

# The empirical content of games with bounded regressors

BRENDAN KLINE

Department of Economics, University of Texas at Austin

This paper develops a strategy for identification and estimation of complete information games that does not require a regressor that has large support or a parametric specification for the distribution of the unobservables. The identification result uses a nonstandard but plausible condition on the unobservables: the assumption that the joint density of the unobservables of all agents is unimodal in the sense of achieving the global maximum at a unique point. Also, a three-step semiparametric estimator is proposed. Under mild regularity conditions, the estimator is consistent and asymptotically normally distributed. The estimator is nonstandard in the sense that the estimators of the intercept and interaction effect parameters converge at slower than the parametric rate. An intermediate result concerns identification and estimation of the direction of the interaction effect.

**KEYWORDS.** Identification, nonstandard estimation, strategic interaction, entry game.

**JEL CLASSIFICATION.** C14, C31, C57, C72, L13.

## 1. INTRODUCTION

### 1.1 *Overview*

It is difficult to establish conditions under which the parameters of a model of a complete information game are point identified, particularly because of problems related to multiple equilibria. Nevertheless, Tamer (2003), Bajari, Hong, and Ryan (2010), Dunker, Hoderlein, and Kaido (2013), Fox and Lazzati (2013), and Kline (2015a), provide various sufficient conditions for point identification.

All those results are based on identification strategies that require a large support regressor, although the way that such a regressor is used may vary across papers.<sup>1</sup> If there are no results that apply without a large support regressor, then point identification in

---

Brendan Kline: [brendan.kline@austin.utexas.edu](mailto:brendan.kline@austin.utexas.edu)

I thank Joel Horowitz, Charles Manski, and Elie Tamer for many helpful comments and discussions. I also thank audiences at the 2014 North American Summer Meeting of the Econometric Society, Texas Econometrics Camp 2014, Texas A&M University, the University of Arizona, and Yale University for helpful comments. Finally, I thank the editor and three anonymous referees for helpful comments and suggestions. Any errors are mine.

<sup>1</sup>Bajari, Hong, and Ryan (2010) give an identification result using only exclusions restrictions, assuming a known distribution of the unobservables, and show that a certain necessary condition for identification is satisfied. Similarly, the seminal earlier work of Bresnahan and Reiss (1990, 1991a, 1991b) does not contain any formal identification results, but is suggestive (based on, for example, the empirical exercises)

Copyright © 2016 Brendan Kline. Licensed under the [Creative Commons Attribution-NonCommercial License 3.0](https://creativecommons.org/licenses/by-nc/3.0/). Available at <http://www.qeconomics.org>.

DOI: 10.3982/QE444

empirical applications without large support regressors may be in question. So this paper shows point identification without a large support regressor and without parametric distributional assumptions, based on a novel location assumption concerning the mode of the unobservables.<sup>2</sup> In establishing this result, this paper uses the important fact that the complete information game framework has nontrivial implications for the data for all values of the explanatory variables, not just “extreme” values of the explanatory variables associated with large support: Nash equilibrium implies some possibly set-valued, but generally nontrivial, restrictions on the outcome of the game, as a mapping from the utility functions to the outcomes, regardless of the “magnitude” of the explanatory variables.<sup>3</sup>

Specifically, the identification strategy is based on 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, even if other outcomes of the game occur as part of a region of multiple equilibria.

The identification strategy uses two main assumptions concerning the unobservables. The first assumption is the standard assumption that the unobservables are independent of the exogenous explanatory variables. The second assumption is the location assumption, which is nonstandard: the assumption that the joint density of the unobservables of all agents is unimodal, in the sense of achieving the global maximum at a unique point. It is allowed that the density has many local maxima.

Moreover, a three-step semiparametric estimator is proposed, and the asymptotic properties are derived. The three steps of the estimator are (i) estimate the direction of

---

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).

<sup>2</sup>Subsequent to the results of this paper, these results (and/or “intuition” for the identification strategy) have already proved useful. Fox and Lazzati (2014), among other contributions, use some of the ideas of the current paper in the context of choice over bundles. Also, subsequently, Zhou (2013) studies some implications of the assumption of “radially symmetric” unobservables, particularly as concerns the rate of convergence for the interaction effect parameters, somewhat similarly to how the strength of symmetry assumptions can be sufficient for positive Fisher information for the intercept in single-agent discrete choice models (e.g., Cosslett (1987)).

<sup>3</sup>Suppose that the parameters of the utility function are specified at known values (i.e., the “theory model” is specified), but the nuisance parameters that are not restricted by the theory are unknown (i.e., the selection mechanism). Because the selection mechanism is a distribution directly over the outcomes, the concern is that there might not be any empirical implications unique to that specification of the theory model, or at least that it is difficult to characterize the restrictions on the data implied by the model, complicating identification of the model. This has led to identification strategies based on large support regressors, which “eliminate” the role of the selection mechanism, in which case the restrictions on the data implied by the model are more straightforward to characterize. In this paper, establishing point identification critically depends on establishing the fact that the theory model implies nontrivial restrictions on the data even without large support regressors. In some econometric models, at least the empirical implications of the model are relatively straightforward to characterize, even if establishing identification is not. In models of games, it tends to be difficult even to characterize (in a useful way) the empirical implications of the model.

the interaction effect (i.e., the sign of the interaction effect parameter  $\text{sgn}(\Delta)$ ), (ii) estimate the slope parameters (i.e., the coefficients on the exogenous explanatory variables  $\beta$ ), and then (iii) estimate the “intercept” parameters (i.e., the interaction effect parameters, and the usual intercepts  $\Delta$  and  $\alpha$ ). The interaction effect parameter is an intercept parameter because it does not appear in the model as a slope coefficient on some exogenous explanatory variable.

