delta_1 \equiv \Delta \equiv \Delta_2$.

Theorem 5.2 (Estimation of α and Δ)**.** *Suppose that $\Delta_1 \leq 0$ and $\Delta_2 \leq 0$. Suppose that β is point identified and that $\hat{\beta}$ is an estimator of β such that $\hat{\beta} - \beta = O_p(M^{-\frac{1}{2}})$.*

Under assumptions 3.2, 3.3, 3.5, 5.5, 5.6, and 5.7,

$$\hat{\alpha}_M \to^p \alpha \text{ and } \hat{\Delta}_M \to^p \Delta.$$

Under the additional assumptions 5.8, 5.9, 5.10, 5.11, 5.12, and 5.13,

$$r_M(\hat{\psi}_M - \psi) \to^d N\left(0, V_0\right).$$

If it is not assumed that $\alpha_1 \equiv \alpha \equiv \alpha_2$ *and* $\Delta_1 \equiv \Delta \equiv \Delta_2$*, then*

$$V_0 = C\left(\left.\frac{\partial^2 Q(\gamma)}{\partial\gamma\partial\gamma'}\right|_{\gamma_0}\right)^{-1}\Omega_0\left(\left.\frac{\partial^2 Q(\gamma)}{\partial\gamma\partial\gamma'}\right|_{\gamma_0}\right)^{-1}C'$$

where $C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}$,

If it is assumed that $\alpha_1 \equiv \alpha \equiv \alpha_2$ *and* $\Delta_1 \equiv \Delta \equiv \Delta_2$*, then*

$$V_0 = C\left(\left.\frac{\partial^2 Q^r(\gamma)}{\partial\gamma\partial\gamma'}\right|_{\gamma_0}\right)^{-1}D\Omega_0 D'\left(\left.\frac{\partial^2 Q^r(\gamma)}{\partial\gamma\partial\gamma'}\right|_{\gamma_0}\right)^{-1}C'$$

where $C = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$, $D = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$, *and*

$$\frac{\partial^2 Q^r(\gamma)}{\partial\gamma\partial\gamma'} = \begin{pmatrix} \frac{\partial^2 Q^r(\gamma)}{\partial a^2} & 0 \\ 0 & \frac{\partial^2 Q^r(\gamma)}{\partial b^2} \end{pmatrix},$$

where

$$\frac{\partial^2 Q^r(\gamma)}{\partial a^2} = \left.\frac{\partial^4 P(y=(0,0)|c_1,c_2)}{\partial c_1^3\partial c_2}\right|_{a,a} + 2\left.\frac{\partial^4 P(y=(0,0)|c_1,c_2)}{\partial c_1^2\partial c_2^2}\right|_{a,a} + \left.\frac{\partial^4 P(y=(0,0)|c_1,c_2)}{\partial c_1\partial c_2^3}\right|_{a,a}$$

and

$$\frac{\partial^2 Q^r(\gamma)}{\partial b^2} = \left.\frac{\partial^4 P(y=(1,1)|c_1,c_2)}{\partial c_1^3\partial c_2}\right|_{b,b} + 2\left.\frac{\partial^4 P(y=(1,1)|c_1,c_2)}{\partial c_1^2\partial c_2^2}\right|_{b,b} + \left.\frac{\partial^4 P(y=(1,1)|c_1,c_2)}{\partial c_1\partial c_2^3}\right|_{b,b}.$$

Remark 5.1 (Rate of convergence)**.** The rate of convergence is the same as the rate of convergence of the derivatives of the objective function, which is essentially the rate of convergence of a third derivative of a regression function, from assumption 5.9, which is slower than $\sqrt{M}$. The reason it is the rate of convergence of a third derivative is that the estimation of γ corresponds to maximization of the second derivative of the regression function, which is essentially equivalent to estimating the third derivative and solving for where it is equal to zero.

Note from the identification strategy that since the regression concerns estimating (the derivatives of) a certain integral of the density of the unobservables, the rate of convergence will depend on the assumed smoothness of the density of the unobservables. (And since integration is a smoothing operator, the estimator is estimating a function that is smoother than the density itself.) Note also that since this regression depends on the *generated regressors* c, which is necessarily in $\mathbb{R}^2$ regardless of the overall number of explanatory variables in the model, estimation of α and Δ is not subject to the curse of dimensionality as a function of the number of explanatory variables. This reflects a dimension reduction strategy.

The slower than $\sqrt{M}$-rate of convergence is not surprising. Khan and Nekipelov (2012) show that there is zero Fisher information about the interaction effect in certain models of complete information games, which implies by Chamberlain (1986) that there cannot be a regular estimator that converges at the parametric rate. Further, there is zero Fisher information about the intercept of a single-agent discrete choice model (i.e., Cosslett (1987) and Pagan and Ullah (1999, Section 7.3)) under standard assumptions. Since a complete information game nests "independent instances" of the single-agent discrete choice model as a special case (i.e., when the interaction effect is zero, and everything else is independent across agents), there is zero Fisher information about α under the same assumptions. (The more important interaction effect parameter Δ disappears by assumption in this sub-model, so critically requires the additional arguments like those of Khan and Nekipelov (2012)). Consequently, combining all of these results, an estimator of the "intercept" parameters that converges at slower than the parametric rate seems unavoidable without stronger assumptions.

Remark 5.2 (Non-parametric kernel regression)**.** Under standard conditions (e.g., Bierens (1987), Andrews (1995), and Pagan and Ullah (1999), among others), these conditions are satisfied by (trimmed) kernel regression estimators. In that case, Ω_0 from assumption 5.9 can be found in standard references, as it is the asymptotic covariance of estimates of the third derivatives of a regression function. The rate of convergence depends on the smoothness assumed by the econometrician and the bandwidth rate, but as noted in remark 5.1, not the number of explanatory variables in the model because of the dimension reduction strategy.

6. Monte Carlo experiment

This section reports the results of a Monte Carlo experiment of the small sample performance of the estimator proposed in section 5. Based on the model studied in this paper, shown in normal form in table 1, the Monte Carlo experiment involves the following specification of the true data generating process:

(1) $\tilde{x} = (\tilde{x}_{11}, \tilde{x}_{12}, \tilde{x}_{21}, \tilde{x}_{22}) \sim N_4(0, \Sigma_{\tilde{x}})$ where $\Sigma_{\tilde{x}} = 0.90 \times I_{4\times4} + 0.10 \times 1_{4\times4}$ (where $1_{a\times b}$ is the $a \times b$ matrix of all 1s)
(2) $x_{ik} = \frac{1}{\pi} \arctan(\tilde{x}_{ik})$
(3) w is void
(4) $\epsilon \sim N_2(0, \Sigma_\epsilon)$ where $\Sigma_\epsilon = 0.03 \times I_{2\times2} + 0.01 \times 1_{2\times2}$.
(5) $\beta_1 = (1, 0.75) = \beta_2$
(6) $\alpha_1 = 0.1 = \alpha_2$
(7) $\Delta_1 = -0.2 = \Delta_2$
(8) $M = 500$ or $M = 1000$

Thus, there are two agent-specific explanatory variables per agent (and therefore a total of four agent-specific explanatory variables). These explanatory variables are a translation (by the arctan function) of positively correlated standard normal random variables. (The components of $\tilde{x}$ have unit variances and covariances of 0.10.) The arctan translation is used to generate explanatory variables that have bounded support, as arctan has bounded range. Because of the scaling by $\frac{1}{\pi}$, the support of each explanatory variable is the unit-length interval $[-\frac{1}{2}, \frac{1}{2}]$. The positive correlation reflects the notion that the various observable "components" of profits are likely positively related. (After the translation by arctan, the explanatory variables x_{ik} have approximately variances 0.0455 and covariances 0.0043.) There are no shared explanatory variables. The unobservables are jointly normally distributed with positive correlation, reflecting an unobserved market fixed effect. (The unobservables have variance 0.04 and covariance 0.01, for a correlation of 0.25.) The game is a game of strategic substitutes (e.g., an entry game), since $\Delta_1 \leq 0$ and $\Delta_2 \leq 0$, and this is known by the econometrician. The experiment is run twice: when the sample has $M = 500$ and when the sample has $M = 1000$ observations of markets. For each sample size, 1500 such samples are generated, and the estimators recorded. The estimator imposes that the parameters are equal across player roles.