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 nonparametric 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. Under realistic assumptions the rate is  $M^{1/4}$  (or faster under more smoothness) in two-agent games. The rate of convergence does not depend on the number of explanatory variables in the model, due to a dimension reduction strategy. However, the rate of convergence does depend negatively on the number of agents in the model.

## 1.2 Outline of the paper

Section 2 introduces the model. Section 3 establishes the identification strategy in the context of an  $N$ -agent, two-action game with continuous explanatory variables. Section 4 shows the identification strategy can be extended to cases involving discrete explanatory variables or games involving more than two actions. Section 5 discusses identification of the direction of the interaction effect. Section 6 discusses estimation. Section 7 reports the results of a Monte Carlo experiment. Section 8 reports the results of a stylized empirical application to entry into airline markets. Section 9 concludes, and includes some identification results on the distribution of the unobservables. As is common in semiparametric models, the main results treat the distribution of the unobservables as a “nuisance parameter.” Additional material is available in supplementary files on the journal website, <http://qeconomics.org/supp/444/supplement.pdf> and [http://qeconomics.org/supp/444/code\\_and\\_data.zip](http://qeconomics.org/supp/444/code_and_data.zip).

## 2. MODEL

The complete information game involves  $N \geq 2$  agents. The actions available to each agent are  $S = \{0, 1\}$ . For example, if this is a model of an entry game, the actions are to enter the market (action 1) or not to enter the market (action 0).

The utility functions are

$$\begin{aligned}
 u_{im}(0, y_{-im}) &= 0 \quad \text{and} \\
 u_{im}(1, y_{-im}) &= \alpha_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i \sum_{j \neq i} y_{jm} + \varepsilon_{im}
 \end{aligned} \tag{1}$$

TABLE 1. Normal form.

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

in market  $m$ , where a “market” is the unit of observation. The subscript  $im$  refers to agent  $i$  in market  $m$  and the subscript  $m$  refers to market  $m$ . The normalization that  $u_{im}(0, y_{-im}) = 0$  is used because only differences in utility are relevant. The solution concept is pure strategy Nash equilibrium play.

If  $N = 2$ , the game is described in normal form in Table 1. The row player is agent 1 and the column player is agent 2. The first payoff is the row payoff, and the second payoff is the column payoff.

The explanatory variables  $x_{im}$  are specific to agent  $im$ , and the explanatory variables  $w_m$  are shared among the agents in market  $m$  (for example, market characteristics). It is allowed that there are no shared variables, but there must be at least one agent-specific variable per agent (i.e.,  $x_{im}$  for each agent  $i$  must contain at least one variable), because the agent-specific variables represent an exclusion restriction in the identification of the model with simultaneity. There are  $D_i \geq 1$  agent-specific explanatory variables for agent  $i$ , and there are  $L$  shared explanatory variables.

The variables  $\beta = (\beta_1, \beta_2, \dots, \beta_N)$ , where  $\beta_i = (\beta_{ix}, \beta_{iw})$ , are the slope parameters, which characterize how utility depends on the exogenous explanatory variables. The variables  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$  are the intercept parameters. Finally,  $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_N)$  are the interaction effect parameters, which characterize how utility depends on the actions of the other agents.

It is assumed that the signs of the  $\Delta$  parameters are weakly equal, in the sense that either  $\Delta_i \leq 0$  for all agents  $i$  (e.g., strategic substitutes) or  $\Delta_i \geq 0$  for all agents  $i$  (e.g., strategic complements). Games of strategic substitutes include such important instances as entry games, oligopoly competition in quantity, public good provision, or information provision. Games of strategic complements include such important instances as social interactions, technology adoption, or oligopoly competition in prices. See, for example, [Bulow, Geanakoplos, and Klemperer \(1985\)](#) or [Vives \(2005\)](#). The results for  $N \geq 3$  require that  $\Delta_i \leq 0$  for all agents  $i$ .

The assumption of complete information, rather than incomplete information, follows that part of the literature: in particular, complete information can be “justified” as an approximation to a long-run interaction. Incomplete information generally results in ex post regret (the action of an agent is not the best response to the *realized* action of the other agents), which suggests that equilibrium cannot be maintained in the long run if agents can adjust their action, whereas there is no ex post regret in pure strategy Nash equilibrium with complete information.

In each market the econometrician observes the outcomes  $y = (y_1, y_2, \dots, y_N)$  and the exogenous explanatory variables  $z = (x_1, x_2, \dots, x_N, w)$ , but does not observe  $\varepsilon =$

$(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)$ . The identification problem concerns recovering the parameters  $\theta = (\alpha, \beta, \Delta)$  from the population distribution of the data,  $P(y, z)$ .

### 3. IDENTIFICATION

#### 3.1 *Sketch of identification strategy*

The following subsection sketches the identification strategy for the case of a nonpositive interaction effect, as in an entry game. (The case of a nonnegative interaction effect, as long as  $N = 2$ , is similar, but uses a different set of unique potential outcomes.) This sketch assumes that all explanatory variables have a continuous distribution, an assumption that is relaxed in Section 4.1. Roughly, if there are discrete explanatory variables, then the identification strategy is to point identify the model conditional on values of the discrete explanatory variables.