The details of the implementation of the estimator are the following. First, the estimator requires the specification of two kernels: a kernel for the estimator corresponding to β, and (implicitly) a kernel for the estimator corresponding to α and Δ as the non-parametric estimate corresponding to theorem 5.2 is chosen to also be a kernel regression estimator. The kernel for estimation of β is the product of four fourth-order Gaussian kernels, where the order is per assumption 5.1, and the product of four is per the fact that there are four explanatory variables. The kernel for estimation of α and Δ is the product of two fourth order Gaussian kernels. Also, the estimator requires two bandwidths: a bandwidth for the estimator corresponding to β, and (implicitly) a bandwidth for the estimator corresponding to α and Δ. Based on results on the optimal rate of convergence of the bandwidths (e.g., Powell and Stoker (1996) and Pagan and Ullah (1999)), the bandwidth for the estimator corresponding to β is proportional to $M^{-\frac{1}{7}}$ (as the optimal bandwidth for density-weighted average derivative estimation with four regressors) and the bandwidth for the estimator corresponding to α and Δ is proportional to $M^{-\frac{1}{16}}$ (as the optimal bandwidth for kernel regression estimation of a third derivative with two regressors and a fourth-order kernel). In particular, the bandwidth for the estimator corresponding to β is based on the "plug-in" estimator for density-weighted average derivatives proposed in Powell and Stoker (1996). Based on these bandwidths, the estimation theorems imply that the rate of convergence of the estimator for β is $M^{\frac{1}{2}}$ and the rate of convergence of the estimator for α and Δ is $M^{\frac{1}{4}}$.

Parameter	Mean	Median	MSE	IQR	Variance	Rej. Rate
$N = 500$						
$\beta = 0.750$	0.745	0.735	0.016	0.172	0.016	0.100
$\alpha = 0.100$	0.084	0.084	0.011	0.142	0.011	0.129
$\Delta = -0.200$	-0.169	-0.180	0.024	0.196	0.023	0.115
$N = 1000$						
$\beta = 0.750$	0.740	0.734	0.008	0.121	0.008	0.104
$\alpha = 0.100$	0.095	0.098	0.009	0.135	0.009	0.093
$\Delta = -0.200$	-0.193	-0.200	0.018	0.168	0.018	0.073

TABLE 2. Numerical results of Monte Carlo experiment

The results of the Monte Carlo experiment are displayed in table 2. Table 2 reports summary statistics for the estimators. Also, in the last column, table 2 reports the empirical probability of rejecting the truth for a test at the $p = 0.10$ significance level. The test is based on the asymptotic covariances derived in the theoretical results, which are estimated (in each Monte Carlo sample) by replacing unknown population quantities by sample analogues. The results suggest the estimators and associated tests have good performance. The estimators seem approximately mean and median unbiased, and the tests reject the truth at roughly the nominal rate of $p = 0.10$. As expected, the performance of the estimators is better for the larger sample size. The rates of the decline of the variances with sample size are similar to that predicted by the theoretical rates of convergence (i.e., the variance of the estimator for β with $M = 1000$ is approximately 0.5 times the variance of the estimator with $M = 500$, and the variance of the estimators for α and Δ with $M = 1000$ are approximately $\frac{1}{\sqrt{2}} = 0.7$ times the variance of the estimator with $M = 500$).

7. Conclusions and discussion

This paper makes two contributions to the literature on the identification and estimation of complete information games. First, it shows that it is possible to point identify the parameters of the utility functions in a complete information game without a large support regressor. This is a consequence of the fact that complete information games have empirical content for all values of the explanatory variables. The identification argument uses the non-standard but plausible assumption that the mode of the joint distribution of the unobservables is zero. And second, it shows that it is possible to semi-parametrically estimate the parameters of the utility function, again without a large support regressor. The resulting estimator is consistent and asymptotically normal, but non-standard in the sense that the estimator of the interaction effect parameter converges at slower than the parametric rate. Apparently, this is the first semiparametric estimator available for complete information games, with or without a large support regressor. An intermediate result of this paper, potentially of independent interest, concerns identification and estimation of the direction of the interaction effect.

Even though this paper focuses on an identification strategy based on "unique potential outcomes," the broader contribution is to show that identification strategies for complete information games can exploit more fully the structural relationship between the exogenous explanatory variables and the outcomes that is implied by Nash equilibrium. Indeed, Kline (2013) shows that point identification with a large support regressor does not require the assumption of Nash equilibrium, but rather only a weaker solution concept related to (and indeed strictly weaker than) rationalizability (i.e., Bernheim (1984) and Pearce (1984)) that relates to the "rationality" of the players but not the additional conditions that are entailed by Nash equilibrium.

In aggregate, adding the identification result of this paper to prior identification results implies that there is a "identification possibility frontier" between the solution concept (e.g., Nash equilibrium or rationalizability) and the assumptions on the explanatory variables (e.g., large support or bounded support), in the sense that point identification is possible under stronger assumptions on the solution concept and weaker assumptions on the explanatory variables, or under weaker assumptions on the solution concept and stronger assumptions on the explanatory variables. The identification strategy used in this paper more fully exploits the structural relationship implied by Nash equilibrium, and therefore is compatible with weaker assumptions on the explanatory variables. The identification strategy "more fully" exploits the assumption of Nash equilibrium, in the sense that under weaker solution concepts like rationalizability there are not "unique potential outcomes."

This paper has focused on identification and estimation of the finite-dimensional parameters of the utility functions, but it is worth remarking on identification of the distribution of the unobservables and selection mechanism. Under only the assumptions in this paper, it is not possible in general to point identify the distribution of the unobservables. In particular, it is not possible to point identify the tails of the distribution of the unobservables. For example, suppose that $\Delta_1 \leq 0$ and $\Delta_2 \leq 0$, and let $\bar{t} = (\min\{x_{1m}\beta_{1x} + w_m\beta_{1w}\}, \min\{x_{2m}\beta_{2x} + w_m\beta_{2w}\})$ where the minimum is taken over the support of the exogenous explanatory variables. Unless there is a regressor with "large support," $\bar{t}$ is finite. Then, note that if $\epsilon_{im} > -\alpha_i - \bar{t}_i - \Delta_i$ for all agents i, then for any realization of the explanatory variables in the support, $\alpha_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i + \epsilon_{im} > 0$, so $(1,1)$ is necessarily the unique pure strategy Nash equilibrium for all such ϵ_m. Consequently, any rearrangement of the probability mass of the distribution of ϵ *within* the region where $\epsilon_i > -\alpha_i - \bar{t}_i - \Delta_i$ for all agents i results in the same distribution over $P(y|z)$, and therefore the distribution of ϵ is not point identified. In contrast, Kline (2013) shows identification using a regressor with large support.