With a nonpositive interaction effect,  $0_{1 \times N} \equiv (0, 0, \dots, 0)$  is a unique potential outcome<sup>4</sup> because  $0_{1 \times N}$  is the Nash equilibrium if and only if  $\varepsilon_i \leq -\alpha_i - x_i \beta_{ix} - w \beta_{iw}$  for all agents  $i$ . Therefore, observing that the outcome is  $0_{1 \times N}$  is equivalent to that condition on the unobservables. Therefore,

$$\begin{aligned} P(y = 0_{1 \times N} | z) &= P(\{\varepsilon_i \leq -\alpha_i - x_i \beta_{ix} - w \beta_{iw}\} \forall i) \\ &= P_0(-\alpha_1 - x_1 \beta_{1x} - w \beta_{1w}, -\alpha_2 - x_2 \beta_{2x} - w \beta_{2w}, \dots, \\ &\quad -\alpha_N - x_N \beta_{Nx} - w \beta_{Nw}), \end{aligned}$$

where  $P_0(t_1, t_2, \dots, t_N) = P(\varepsilon_1 \leq t_1, \varepsilon_2 \leq t_2, \dots, \varepsilon_N \leq t_N)$ . Let  $N_0 = N - 1$ . Similarly,  $1_{1 \times N} \equiv (1, 1, \dots, 1)$  is the Nash equilibrium if and only if  $\varepsilon_i \geq -\alpha_i - x_i \beta_{ix} - w \beta_{iw} - N_0 \Delta_i$  for all agents  $i$ . Therefore,

$$\begin{aligned} P(y = 1_{1 \times N} | z) &= P(\{\varepsilon_i \geq -\alpha_i - x_i \beta_{ix} - w \beta_{iw} - N_0 \Delta_i\} \forall i) \\ &= P_1(-\alpha_1 - x_1 \beta_{1x} - w \beta_{1w} - N_0 \Delta_1, -\alpha_2 - x_2 \beta_{2x} - w \beta_{2w} - N_0 \Delta_2, \dots, \\ &\quad -\alpha_N - x_N \beta_{Nx} - w \beta_{Nw} - N_0 \Delta_N), \end{aligned}$$

where  $P_1(t_1, t_2, \dots, t_N) = P(\varepsilon_1 \geq t_1, \varepsilon_2 \geq t_2, \dots, \varepsilon_N \geq t_N)$ . This uses the assumption that the unobservables are independent of the explanatory variables.

<sup>4</sup>The observation that there are unique potential outcomes is not the innovation in this paper. Indeed, for example, the same observation is made in [Bresnahan and Reiss \(1990\)](#) and [Ciliberto and Tamer \(2009\)](#). The innovation in this paper is using the existence of unique potential outcomes as an ingredient in a strategy to point identify and estimate the model parameters. Particularly, the innovation is showing that using unique potential outcomes allows identification and estimation without a regressor with large support. For example, despite using unique potential outcomes, the point identification result in [Ciliberto and Tamer \(2009\)](#) uses a regressor with large support, leading them to use partial identification inference results in practice. Note that the existence of unique potential outcomes relies critically on knowing (or point identifying; see Section 5) the sign of the interaction effect. See also, for example, [Bresnahan and Reiss \(1991a\)](#).

Therefore, for explanatory variable  $k$  of agent  $i$ ,  $\frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{ik}} = P_0^{(i)}(-\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}, \dots, -\alpha_N - x_N\beta_{Nx} - w\beta_{Nw})(-\beta_{ixk})$ , where  $P_0^{(i)}(\cdot)$  indicates the evaluation of the derivative of  $P_0(\cdot)$  with respect to its  $i$ th argument, and so

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

so  $\beta_{ix}$  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 similar, but more technical, identification strategy shows that  $\beta_{iw}$  is point identified up to the same scale normalization. Identification of  $\beta_{iw}$  is more complicated because shared explanatory variables affect all agents' utilities, but only the total effect of the explanatory variables is "observed," so some additional work is necessary to point identify the effect on each agent separately.

The intercept parameters are not coefficients on exogenous explanatory variables, and therefore cannot be point identified using this strategy. Identification of the intercept parameters is important, particularly as the interaction effect parameter  $\Delta$  is an intercept parameter. More generally, note that identification (and/or estimation) of intercept parameters of 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 \(1998\)](#) or Remark 6.2).

Define  $c_i = -x_i\beta_{ix} - w\beta_{iw}$  and  $c = (c_1, c_2, \dots, c_N)$ , which are point identified given that  $\beta$  is point identified. Then  $P(y = 0_{1 \times N}|c) = P_0(-\alpha_1 + c_1, -\alpha_2 + c_2, \dots, -\alpha_N + c_N)$  and  $P(y = 1_{1 \times N}|c) = P_1(-\alpha_1 - N_0\Delta_1 + c_1, -\alpha_2 - N_0\Delta_2 + c_2, \dots, -\alpha_N - N_0\Delta_N + c_N)$ . Consequently,

$$\frac{\partial^N P(y = 0_{1 \times N}|c)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{(a_1, a_2, \dots, a_N)} = P_0^{(12 \dots N)}(-\alpha_1 + a_1, -\alpha_2 + a_2, \dots, -\alpha_N + a_N),$$

where  $P_0^{(12 \dots N)}$  is the  $N$ th partial derivative of  $P_0$ , taking a partial derivative with respect to each argument exactly once. The left hand side is observed, and the right hand side is an unknown function of the parameters of interest,  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$ . In general, it is impossible to use such an equation to identify the intercept parameters, since the unknown  $P_0^{(12 \dots N)}$  can "absorb" the intercept terms.<sup>5</sup>

But  $P_0^{(12 \dots N)}(t_1, t_2, \dots, t_N) = f_\varepsilon(t_1, t_2, \dots, t_N)$ , where  $f_\varepsilon$  is the density of  $\varepsilon$ . So  $\frac{\partial^N P(y=0_{1 \times N}|c)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{(a_1, a_2, \dots, a_N)}$  is the joint density of  $\varepsilon$  evaluated at  $(-\alpha_1 + a_1, -\alpha_2 + a_2, \dots, -\alpha_N + a_N)$ . Suppose that the joint density of  $\varepsilon$  is unimodal, in the sense of achieving the global maximum at a unique point, allowing potentially many local maxima.

<sup>5</sup>It could be that  $P_0^{(12 \dots N)}(-\alpha_1 + a_1, -\alpha_2 + a_2, \dots, -\alpha_N + a_N) = Q_0^{(12 \dots N)}(-\alpha'_1 + a_1, -\alpha'_2 + a_2, \dots, -\alpha'_N + a_N)$  for all  $(a_1, a_2, \dots, a_N)$  if  $Q_0^{(12 \dots N)}(t_1, t_2, \dots, t_N) \equiv P_0^{(12 \dots N)}(t_1 - \alpha_1 + \alpha'_1, t_2 - \alpha_2 + \alpha'_2, \dots, t_N - \alpha_N + \alpha'_N)$ , implying that  $\alpha$  would not be point identified, since  $\{\alpha, \alpha'\}$  would be observationally equivalent (relative to that equation) to  $\{Q_0, \alpha'\}$ .

As a normalization, it can be assumed that  $0_{1 \times N}$  is the unique global maximizer of the joint density of  $\varepsilon$ . Then  $\frac{\partial^N P(y=0_{1 \times N}|c)}{\partial c_1 \partial c_2 \dots \partial c_N} |_{(a_1, a_2, \dots, a_N)}$  is uniquely maximized as a function of  $(a_1, a_2, \dots, a_N)$  when  $a_i = \alpha_i$  for all agents  $i$ . Since  $\frac{\partial^N P(y=0_{1 \times N}|c)}{\partial c_1 \partial c_2 \dots \partial c_N} |_{(a_1, a_2, \dots, a_N)}$  is observed, this implies that  $\alpha$  is point identified. Similarly,  $(-1)^N \frac{\partial^N P(y=1_{1 \times N}|c)}{\partial c_1 \partial c_2 \dots \partial c_N} |_{(b_1, b_2, \dots, b_N)}$  is uniquely maximized as a function of  $(b_1, b_2, \dots, b_N)$  when  $b_i = \alpha_i + N_0 \Delta_i$  for all agents  $i$ , so  $\Delta$  is point identified.

### 3.2 Formal identification results

The following assumptions are sufficient for point identification of the parameters of the utility function.

**ASSUMPTION 3.1 (Scale Normalization).** *It holds that  $\beta_{ix1} = 1$  for all agents  $i$ .*

The scale of the utility functions has no observable implications. See Remark 3.7 on the fact that this is also a sign assumption. (In short, the proof of Theorem 3.1 shows that the sign of  $\beta_{ix1}$  is point identified, and so to avoid distracting accounting details related to keeping track of the sign, it is assumed the signs are positive. If not, the signs of  $x_{i1}$  can be “flipped” by multiplying them by  $-1$ .)

**ASSUMPTION 3.2 (Independence of Unobservables From Explanatory Variables).** *It holds that  $\varepsilon \perp z$ .*

Assumption 3.2 allows that  $\varepsilon_{im}$  is correlated with  $\varepsilon_{jm}$  within market  $m$ .

**ASSUMPTION 3.3 (Unobservables Have Mode at Zero).** *The distribution of  $\varepsilon$  admits an ordinary continuous density  $f_\varepsilon(\cdot)$  with respect to Lebesgue measure. The unique mode of  $\varepsilon$  is  $0_{1 \times N}$ , defined by  $0_{1 \times N} = \operatorname{argmax}_{t_1, t_2, \dots, t_N} f_\varepsilon(t_1, t_2, \dots, t_N)$ .*

The assumption on the mode of  $\varepsilon$  deals with the fact that a necessary condition for point identification of  $\alpha$  and  $\Delta$  is an assumption on the location of  $\varepsilon$ . The condition that the mode is at  $0_{1 \times N}$  rather than some other point is a normalization, as the true mode is absorbed by  $\alpha$ . The substantive condition is that the density of  $\varepsilon$  is maximized at a unique point (i.e.,  $\operatorname{argmax}_t f_\varepsilon(t)$  is a singleton set), a condition satisfied by many important distributions.<sup>6</sup>

Assumption 3.3 requires only the uniqueness of the point achieving the *global* maximum of the density of  $\varepsilon$ , and therefore allows multiple *local* maxima of the density of  $\varepsilon$ .<sup>7</sup>

<sup>6</sup>The use of the mode as a location assumption in econometrics seems rare, but one notable exception is the mode regression setup of Lee (1989). Of course, the use of the mode differs significantly between this paper and Lee (1989).