Of course, it may be possible to point identify the distribution of ϵ under stronger parametric assumptions, but without a large support regressor. Indeed, in general parametric distributional assumptions on the unobservables have significant identifying power. For example, arguments like in the proof of theorem 3.1 show that the values of the cumulative distribution function of ϵ are point identified in a certain region of its argument, as long as the finite-dimensional parameters are point identified. The region where the cumulative distribution function is point identified depends on the support of $(\alpha_1 + x_{1m}\beta_{1x} + w_m\beta_{1w}, \alpha_2 + x_{2m}\beta_{2x} + w_m\beta_{2w})$. Therefore, if a parametric family for the distribution of ϵ is sufficiently small so that there is a one-to-one mapping between the cumulative distribution functions restricted only to being evaluated on that support, and all cumulative distribution functions in the family, then the distribution of ϵ is point identified. It is not difficult to show, for example, conditions under which this is true for the normal distribution with mean zero but unknown covariance Σ, by exploiting the fact that if $f_\epsilon(t)$ is the density at t, and $\frac{f_\epsilon(t)}{dt}$ is the 2×1 vector of derivatives of the density with respect to the arguments, then $\frac{-\frac{f_\epsilon(t)}{dt}}{f_\epsilon(t)} = \Sigma t$, so Σ can be recovered.

If the distribution of the unobservables can be point identified, then it is almost immediate to point identify the selection mechanism. This is because it is possible to write (for example in the case of a non-positive interaction effect) $P(y = (0,1)|z) = P(y = (0,1) \text{ is a unique equilibrium}|z) + P(y = (0,1)|\epsilon \in \mathcal{R}^-(z,\theta), z)P(\epsilon \in \mathcal{R}^-(z,\theta))$, where $\mathcal{R}^-(z,\theta), z)$ was defined in section 4 as the set of ϵ such that there are multiple equilibria. Since the condition that a certain outcome is a unique equilibrium is simply a probability of a certain region of ϵs (the probability the ϵs are in the region where that outcome is the unique equilibrium), once θ and the distribution of the unobservables are point identified, all terms in this expression are known except for the selection mechanism $P(y = (0,1)|\epsilon \in \mathcal{R}^-(z,\theta), z)$, and therefore the selection mechanism is indeed point identified. And so, if the econometrician had a parametric model for the distribution of the unobservables and the selection mechanism, then maximum likelihood could be used to estimate the model, as all unknowns would be point identified.

8. Appendix: Identification of the direction of the interaction effect with independence

It is possible to *non-parametrically* point identify the sign of the interaction effect using a strategy similar to de Paula and Tang (2012) for incomplete information games; see below for a comparison. The setup is the game in normal form in table 3. Note that these utility functions are non-parametric.

	0	1
0	$(0,0)$	$(0, u_{2m}(0,1))$
1	$(u_{1m}(1,0), 0)$	$(u_{1m}(1,1), u_{2m}(1,1))$

TABLE 3. Non-parametric specification of utility functions

Then, let

$$\tilde{\Delta} = \begin{cases} 1 & \text{if } P(u_{1m}(1,1) > u_{1m}(1,0), u_{2m}(1,1) > u_{2m}(0,1)) = 1 \\ 0 & \text{if } P(u_{1m}(1,1) = u_{1m}(1,0), u_{2m}(1,1) = u_{2m}(0,1)) = 1 \\ -1 & \text{if } P(u_{1m}(1,1) < u_{1m}(1,0), u_{2m}(1,1) < u_{2m}(0,1)) = 1. \end{cases}$$

This section shows how to identify $\tilde{\Delta}$. It is assumed that one of these cases hold (i.e., it cannot be that the interaction effect is sometimes positive and sometimes negative). Otherwise, the interaction effect (i.e., $u_{1m}(1,1) - u_{1m}(1,0)$ and $u_{2m}(1,1) - u_{2m}(0,1)$) can have heterogeneity of unrestricted form. For example, in the canonical linear model $u_{im}(1, y_{(-i)m}) = \alpha_i + x_{im}\beta_i + \Delta y_{(-i)m} + \epsilon_{im}$, it follows that sgn $(\Delta) = \tilde{\Delta}$.

Theorem 8.1 (Non-parametric identification of the sign of the interaction effect with independence)**.** *Suppose that the model of the interaction is given in normal form in table 3, and suppose there is pure strategy Nash equilibrium play. Suppose that there is zero probability that any component of* $u = (u_1(1,0), u_1(1,1), u_2(0,1), u_2(1,1))$ *equals zero, and* $(u_1(1,0), u_1(1,1)) \perp (u_2(0,1), u_2(1,1))$. *Also suppose that* $0 < P(y_1, y_2) < 1$ *for all* $(y_1, y_2) \in \{0,1\}^2$. *Then the following holds:*

(1) If: either $\tilde{\Delta} = 0$, *or both* $P(\text{sgn } (u_1(1,0)) \neq \text{sgn } (u_1(1,1))) > 0$ *and* $P(\text{sgn } (u_2(0,1)) \neq \text{sgn } (u_2(1,1))) > 0$: $\tilde{\Delta} = \text{sgn } \log\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}\right)$.

(2) In general: if $\tilde{\Delta} \leq 0$ *then* $\log\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}\right) \leq 0$; *and if* $\tilde{\Delta} \geq 0$ *then* $\log\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}\right) \geq 0$. *Also, if* $\log\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}\right) < 0$ *then* $\tilde{\Delta} < 0$; *and if* $\log\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}\right) > 0$ *then* $\tilde{\Delta} > 0$.

Remark 8.1 (Intuition for identification strategy)**.** $\tilde{\Delta} = 1$ induces a positive correlation between y_1 and y_2, since $y_j = 1$ increases the probability that action 1 is the utility maximizing action of agent i; conversely, $\tilde{\Delta} = -1$ induces a negative correlation between y_1 and y_2, since $y_j = 1$ decreases the probability that action 1 is the utility maximizing action of agent i. This can be captured by the "pointwise mutual information" statistic $\log\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}\right)$ from information theory.

Remark 8.2 (Condition on sign of utility functions)**.** The condition in part 1 rules out a situation in which there is a non-zero interaction effect (i.e., $\tilde{\Delta} \neq 0$), but effectively no strategic interaction because agents have utility functions that are always positive (or always negative) no matter what the other agent does. In such a case, there is "effectively" a zero interaction effect. In the canonical linear model, with $\Delta < 0$, this condition is equivalent to $P(\alpha_i + x_i\beta_i + \Delta_i < -\epsilon_i < \alpha_i + x_i\beta_i) > 0$, which would be implied by unobservables with support on the real line.

Remark 8.3 (Equilibrium existence)**.** A pure strategy Nash equilibrium exists (with probability one), by an argument similar to Kline and Tamer (2012).

Remark 8.4 (Explanatory variables)**.** It is possible to do the analysis conditional on some explanatory variables x, which may increase the credibility of the independence assumption. In that case, the assumption is that the "unobservables" are independent across agents, which is essentially also assumed by de Paula and Tang (2012) but in the context of incomplete information games. Alternatively, theorem 4.1 provides a different result that uses explanatory variables but does not require independence.

Remark 8.5 (Comparison to de Paula and Tang (2012))**.** de Paula and Tang (2012) show under their (similar) assumptions in the context of incomplete information games: if $\log\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}\right) > 0$ then $\tilde{\Delta} = 1$, and if $\log\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}\right) < 0$ then $\tilde{\Delta} = -1$.[8] As long as $\log\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}\right) \neq 0$ the results overlap, implying that the use of $\log\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}\right)$ as a statistic for $\tilde{\Delta}$ is partially robust to different conditions on information (i.e., incomplete information versus complete information). Despite the similarity in the results, the proofs are quite different, because they apply to different conditions on information.