<sup>7</sup>The definition of *unimodal* is not standardized, as sometimes it means achieving the maximal density at a unique point, but allowing multiple local maxima, while other times it means having only one local (and therefore global) maximum. Assumption 3.3 uses unimodal in the weaker sense of achieving the global maximum at a unique point, while having possibly many local maxima. Remark 3.6 discusses some additional implications of assuming the stronger sense of “strictly unimodal.”

Many distributions, for example, mixture distributions, will tend to have multiple local maxima of the density, but will achieve the global maximum at a unique point.

Typically, the location assumption concerns either the mean or median of the unobservables. However,  $E(\varepsilon) = 0_{1 \times N}$  and  $\varepsilon \perp z$  is not sufficient for point identification because there is no regressor with large support.<sup>8</sup> An assumption on the median, in the sense of  $\text{Med}(\varepsilon) \equiv \{\text{Med}(\varepsilon_i)\}_i$ , was shown to be compatible with point identification in [Kline \(2015a\)](#) with a large support regressor (and allowing  $\varepsilon$  to not be independent from  $z$ ), but without a large support regressor it seems difficult (although not necessarily impossible) to achieve point identification using this assumption.<sup>9</sup>

The identifying power of an assumption on the mode of  $\varepsilon$  is due to the fact that the condition that  $0_{1 \times N}$  is the mode of  $\varepsilon$  is a “local” property. Solving

$$\underset{t_1, t_2, \dots, t_N}{\text{argmax}} f_{\varepsilon_1+c_1, \varepsilon_2+c_2, \dots, \varepsilon_N+c_N}(t_1, t_2, \dots, t_N),$$

which is the problem of finding the mode of  $(\varepsilon_1 + c_1, \varepsilon_2 + c_2, \dots, \varepsilon_N + c_N)$  so as to identify  $(c_1, c_2, \dots, c_N)$ , requires only that the density  $f_{\varepsilon_1+c_1, \varepsilon_2+c_2, \dots, \varepsilon_N+c_N}(\cdot)$  is known in a neighborhood of the mode. For example, suppose that the density of  $(\varepsilon_1 + \alpha_1, \varepsilon_2 + \alpha_2, \dots, \varepsilon_N + \alpha_N)$  is point identified on some set  $\mathcal{E}$ . This is established in the identification result, and generically  $\mathcal{E} \subsetneq \mathbb{R}^N$  because there is no regressor with “large support,” and the density of  $(\varepsilon_1 + \alpha_1, \varepsilon_2 + \alpha_2, \dots, \varepsilon_N + \alpha_N)$  is point identified relative to the support of  $z$ . Nevertheless, as long as  $\mathcal{E}$  contains  $(\alpha_1, \alpha_2, \dots, \alpha_N)$ , the density of  $(\varepsilon_1 + \alpha_1, \varepsilon_2 + \alpha_2, \dots, \varepsilon_N + \alpha_N)$  is point identified at its mode, so the mode of  $(\varepsilon_1 + \alpha_1, \varepsilon_2 + \alpha_2, \dots, \varepsilon_N + \alpha_N)$  can be identified as the point in  $\mathcal{E}$  with highest density. The mode is located at  $(\alpha_1, \alpha_2, \dots, \alpha_N)$ , so  $\alpha$  is point identified. The interaction effects  $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_N)$  are point identified similarly.

<sup>8</sup>[Manski \(1988\)](#) shows that mean independence is not sufficient for point identification in a single-agent discrete choice model. [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. If there is not a regressor with large support, then the intercept is not point identified even if  $E(\varepsilon) = 0_{1 \times N}$  and  $\varepsilon \perp z$ . 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  $\varepsilon$ , called  $\mathcal{E}_y$ , such that for any realization of  $\varepsilon \in \mathcal{E}_y$  and any realization of  $z$ , the outcome of the game is  $y$ . For example, when  $N = 2$ , for  $y = (1, 1)$  the set of  $\mathcal{E}_y$  are  $\varepsilon$  such that  $\varepsilon_1$  and  $\varepsilon_2$  are both very large. By rearranging probability mass of  $\varepsilon$  within these regions of  $\varepsilon$  space it is possible to construct observationally equivalent models that have different location parameters  $\alpha$ , because the mean of the constructed distribution of  $\varepsilon$  can take arbitrarily positive or negative values since the mean functional is infinitely sensitive to sufficiently large outliers even if they have only very small probability. Note that the identification strategy for the slope parameters depends less critically on the support of the explanatory variables, both in this paper and in (some) papers on single-agent discrete choice models. See, for example, the discussion of the support of the explanatory variables in identifiability of the slope parameters in [Cosslett \(1983\)](#).

<sup>9</sup>The difficulty is that identification of the joint cumulative distribution function of  $\varepsilon$  on a set of points  $\mathcal{E}$  does not necessarily imply identification of the marginal cumulative distribution function of  $\varepsilon_i$  for any particular agent  $i$  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 3.4 (Continuous Explanatory Variables). *The distribution of  $z$  is supported on a convex set with nonempty interior, and admits an ordinary continuous density with respect to Lebesgue measure that is strictly positive on the support, except possibly the boundary.*<sup>10</sup> *The support of  $z$  includes an open set  $\mathcal{Z}_0$  with  $P(z \in \mathcal{Z}_0) > 0$  such that for  $z \in \mathcal{Z}_0$ ,  $\frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{i1}} \neq 0$  for all agents  $i$ .*