But also there are some important differences in the actual results; in particular, the de Paula and Tang (2012) result is silent about $\tilde{\Delta}$ if $\log\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}\right) = 0$. They also show that, when there is incomplete information, $\log\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}\right) \neq 0$ if and only if there are multiple Bayesian Nash equilibria that are played in the data generating process. In contrast, with complete information, the results in this paper show that $\log\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}\right) = 0$ is equivalent to $\tilde{\Delta} = 0$. One implication of this

[8] de Paula and Tang (2012) do not give literally this result, but the equivalence is evident after translating the notation, and some algebra.

difference is that in the incomplete information model of de Paula and Tang (2012), it is not possible to "prove" that the interaction effect is zero, essentially because, with incomplete information, it is not possible to distinguish between the possibility of a zero interaction effect, and the possibility that there is just one Bayesian Nash equilibrium being played in the data generating process.[9] These results provide an interesting contrast of the empirical content of complete and incomplete information games.

8.1. **Non-parametric estimation of the sign of the interaction effect.** The following is a consistent estimator for $\tilde{\Delta}$, using the notation that $P_M(\cdot)$ is the sample distribution of M independent instances of the game.

Theorem 8.2 (Non-parametric estimation of $\tilde{\Delta}$ with bounded regressors)**.** *Under the same conditions as part 1 of theorem 8.1,*

$$\hat{\tilde{\Delta}}_M = \text{sgn}\left(1\left\{M^{\frac{1}{4}}\left|\frac{P_M(y_1=1,y_2=1)}{P_M(y_1=1)P_M(y_2=1)}-1\right|\geq 1\right\}\log\left(\frac{P_M(y_1=1,y_2=1)}{P_M(y_1=1)P_M(y_2=1)}\right)\right) \to^{a.s.} \tilde{\Delta}.$$

The indicator is used because when $\tilde{\Delta}=0$ generally $\text{sgn}\left(\log\left(\frac{P_M(y_1=1,y_2=1)}{P_M(y_1=1)P_M(y_2=1)}\right)\right)\neq 0$, since $\text{sgn}(\cdot)$ is discontinuous at 0. This is similar to tests for moment inequalities in Andrews and Soares (2010) and Kline (2011), where it is necessary to know which moment inequalities bind. Since the support of $\hat{\tilde{\Delta}}$ is finite, with probability one, in large enough samples, $\hat{\tilde{\Delta}}=\tilde{\Delta}$. So, the rate of convergence is arbitrarily fast. Therefore, estimation of $\tilde{\Delta}$ has no asymptotic effect on subsequent steps of estimation.

9. Appendix: proofs

Proof of theorem 3.1. *Proof of identification of β:*

By assumption 3.3, $F_\epsilon(t_1,t_2)=\int_{-\infty}^{t_1}\int_{-\infty}^{t_2} f_\epsilon(e_1,e_2)de_2de_1$. Use the notation that $F_\epsilon^{[i,1]}(t_1,t_2)$ is the first derivative of $F_\epsilon(\cdot)$ with respect to t_i, evaluated at $t=(t_1,t_2)$. So, $F_\epsilon^{[i=1,1]}(t_1,t_2)=\int_{-\infty}^{t_2} f_{\epsilon_1,\epsilon_2}(t_1,e_2)de_2$ and $F_\epsilon^{[i=2,1]}(t_1,t_2)=\int_{-\infty}^{t_1} f_{\epsilon_1,\epsilon_2}(e_1,t_2)de_1$. Since the density is strictly positive everywhere, it holds that $0<F_\epsilon^{[i,1]}(t_1,t_2)<f_{\epsilon_i}(t_i)$.

Suppose that $\alpha_i+x_i\beta_{ix}+w\beta_{iw}+\epsilon_i<0$ for all i. Then, the outcome $(0,0)$ is a pure strategy Nash equilibrium. Moreover, since $\Delta_1\leq 0$ and $\Delta_2\leq 0$, it is the

[9] This discussion is focused only on proposition 1 of de Paula and Tang (2012); their paper includes many other results on related issues. In particular, their corollary 1 and proposition 2 deals with issues related to the uniqueness of Bayesian Nash equilibrium. However, neither of those results apply in the case of a zero interaction effect.

unique pure strategy Nash equilibrium, since $u_i(1, y_{(-i)}) \leq \alpha_i + x_i\beta_{ix} + w\beta_{iw} + \epsilon_i < 0$. Conversely, suppose that $(0,0)$ is a pure strategy Nash equilibrium. Then, it must be that $\alpha_i + x_i\beta_{ix} + w\beta_{iw} + \epsilon_i \leq 0$. Therefore, since by assumption 3.3 there is zero probability that $\alpha_i + x_i\beta_{ix} + w\beta_{iw} + \epsilon_i = 0$ conditional on any z, and using assumption 3.2, $P(y = (0,0)|z) = P(\epsilon_1 \leq -\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, \epsilon_2 \leq -\alpha_2 - x_2\beta_{2x} - w\beta_{2w})$.

Then, using assumption 3.3, note that $\frac{\partial P(y=(0,0)|z)}{\partial x_{ik}} = F_\epsilon^{[i,1]}(-\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w})(-\beta_{ixk})$. Also note that, using the chain rule for total derivatives, $\frac{\partial P(y=(0,0)|z)}{\partial w_l} = \sum_i^2 \left(F_\epsilon^{[i,1]}(-\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w})(-\beta_{iwl})\right)$.

Since z has an ordinary density by assumption 3.4, the derivatives on the left hand sides of these expressions are observed in the population. Use the notation that $F_\epsilon^{[i,1]}(z,\theta) \equiv F_\epsilon^{[i,1]}(-\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w})$.

Then, since $\beta_{ix1} = 1$ by assumption 3.1, $\beta_{ixk} = \frac{\frac{\partial P(y=(0,0)|z)}{\partial x_{ik}}}{\frac{\partial P(y=(0,0)|z)}{\partial x_{i1}}}$. Further, assuming that $\beta_{1w} = \beta_{2w}$, $\beta_{wl} = \frac{\frac{\partial P(y=(0,0)|z)}{\partial w_l}}{\sum_i^2 \frac{\partial P(y=(0,0)|z)}{\partial x_{i1}}}$. The denominators in these expressions are strictly negative by assumption 3.3.

Also, for any weight function $\pi(\cdot)$, and assuming that the expectations exist, $E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial x_{ik}}\right) = E\left(\pi(z)F_\epsilon^{[i,1]}(z,\theta)\right)(-\beta_{ixk})$. If $\beta_{1w} = \beta_{2w}$, $E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial w_l}\right) = E\left(\pi(z)\sum_i^2 F_\epsilon^{[i,1]}(z,\theta)\right)(-\beta_{wl})$. Therefore, it holds that $\beta_{ixk} = \frac{E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial x_{ik}}\right)}{E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial x_{i1}}\right)}$ and $\beta_{wl} = \frac{E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial w_l}\right)}{\sum_i^2 E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial x_{i1}}\right)}$. In order for the division to be justified, it must be that $E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial x_{i1}}\right)$ is non-zero. It is, because $\frac{\partial P(y=(0,0)|z)}{\partial x_{i1}} < 0$ for all z, as noted above, and $\pi(z)$ is non-negative for all z, and is strictly positive on a set of z that has positive measure. Consequently, $E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial x_{i1}}\right) < 0$.