This assumption is understood to apply to the variables that *actually exist*, so it allows that there are no shared variables. The requirement of continuous explanatory variables arises because the identification strategy uses derivatives with respect to the explanatory variables. The extension to allowing discrete explanatory variables is developed in Section 4.1. The strategy with discrete explanatory variables is to identify the model conditional on each value of the discrete explanatory variables, using the identification strategy in this section. The set  $\mathcal{Z}_0$  is the entire support of  $z$  if  $\varepsilon$  has everywhere positive density, under the maintained assumption that  $\beta_{ix1} \neq 0$  for all agents  $i$ . Conversely, if  $\frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{i1}} = 0$  for all agents  $i$ , and under the maintained assumption that  $\beta_{ix1} \neq 0$  for all agents  $i$ , it would follow that  $\frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{ik}} = 0 = \frac{\partial P(y=0_{1 \times N}|z)}{\partial w_l}$  for all agents  $i$ , and all  $k$  and  $l$ , so the existence of  $\mathcal{Z}_0$  essentially is the condition that different values of  $z$  result in different probabilities of the  $0_{1 \times N}$  equilibrium outcome. (See the proof of Theorem 3.1 for the details.)

ASSUMPTION 3.5 (Sufficient Variation of Explanatory Variables). *The density of  $(-x_1\beta_{1x} - w\beta_{1w}, -x_2\beta_{2x} - w\beta_{2w}, \dots, -x_N\beta_{Nx} - w\beta_{Nw})$  exists and is strictly positive on a convex open set that contains  $(\alpha_1, \alpha_2, \dots, \alpha_N)$  and  $(\alpha_1 + N_0\Delta_1, \alpha_2 + N_0\Delta_2, \dots, \alpha_N + N_0\Delta_N)$ .*

This assumption requires that the densities of  $\varepsilon + \alpha$  and  $\varepsilon + \alpha + N_0\Delta$  are observed at their modes, where “observed” is meant in the sense of the identification strategy above, which results in identifying  $\alpha$  and  $\alpha + N_0\Delta$  since the mode of  $\varepsilon$  is zero. 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.<sup>11</sup> This assumption can have straightforward observable implications (under an additional assumption on the unobservables), as discussed in Remark 3.6. Also, since identification and estimation of  $\beta$  do not depend on this assumption, it is possible to investigate the credibility of this assumption by estimating (heuristically, not as a formal estimation problem) the support of  $(-x_1\beta_{1x} - w\beta_{1w}, -x_2\beta_{2x} - w\beta_{2w}, \dots, -x_N\beta_{Nx} - w\beta_{Nw})$  based on the estimates of  $\beta$ . Some further remarks about this assumption are relegated to a footnote.<sup>12</sup>

<sup>10</sup>The boundary has Lebesgue measure zero, so it is irrelevant (e.g., Dudley (1999, Lemma 2.4.3)).

<sup>11</sup>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  $\alpha$ .

<sup>12</sup>It seems implausible that point identification is possible without this assumption or a stronger assumption. To explain that claim, suppose that  $\varepsilon \equiv 0$ , which can be viewed as an extremely strong lo-

ASSUMPTION 3.6 (Full Rank Marginal Effects of Agent-Specific Explanatory Variables). *If there is a shared explanatory variable (i.e.,  $w$  is not void), either of the following statements holds:*

- (1) *It holds that  $\beta_{iw} = \beta_w$  for all agents  $i$ .*
- (2) *The term  $E(P_x(z))$  exists and has full rank, where  $P_x(z)$  is the  $N \times N$  matrix with  $P_{x,ij}(z) = \frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{i1}} \frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{j1}}$ .*

*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 satisfied. This is implied by the assumption that  $\beta_i = \beta$  for all agents  $i$ , which often arises in models where the “labeling” of the agents has no economic content. Otherwise, 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 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.

The following theorem gives an identification result for  $\Delta \leq 0$ , as in an entry game.

REMARK 3.1 (Nonnegative Interaction Effect). The case of  $\Delta \geq 0$  can be addressed similarly when  $N = 2$ , using outcomes  $(0, 1)$  and  $(1, 0)$ , and adjusting the statement of the theorem and the assumptions. Specifically,  $P((1, 0)|z) = P(\varepsilon_1 \geq -\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, \varepsilon_2 \leq -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2)$  and  $P((0, 1)|z) = P(\varepsilon_1 \leq -\alpha_1 - x_1\beta_{1x} - w\beta_{1w} - \Delta_1, \varepsilon_2 \geq -\alpha_2 - x_2\beta_{2x} - w\beta_{2w})$ . If  $N > 2$  and  $\Delta \geq 0$ , then the identification strategy seems to not directly extend, due to lack of unique potential outcomes.<sup>13</sup>

cation assumption. Suppose  $N = 2$ . Suppose  $\alpha_1 = \alpha = \alpha_2$  and  $\Delta_1 = \Delta = \Delta_2$  and suppose the support of  $(-x_1\beta_{1x} - w\beta_{1w}, -x_2\beta_{2x} - w\beta_{2w})$  is  $[\underline{c}, \bar{c}]^2$ , a “box” in  $\mathbb{R}^2$ . Then suppose, contrary to Assumption 3.5, that  $[\underline{c}, \bar{c}]^2$  is contained within  $[\alpha + \Delta, \alpha]^2$ . That implies that for all values of the explanatory variables, the utility functions are in the region of multiple equilibria, so any more negative  $\Delta$  could generate exactly the same distribution of observables, since the utility functions would remain in the region of multiple equilibria with more negative  $\Delta$ . As a practical matter, this assumption means that the econometrician should assume that there is a bounded parameter space for  $(\alpha_1, \alpha_2, \Delta_1, \Delta_2)$ , which implies that the possible values of  $(\alpha_1, \alpha_2)$  and  $(\alpha_1 + \Delta_1, \alpha_2 + \Delta_2)$  must be in some bounded set  $\Gamma$ . The assumption of a bounded parameter space is ubiquitous in estimation, and is assumed in this paper in Section 6. Assumption 3.5 requires that the support of the (negative of the) linear combination of the explanatory variables appearing in the utility function contains  $\Gamma$ .

<sup>13</sup>Extrapolating from  $(0, 1)$  being a unique potential outcome when  $N = 2$ , it is tempting but false to claim that  $(0, 0, 1)$  would be a unique potential outcome when  $N = 3$ . Suppose that agent 3 is such that  $y_3 = 1$  is a dominant action. Then agents 1 and 2 essentially play the two-agent game, in which  $(0, 0)$  can be part of the region of multiple equilibria with  $(1, 1)$ . So  $(0, 0, 1)$  would be part of the region of multiple equilibria with  $(1, 1, 1)$ . It is similar for other candidate unique potential outcomes, when  $N > 3$ .

The following theorem allows a certain weight function  $\pi(\cdot)$ . For estimation purposes in Section 6,  $\pi(\cdot)$  is taken to be the density of  $z$ . Also, in the statement of the theorem, let  $\beta_{wl} = (\beta_{1wl}, \beta_{2wl}, \dots, \beta_{Nwl})$  for any  $l \in \{1, 2, \dots, L\}$ , let  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$ , let  $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_N)$ , and let  $P^{[1]}(z)$  be the  $1 \times N$  matrix whose  $i$ th entry is  $\frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{i1}}$ , so that  $E(P_x(z)) = E(P^{[1]}(z)'P^{[1]}(z))$ .

**THEOREM 3.1 (Identification).** *Suppose that  $\Delta_i \leq 0$  for all agents  $i$ . Under Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6,  $\theta = (\alpha, \beta, \Delta)$  is point identified, and can be expressed in terms of observables as follows.*

*Let  $\pi(\cdot)$  be any function of  $z$  that is strictly positive on the support of  $z$ , except possibly the boundary. Then assuming that the expectations in these expressions exist (see Remark 3.3), for any  $i$  and  $k \in \{1, 2, \dots, D_i\}$ ,*

$$\beta_{ixk} = \frac{E\left(\pi(z) \frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{ik}}\right)}{E\left(\pi(z) \frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{i1}}\right)}.$$

*If  $\beta_{iw} = \beta_w$  for all agents  $i$ , then for any  $l \in \{1, 2, \dots, L\}$ ,*

$$\beta_{wl} = \frac{E\left(\pi(z) \frac{\partial P(y=0_{1 \times N}|z)}{\partial w_l}\right)}{\sum_i^N E\left(\pi(z) \frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{i1}}\right)}.$$

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

$$\beta_{wl} = (E(P^{[1]}(z)'P^{[1]}(z)))^{-1} E\left(P^{[1]}(z)' \frac{dP(y=0_{1 \times N}|z)}{dw_l}\right).$$

*Further, set  $c_i \equiv -x_i \beta_{ix} - w \beta_{iw}$  and  $c = (c_1, c_2, \dots, c_N)$ , which are point identified by the above statements. Then*

$$\alpha = \operatorname{argmax}_{a=(a_1, a_2, \dots, a_N)} \frac{\partial^N P(y=0_{1 \times N}|c)}{\partial c_1 \partial c_2 \cdots \partial c_N} \Big|_a$$

*and*

$$\Delta = \frac{1}{N_0} \left( \operatorname{argmax}_{b=(b_1, b_2, \dots, b_N)} (-1)^N \frac{\partial^N P(y=1_{1 \times N}|c)}{\partial c_1 \partial c_2 \cdots \partial c_N} \Big|_b \right. \\ \left. - \operatorname{argmax}_{a=(a_1, a_2, \dots, a_N)} \frac{\partial^N P(y=0_{1 \times N}|c)}{\partial c_1 \partial c_2 \cdots \partial c_N} \Big|_a \right),$$

*where the maximization is over the support of  $(c_1, c_2, \dots, c_N)$ .*

*The same identification results obtain for  $\beta$  even without Assumption 3.5.*

REMARK 3.2 (Small Extensions). For purposes of identifying  $\beta$ , analogous results apply to  $1_{1 \times N}$  in place of  $0_{1 \times N}$ , increasing efficiency of estimation.<sup>14</sup> The expression for  $\beta_{wl}$  is also valid with  $\pi(\cdot)$  inside all of the expectations, as long as  $E(\pi(z)P_x(z))$  has full rank. (The details of this claim are obvious from the proof.)

REMARK 3.3 (Existence of the Expectations). Theorem 3.1 requires existence of expectations. A sufficient condition is that the densities of  $\varepsilon_i$  for each agent  $i$  are bounded above (which is essentially implied by the mode assumption), and that  $\pi(\cdot)$  is integrable with respect to the data generating process for  $z$ . See Lemma 10.1.

REMARK 3.4 (Equilibrium Existence). A pure strategy Nash equilibrium exists, by an argument detailed in the Supplement, based on comparative statics.