Alternatively, if $E(P_x(z))$ has full rank, let $F_\epsilon^{[1]}(z,\theta)$ be the 1×2 matrix whose ith entry is $F_\epsilon^{[i,1]}(-\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w})$. Then, it follows that $F_\epsilon^{[1]}(z,\theta)'\frac{dP(y=(0,0)|z)}{dw_l} = F_\epsilon^{[1]}(z,\theta)'F_\epsilon^{[1]}(z,\theta)(-\beta_{wl})$ where β_{wl} is the 2×1 matrix whose ith entry is β_{iwl}. So, taking expectations and using assumption 3.6,

$$\beta_{wl} = -\left(E\left(F_\epsilon^{[1]}(z,\theta)'F_\epsilon^{[1]}(z,\theta)\right)\right)^{-1} E\left(F_\epsilon^{[1]}(z,\theta)'\frac{dP(y=(0,0)|z)}{dw_l}\right).$$

Then, let $P^{[1]}(z)$ be the 1×2 matrix whose ith entry is $\frac{\partial P(y=(0,0)|z)}{\partial x_{i1}}$. And recall from above that $-F_\epsilon^{[i,1]}(-\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}) = \frac{\partial P(y=(0,0)|z)}{\partial x_{i1}}$, so $F_\epsilon^{[1]}(z,\theta) = -P^{[1]}(z)$.

Proof of identification of α and Δ:

As above, $P(y=(0,0)|z) = P(\epsilon_1+\alpha_1 \leq -x_1\beta_{1x} - w\beta_{1w}, \epsilon_2+\alpha_2 \leq -x_2\beta_{2x} - w\beta_{2w})$ and similarly $P(y=(1,1)|z) = P(\epsilon_1+\alpha_1+\Delta_1 \geq -x_1\beta_{1x} - w\beta_{1w}, \epsilon_2+\alpha_2+\Delta_2 \geq -x_2\beta_{2x} - w\beta_{2w})$. Thus, $f_{\epsilon_1+\alpha_1,\epsilon_2+\alpha_2}(t_1,t_2) = \left.\frac{\partial^2 P(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2}\right|_{t_1,t_2}$ is point identified at points (t_1,t_2) such that the density of (c_1,c_2) is positive on an open set that contains (t_1,t_2). Therefore by assumption 3.5 this second cross partial derivative is identified on an open set of (t_1,t_2) that contains (α_1,α_2). By assumption 3.3, the mode of $(\epsilon_1+\alpha_1,\epsilon_2+\alpha_2)$ is (α_1,α_2). Therefore, $(\alpha_1,\alpha_2) = \arg\max_{a_1,a_2} \left.\frac{\partial^2 P(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2}\right|_{a_1,a_2}$. Similarly, $(\alpha_1+\Delta_1,\alpha_2+\Delta_2) = \arg\max_{b_1,b_2} \left.\frac{\partial^2 P(y=(1,1)|c_1,c_2)}{\partial c_1 \partial c_2}\right|_{b_1,b_2}$. The maximization is over the support of (c_1,c_2). □

Lemma 9.1. *Under the same conditions as theorem 3.1, if the densities of ϵ_1 and ϵ_2 are bounded above, and $\pi(\cdot)$ is integrable with respect to the data generating process for z, then the expectations appearing in the statement of theorem 3.1 all exist.*

Proof of lemma 9.1. For $E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial x_{ik}}\right)$, note that by the arguments of theorem 3.1, $\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial x_{ik}} = \left(\pi(z)F_\epsilon^{[i,1]}(z,\theta)\right)(-\beta_{ixk})$. Also, if there is an upper bound $\overline{f}$ for the densities of ϵ_1 and ϵ_2, then also by the arguments of theorem 3.1, $F_\epsilon^{[i,1]}(z,\theta) \leq \overline{f}$. Consequently, $\left(\pi(z)F_\epsilon^{[i,1]}(z,\theta)\right)(-\beta_{ixk})$ is bounded above by a constant multiple of $\pi(z)$, so is integrable as long as $\pi(z)$ is integrable.

It is similar for $E\left(\pi(z)\frac{\partial P(y=(0,0)|z)}{\partial w_l}\right)$. And for $E(P_x(z))$ and $E\left(P^{[1]}(z)'\frac{dP(y=(0,0)|z)}{dw_l}\right)$, note similarly that the integrand is bounded above by a constant if there is an upper bound $\overline{f}$ for the densities of ϵ_1 and ϵ_2, and therefore is integrable. □

Proof of theorem 4.1. Suppose that $\Delta_1 \leq 0$ and $\Delta_2 \leq 0$. Then, inspecting the condition for $y_2 = 1$ to be part of a pure strategy Nash equilibrium, and using assumptions 3.2 and 3.3,

$$
\begin{aligned}
P(y_2=1|z) = {} & P(\epsilon_2 \geq -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2) \\
& + P(\epsilon_1 \leq -\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} \leq \epsilon_2 \leq -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2) \\
& + P(y_2=1|\epsilon \in \mathcal{R}^-(z,\theta), z)P(\epsilon \in \mathcal{R}^-(z,\theta)),
\end{aligned}
$$

where the last term corresponds to the region of multiple equilibria, and the conditioning on z in $P(\epsilon \in \mathcal{R}^-(z,\theta))$ can be dropped by assumption 3.2. Let $P_2(z) \equiv P(y_2=1|\epsilon \in \mathcal{R}^-(z,\theta), z)$ be the selection mechanism.

Therefore, using assumption 3.4, and letting $c_i = -\alpha_i - x_i\beta_{ix} - w\beta_{iw}$,

$$\begin{aligned}
\frac{\partial P(y_2=1|z)}{\partial x_{11}} &= -\beta_{1x1}\int_{c_2}^{c_2-\Delta_2} f_{\epsilon_1,\epsilon_2}(c_1,e_2)de_2 \\
&+ P_2(z)\left(-\beta_{1x1}\int_{c_2}^{c_2-\Delta_2} f_{\epsilon_1,\epsilon_2}(c_1-\Delta_1,e_2)de_2 + \beta_{1x1}\int_{c_2}^{c_2-\Delta_2} f_{\epsilon_1,\epsilon_2}(c_1,e_2)de_2\right) \\
&+ \frac{\partial P_2(z)}{\partial x_{11}} P(\epsilon\in\mathcal{R}^-(z,\theta)) \\
&= (1-P_2(z))\left(-\beta_{1x1}\int_{c_2}^{c_2-\Delta_2} f_{\epsilon_1,\epsilon_2}(c_1,e_2)de_2\right) \\
&+ P_2(z)\left(-\beta_{1x1}\int_{c_2}^{c_2-\Delta_2} f_{\epsilon_1,\epsilon_2}(c_1-\Delta_1,e_2)de_2\right) + \frac{\partial P_2(z)}{\partial x_{11}} P(\epsilon\in\mathcal{R}^-(z,\theta)).
\end{aligned}$$

Note that $\beta_{1x1} = 1$. If $\Delta_2 < 0$, then the integration bounds are not equal, so by assumption 3.3, the integrals in this expression are strictly positive. Obviously, $0 \le P_2(z) \le 1$. Finally, by assumption 4.1, $\frac{\partial P_2(z)}{\partial x_{11}} \le 0$, since in the region of multiple equilibria, $y_2 = 1$ happens exactly in case $y_1 = 0$. So, this expression is strictly negative. Similarly, if $\Delta_1 < 0$, then $\frac{\partial P(y_1=1|z)}{\partial x_{21}} < 0$.

Alternatively, suppose that $\Delta_1 \ge 0$ and $\Delta_2 \ge 0$. By symmetric arguments, if $\Delta_2 > 0$, then $\frac{\partial P(y_2=1|z)}{\partial x_{11}}$ is strictly positive; and if $\Delta_1 > 0$, then $\frac{\partial P(y_1=1|z)}{\partial x_{21}}$ is strictly positive.

Suppose that $\Delta_2 = 0$. Then $P(y_2 = 1|z) = P(\epsilon_2 \ge -\alpha_2 - x_2\beta_{2x} - w\beta_{2w})$, so $\frac{\partial P(y_2=1|z)}{\partial x_{11}} = 0$. Similarly, if $\Delta_1 = 0$, then $\frac{\partial P(y_1=1|z)}{\partial x_{21}} = 0$

Since these results hold pointwise in z, they also hold for weighted averages. □

Proof of theorem 5.1. By Powell, Stock, and Stoker (1989), $\sqrt{M}(\hat\delta_M - \delta) \to^d N(0, V(\delta))$. So, by the delta method, the asymptotic covariance is $C(\delta)V(\delta)C(\delta)'$. □

Proof of theorem 5.2. By a Taylor series approximation, which exists by assumption 5.7, $\widehat{\Delta Q}_M(\gamma) = \left(\left.\frac{\partial \tilde R_{M,1}(a_1,a_2,B)}{\partial B}\right|_{\tilde\beta_1}\left(\beta-\hat\beta\right) \quad \left.\frac{\partial \tilde R_{M,2}(b_1,b_2,B)}{\partial B}\right|_{\tilde\beta_2}\left(\beta-\hat\beta\right)\right)'$ where $||\tilde\beta_1 - \beta|| \le ||\hat\beta - \beta||$ and $||\tilde\beta_2 - \beta|| \le ||\hat\beta - \beta||$. So, $\sup_\gamma \left\|\widehat{\Delta Q}_M(\gamma)\right\| \lesssim \max\left(\sup_\gamma\left\|\left.\frac{\partial \tilde R_{M,1}(a_1,a_2,B)}{\partial B}\right|_{\tilde\beta_1}\right\| \left\|\beta-\hat\beta\right\|, \sup_\gamma\left\|\left.\frac{\partial \tilde R_{M,2}(b_1,b_2,B)}{\partial B}\right|_{\tilde\beta_2}\right\| \left\|\beta-\hat\beta\right\|\right) = O_p(M^{-\frac{1}{2}})$ by assumption 5.7, where $\lesssim$ means less than or equal up to a positive constant. Therefore, by assumption 5.6, $\hat Q_M(\gamma)$ converges in probability to $Q(\gamma)$ uniformly over Γ. Consequently, because of assumptions 3.3 and 5.5 and the conclusion of theorem 3.1, Newey and McFadden (1994, Theorem 2.1) applies, so $\hat\gamma_M \to^p \gamma_0$. (Using the notation that γ_0 is the true parameter, to distinguish between the use of "γ" as the argument of a function.)

The proof of asymptotic normality is given for the case where it is assumed that $\alpha_1 \equiv \alpha \equiv \alpha_2$ and $\Delta_1 \equiv \Delta \equiv \Delta_2$, which implicitly entails deriving the asymptotic covariance without that assumption.

Since the parameters are assumed in estimation to be equal across agent roles, with corresponding constrained objective function $Q^r(\gamma^r) = Q(\gamma_1^r, \gamma_1^r, \gamma_2^r, \gamma_2^r)$, where $\gamma^r = (a, b)$ constrains the parameters to be equal across agent roles,

$$\frac{\partial \hat{Q}_M^r(\gamma)}{\partial \gamma} = \begin{pmatrix} \frac{\partial^3 P_M(y=(0,0)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1^2 \partial \hat{c}_2}\Big|_{a,a} + \frac{\partial^3 P_M(y=(0,0)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1 \partial \hat{c}_2^2}\Big|_{a,a} \\ \frac{\partial^3 P_M(y=(1,1)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1^2 \partial \hat{c}_2}\Big|_{b,b} + \frac{\partial^3 P_M(y=(1,1)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1 \partial \hat{c}_2^2}\Big|_{b,b} \end{pmatrix}$$

and $\frac{\partial^2 \hat{Q}_M^r(\gamma)}{\partial \gamma \partial \gamma'} = \begin{pmatrix} \frac{\partial^2 \hat{Q}_M^r(\gamma)}{\partial a^2} & 0 \\ 0 & \frac{\partial^2 \hat{Q}_M^r(\gamma)}{\partial b^2} \end{pmatrix}$, where

$$\frac{\partial^2 \hat{Q}_M^r(\gamma)}{\partial a^2} = \frac{\partial^4 P_M(y=(0,0)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1^3 \partial \hat{c}_2}\Big|_{a,a} + 2\frac{\partial^4 P_M(y=(0,0)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1^2 \partial \hat{c}_2^2}\Big|_{a,a} + \frac{\partial^4 P_M(y=(0,0)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1 \partial \hat{c}_2^3}\Big|_{a,a}$$

and

$$\frac{\partial^2 \hat{Q}_M^r(\gamma)}{\partial^2 b} = \frac{\partial^4 P_M(y=(1,1)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1^3 \partial \hat{c}_2}\Big|_{b,b} + 2\frac{\partial^4 P_n(y=(1,1)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1^2 \partial \hat{c}_2^2}\Big|_{b,b} + \frac{\partial^4 P_M(y=(1,1)|\hat{c}_1,\hat{c}_2)}{\partial \hat{c}_1 \partial \hat{c}_2^3}\Big|_{b,b}.$$

By a Taylor series approximation, which exists by assumption 5.13, and since $\hat{\gamma}_M$ solves the first order condition by assumption 5.8, with probability approaching 1, $0 = \frac{\partial \hat{Q}_M^r(\gamma)}{\partial \gamma}\Big|_{\hat{\gamma}_M} = \frac{\partial \hat{Q}_M^r(\gamma)}{\partial \gamma}\Big|_{\gamma_0} + \frac{\partial^2 \hat{Q}_M^r(\gamma)}{\partial \gamma \partial \gamma'}\Big|_{\tilde{\gamma}_M} (\hat{\gamma}_M - \gamma_0)$. So, $r_M \left(\left(- \frac{\partial \hat{Q}_M^r(\gamma)}{\partial \gamma}\Big|_{\gamma_0} + \frac{\partial Q_M^r(\gamma)}{\partial \gamma}\Big|_{\gamma_0} \right) - \frac{\partial Q_M^r(\gamma)}{\partial \gamma}\Big|_{\gamma_0} \right) = r_M \left(\left(\frac{\partial^2 \hat{Q}_M^r(\gamma)}{\partial \gamma \partial \gamma'}\Big|_{\tilde{\gamma}_M} - \frac{\partial^2 Q_M^r(\gamma)}{\partial \gamma \partial \gamma'}\Big|_{\tilde{\gamma}_M} \right) (\hat{\gamma}_M - \gamma_0) + \frac{\partial^2 Q_M^r(\gamma)}{\partial \gamma \partial \gamma'}\Big|_{\tilde{\gamma}_M} (\hat{\gamma}_M - \gamma_0) \right)$. By assumption 5.11, and arguments similar to the above, $\left(- \frac{\partial \hat{Q}_M^r(\gamma)}{\partial \gamma}\Big|_{\gamma_0} + \frac{\partial Q_M^r(\gamma)}{\partial \gamma}\Big|_{\gamma_0} \right) = O_p(M^{-\frac{1}{2}})$ and $\left(\frac{\partial^2 \hat{Q}_M^r(\gamma)}{\partial \gamma \partial \gamma'}\Big|_{\tilde{\gamma}_M} - \frac{\partial^2 Q_M^r(\gamma)}{\partial \gamma \partial \gamma'}\Big|_{\tilde{\gamma}_M} \right) = O_p(M^{-\frac{1}{2}})$.

By assumption 5.10, $\frac{\partial^2 Q_M^r(\gamma)}{\partial \gamma \partial \gamma'}\Big|_{\tilde{\gamma}_M} \to^p \frac{\partial^2 Q^r(\gamma)}{\partial \gamma \partial \gamma'}\Big|_{\gamma_0}$, where $\frac{\partial^2 Q^r(\tau)}{\partial \gamma \partial \gamma'} = \begin{pmatrix} \frac{\partial^2 Q^r(\gamma)}{\partial a^2} & 0 \\ 0 & \frac{\partial^2 Q^r(\gamma)}{\partial b^2} \end{pmatrix}$, where $\frac{\partial^2 Q^r(\gamma)}{\partial a^2} = \frac{\partial^4 P(y=(0,0)|c_1,c_2)}{\partial c_1^3 \partial c_2}\Big|_{a,a} + 2\frac{\partial^4 P(y=(0,0)|c_1,c_2)}{\partial c_1^2 \partial c_2^2}\Big|_{a,a} + \frac{\partial^4 P(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2^3}\Big|_{a,a}$ and $\frac{\partial^2 Q^c(\gamma)}{\partial b^2} = \frac{\partial^4 P(y=(1,1)|c_1,c_2)}{\partial c_1^3 \partial c_2}\Big|_{b,b} + 2\frac{\partial^4 P(y=(1,1)|c_1,c_2)}{\partial c_1^2 \partial c_2^2}\Big|_{b,b} + \frac{\partial^4 P(y=(1,1)|c_1,c_2)}{\partial c_1 \partial c_2^3}\Big|_{b,b}$.

The Hessian of the density of $\epsilon + \alpha$ evaluated at α is

$$\begin{pmatrix} \frac{\partial^4 P(y=(0,0)|c_1,c_2)}{\partial c_1^3 \partial c_2}\Big|_{\alpha,\alpha} & \frac{\partial^4 P(y=(0,0)|c_1,c_2)}{\partial c_1^2 \partial c_2^2}\Big|_{\alpha,\alpha} \\ \frac{\partial^4 P(y=(0,0)|c_1,c_2)}{\partial c_1^2 \partial c_2^2}\Big|_{\alpha,\alpha} & \frac{\partial^4 P(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2^3}\Big|_{\alpha,\alpha} \end{pmatrix}.$$

Therefore assumption 5.12 implies that $\frac{\partial^4 P(y=(0,0)|c_1,c_2)}{\partial c_1^3 \partial c_2}\Big|_{\alpha,\alpha} + 2\frac{\partial^4 P(y=(0,0)|c_1,c_2)}{\partial c_1^2 \partial c_2^2}\Big|_{\alpha,\alpha} + \frac{\partial^4 P(y=(0,0)|c_1,c_2)}{\partial c_1 \partial c_2^3}\Big|_{\alpha,\alpha} < 0$. By similar arguments for the Hessian of the density of $\epsilon +$

$\alpha + \Delta$ evaluated at α and Δ, $\frac{\partial^4 P(y=(1,1)|c_1,c_2)}{\partial c_1^3 \partial c_2}\Big|_{\alpha+\Delta,\alpha+\Delta} + 2\,\frac{\partial^4 P(y=(1,1)|c_1,c_2)}{\partial c_1^2 \partial c_2^2}\Big|_{\alpha+\Delta,\alpha+\Delta} + \frac{\partial^4 P(y=(1,1)|c_1,c_2)}{\partial c_1 \partial c_2^3}\Big|_{\alpha+\Delta,\alpha+\Delta} < 0$. Therefore, with probability approaching 1, $\frac{\partial^2 Q_M^r(\gamma)}{\partial\gamma\partial\gamma'}\Big|_{\tilde{\gamma}_M}$ is invertible.

Therefore, $r_M(\hat{\gamma}_M - \gamma_0) \to^d N\left(0, \left(\frac{\partial^2 Q^r(\gamma)}{\partial\gamma\partial\gamma'}\Big|_{\gamma_0}\right)^{-1} D\Omega_0 D' \left(\frac{\partial^2 Q^r(\gamma)}{\partial\gamma\partial\gamma'}\Big|_{\gamma_0}\right)^{-1}\right)$ by assumption 5.9, where $D = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$. Therefore it follows by the delta method that $r_M(\hat{\psi}_M - \psi) \to^d N\left(0, C\left(\frac{\partial^2 Q^r(\gamma)}{\partial\gamma\partial\gamma'}\Big|_{\gamma_0}\right)^{-1} D\Omega_0 D' \left(\frac{\partial^2 Q^r(\gamma)}{\partial\gamma\partial\gamma'}\Big|_{\gamma_0}\right)^{-1} C'\right)$ where $C = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$. □

Proof of theorem 8.1. It is equivalent to show that $\tilde{\Delta} = \text{sgn}\left(\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)} - 1\right)$.

Suppose that $\tilde{\Delta} = -1$. Since $P(y_1 = 1, y_2 = 1) > 0$, it follows that $P(u_1(1,1) \geq 0, u_2(1,1) \geq 0) > 0$.[10] So, $P(u_1(1,1) \geq 0) > 0$ and $P(u_2(1,1) \geq 0) > 0$. Similarly, since $P(y_1 = 0, y_2 = 0) > 0$, $P(u_1(1,0) \leq 0) > 0$ and $P(u_2(0,1) \leq 0) > 0$. By the assumptions in part 1, $P(u_1(1,0) > 0 > u_1(1,1)) > 0$ and $P(u_2(0,1) > 0 > u_2(1,1)) > 0$. Since $u_1 \perp u_2$, this implies (among other things) that $P(u_1(1,0) > 0 > u_1(1,1), u_2(0,1) < 0) > 0$.

Then, $P(y_1 = 1|y_2 = 1) = \frac{P(y_1=1,y_2=1)}{P(y_2=1)} = \frac{P(u_1(1,1)\geq 0,u_2(1,1)\geq 0)}{P(y_2=1)}$. Further, $P(y_2 = 1) \geq P(u_2(1,1) \geq 0)$, since whenever $u_2(1,1) > 0$, $y_2 = 1$ is a strictly dominant strategy, and $P(u_2(1,1) = 0) = 0$. Therefore, $P(y_1 = 1|y_2 = 1) \leq \frac{P(u_1(1,1)\geq 0,u_2(1,1)\geq 0)}{P(u_2(1,1)\geq 0)} = P(u_1(1,1) \geq 0) \leq P(y_1 = 1)$ since $u_1 \perp u_2$. This implies $\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)} \leq 1$. The inequality is strict under the assumption in part 1, because $P(u_1(1,0) > 0 > u_1(1,1), u_2(0,1) < 0) > 0$, which also results in the Nash equilibrium outcome $y_1 = 1$, so $P(y_1 = 1) > P(u_1(1,1) \geq 0)$.

Suppose that $\tilde{\Delta} = 1$. By symmetric arguments, $\frac{P(y_1=0,y_2=1)}{P(y_1=0)P(y_2=1)} < (\leq) 1$, where the inequality is strict or weak depending on whether the assumption in part 1 is maintained. This is equivalent to $P(y_1 = 0, y_2 = 1) < (\leq) P(y_2 = 1) - P(y_1 = 1)P(y_2 = 1)$, which is equivalent to $\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)} > (\geq) 1$.

Suppose that $\tilde{\Delta} = 0$. Then, $P(y_1 = 1, y_2 = 1) = P(u_1(1,0) \geq 0, u_2(0,1) \geq 0)$ and $P(y_1 = 1) = P(u_1(1,0) \geq 0)$ and $P(y_2 = 1) = P(u_2(0,1) \geq 0)$. So, since $u_1 \perp u_2$, this implies that $\frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)} = 1$. □

[10] This is because $P(y_1 = 1, y_2 = 1) = P(u_1(1,1) \geq 0, u_2(1,1) \geq 0)$ since if $(1,1)$ is the pure strategy Nash equilibrium outcome then $u_1(1,1) \geq 0$ and $u_2(1,1) \geq 0$. Conversely, if $u_1(1,1) > 0$ and $u_2(1,1) > 0$, then $(1,1)$ is the unique pure strategy Nash equilibrium. The event that $u_1(1,1) = 0$ or $u_2(1,1) = 0$ has zero probability by assumption.

Proof of theorem 8.2. By the law of large numbers, $\frac{P_M(y_1=1,y_2=1)}{P_M(y_1=1)P_M(y_2=1)} \to^{a.s} \frac{P(y_1=1,y_2=1)}{P(y_1=1)P(y_2=1)}$. So, if $\tilde{\Delta} \neq 0$, then $\frac{P_M(y_1=1,y_2=1)}{P_M(y_1=1)P_M(y_2=1)} - 1$ converges almost surely to a non-zero number, so the left hand side of the argument in the indicator function converges to $+\infty$ almost surely. Therefore, w.p.1, the indicator function is 1 for sufficiently large sample size. Thus, $\hat{\tilde{\Delta}} \to^{a.s.} \tilde{\Delta}$ when $\tilde{\Delta} \neq 0$. Alternatively, supposing that $\tilde{\Delta} = 0$,

$$1\left\{M^{\frac{1}{4}}\left|\frac{P_M(y_1=1,y_2=1)}{P_M(y_1=1)P_M(y_2=1)} - 1\right| \geq 1\right\} = 1\left\{M^{\frac{1}{4}}\left|P_M(y_1=1,y_2=1) - P_M(y_1=1)P_M(y_2=1)\right| \geq P_M(y_1=1)P_M(y_2=1)\right\}$$
$$= 1\left\{M^{\frac{1}{4}}\left|P_M(y_1=1,y_2=1) - P(y_1=1,y_2=1) + P(y_1=1)P(y_2=1) - P_M(y_1=1)P_M(y_2=1)\right| \geq P_M(y_1=1)P_M(y_2=1)\right\}$$
$$= 1\left\{M^{\frac{1}{4}}\left|P_M(y_1=1,y_2=1) - P(y_1=1,y_2=1) + (P(y_1=1) - P_M(y_1=1))P_M(y_2=1)\right.\right.$$
$$\left.\left.+(P(y_2=1) - P_M(y_2=1))P(y_1=1)\right| \geq P_M(y_1=1)P_M(y_2=1)\right\}$$
$$\leq 1\left\{M^{\frac{1}{4}}\left|P_M(y_1=1,y_2=1) - P(y_1=1,y_2=1)\right| + M^{\frac{1}{4}}\left|P(y_1=1) - P_M(y_1=1)\right| P_M(y_2=1)\right.$$
$$\left.+M^{\frac{1}{4}}\left|P(y_2=1) - P_M(y_2=1)\right| P(y_1=1) \geq P_M(y_1=1)P_M(y_2=1)\right\}.$$

The second equality follows from $\tilde{\Delta} = 0$ implies that $P(y_1 = 1, y_2 = 1) = P(y_1 = 1)P(y_2 = 1)$. By the Marcinkiewicz and Zygmund (1937) strong law of large numbers, $M^{\frac{1}{4}}(P_M(y_1 = 1, y_2 = 1) - P(y_1 = 1, y_2 = 1)) \to^{a.s} 0$, $M^{\frac{1}{4}}(P(y_1 = 1) - P_M(y_1 = 1)) \to^{a.s} 0$, and $M^{\frac{1}{4}}(P(y_2 = 1) - P_M(y_2 = 1)) \to^{a.s.} 0$. Therefore, since also $P_M(y_1 = 1)P_M(y_2 = 1) \to^{a.s.} P(y_1 = 1)P(y_2 = 1) \neq 0$, the entire indicator function converges almost surely to 0. This means that w.p.1, for sufficiently large sample size, the indicator function is 0. Therefore, $\hat{\tilde{\Delta}} \to^{a.s.} \tilde{\Delta}$. □

References

ANDREWS, D. W. (1995): "Nonparametric kernel estimation for semiparametric models," *Econometric Theory*, 11(3), 560–596.

ANDREWS, D. W., AND M. M. SCHAFGANS (1998): "Semiparametric estimation of the intercept of a sample selection model," *The Review of Economic Studies*, 65(3), 497–517.

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

ARADILLAS-LOPEZ, A., AND A. M. ROSEN (2013): "Inference in ordered response games with complete information," Discussion paper, working paper.

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

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

BIERENS, H. J. (1987): "Kernel estimators of regression functions," in *Advances in Econometrics: Fifth World Congress*, ed. by T. F. Bewley, vol. I, pp. 99–144. Cambridge University Press.

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

——— (1991a): "Empirical models of discrete games," *Journal of Econometrics*, 48(1), 57–81.

——— (1991b): "Entry and competition in concentrated markets," *Journal of Political Economy*, pp. 977–1009.

Chamberlain, G. (1986): “Asymptotic efficiency in semi-parametric models with censoring,” *Journal of Econometrics*, 32(2), 189–218.

Cosslett, S. (1987): “Efficiency bounds for distribution-free estimators of the binary choice and the censored regression models,” *Econometrica*, pp. 559–585.

Dunker, F., S. Hoderlein, AND H. Kaido (2013): “Random coefficients in static games of complete information,” Discussion paper, working paper.

Fox, J. T., AND N. Lazzati (2013): “Identification of discrete choice models for bundles and binary games,” Discussion paper, working paper.

French, E., AND C. Taber (2011): “Identification of models of the labor market,” *Handbook of Labor Economics*, 4, 537–617.

Horowitz, J. L. (2009): *Semiparametric and Nonparametric Methods in Econometrics*. Springer Series in Statistics.

Khan, S., AND D. Nekipelov (2012): “Information structure and statistical information in discrete response models,” working paper.

Khan, S., AND E. Tamer (2010): “Irregular identification, support conditions, and inverse weight estimation,” *Econometrica*, 78(6), 2021–2042.

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.

——— (2013): “Identification of complete information games,” working paper.

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

Lewbel, A. (2000): “Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables,” *Journal of Econometrics*, 97(1), 145–177.

Magnac, T., AND E. Maurin (2007): “Identification and information in monotone binary models,” *Journal of Econometrics*, 139(1), 76–104.

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

Marcinkiewicz, J., AND A. Zygmund (1937): “Sur les Fonctions Indépendantes,” *Fundamenta Mathematicae*, 29, 60–90.

Newey, W. K., AND D. McFadden (1994): “Large sample estimation and hypothesis testing,” in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. IV, chap. 36, pp. 2111–2245. Elsevier Science B.V.

Pagan, A., AND A. Ullah (1999): *Nonparametric Econometrics*. Cambridge University Press.

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.

Powell, J., J. Stock, AND T. Stoker (1989): “Semiparametric estimation of index coefficients,” *Econometrica*, pp. 1403–1430.

Powell, J. L., AND T. M. Stoker (1996): “Optimal bandwidth choice for density-weighted averages,” *Journal of Econometrics*, 75(2), 291–316.

Samarov, A. M. (1993): “Exploring regression structure using nonparametric functional estimation,” *Journal of the American Statistical Association*, 88(423), 836–847.

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