REMARK 3.5 (Large Number of Agents). One possible concern with this result in the case of large  $N$  (i.e.,  $N \gg 2$ ) is that unique potential outcomes might be observed only rarely in the data, even though the model implies they do exist in principle. This does not threaten the validity of the identification strategy, but it might result in imprecise estimates. In particular, Section 8 establishes that the rate of convergence of  $\alpha$  and  $\Delta$  is slower with larger  $N$ . Of course, alternative identification strategies based on regressors with large support may also result in imprecise estimates, since observations with extreme realizations of the regressors might be rare even when they do exist in principle (e.g., Khan and Tamer (2010)).

One relative advantage of the identification strategy in this paper is that the observation in the data of unique potential outcomes is implied by extreme realizations of the regressors, but extreme realizations of the regressors is not implied by the observation of unique potential outcomes in the data. For example,  $0_{1 \times N}$  is the unique potential outcome whenever  $\varepsilon_i \leq -\alpha_i - x_i\beta_{ix} - w\beta_{iw}$  for all agents  $i$ , which in particular would arise for extreme values of the explanatory variables that come from a regressor with large support. But, clearly, the  $0_{1 \times N}$  outcome arises even for moderate values of the explanatory variables. And so, compared to identification using a regressor with large support, this identification strategy seems attractive even in the case of many agents.<sup>15</sup>

REMARK 3.6 (Conditions on the Support of the Explanatory Variables). Assumption 3.5 has observable implications under the strengthened assumption that the density of  $\varepsilon$  is strictly unimodal in the sense that it has a global maximum achieved at a unique point and no other local maxima. Note that Assumption 3.3 requires only that the global maximum is achieved at a unique point, allowing many local maxima.

<sup>14</sup>Per Assumption 3.4, that requires that the support of  $z$  includes an open set  $Z'_0$  with  $P(z \in Z'_0) > 0$  such that, for  $z \in Z'_0$ ,  $\frac{\partial P(y=1_{1 \times N}|z)}{\partial x_{i1}} \neq 0$  for all agents  $i$ .

<sup>15</sup>Another approach is partial identification: Aradillas-Lopez and Tamer (2008), Ciliberto and Tamer (2009), Beresteanu, Molchanov, and Molinari (2011), Galichon and Henry (2011), Kline and Tamer (2012), Aradillas-Lopez and Rosen (2013), or Kline and Tamer (2015). Even from that perspective, the results of this paper are still useful (as are any results establishing point identification), as they establish that there is point identifying information about the parameters under the provided conditions.

More specifically, by *strictly unimodal*, this remark means that the density of  $\varepsilon$  has an everywhere negative definite Hessian matrix, and achieves the global maximum at a unique point and has no other local maxima. Therefore, implicitly it is assumed in this remark that the density of  $\varepsilon$  is twice continuously differentiable.

Theorem 3.1 (together with the proof) shows that  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$  and  $\alpha + N_0\Delta = (\alpha_1 + N_0\Delta_1, \alpha_2 + N_0\Delta_2, \dots, \alpha_N + N_0\Delta_N)$  are point identified as

$$\begin{aligned}\alpha &= \operatorname{argmax}_{a_1, a_2, \dots, a_N} \left. \frac{\partial^N P(y = 0_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \cdots \partial c_N} \right|_{a_1, a_2, \dots, a_N} \\ &= \operatorname{argmax}_{a_1, a_2, \dots, a_N} f_\varepsilon(-\alpha_1 + a_1, -\alpha_2 + a_2, \dots, -\alpha_N + a_N)\end{aligned}$$

and

$$\begin{aligned}\alpha + N_0\Delta &= \operatorname{argmax}_{b_1, b_2, \dots, b_N} (-1)^N \left. \frac{\partial^N P(y = 1_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \cdots \partial c_N} \right|_{b_1, b_2, \dots, b_N} \\ &= \operatorname{argmax}_{b_1, b_2, \dots, b_N} f_\varepsilon(-\alpha_1 - N_0\Delta_1 + b_1, -\alpha_2 - N_0\Delta_2 + b_2, \dots, -\alpha_N - N_0\Delta_N + b_N).\end{aligned}$$

Under the assumption that the density of  $\varepsilon$  is strictly unimodal, the first order conditions (e.g.,  $\frac{\partial^{N+1} P(y=0_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial^2 c_1 \partial c_2 \cdots \partial c_N} |_{a_1, a_2, \dots, a_N} = 0$ ) to these maximization problems will be *uniquely* satisfied when evaluated at  $\alpha$  and  $\alpha + N_0\Delta$ , 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.)

Further, if Assumption 3.5 is satisfied, then the parameters defined by the above maximization problems over the support of  $(c_1, c_2, \dots, c_N)$  will indeed equal  $\alpha$  and  $\alpha + N_0\Delta$ . In contrast, if Assumption 3.5 is not satisfied, then since the maximization is by construction over the support of  $(c_1, c_2, \dots, c_N)$ , the parameters defined by the above maximization cannot equal  $\alpha$  and/or  $\alpha + N_0\Delta$ . So 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.

**REMARK 3.7 (Scale Normalization).** Assumption 3.1 implies a sign assumption. However, the sign can be identified by the same identification strategy, because the proof of identification shows that  $\frac{\partial P(y=0_{1 \times N} | z)}{\partial x_{i1}} = F_\varepsilon^{[i,1]}(-\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}, \dots, -\alpha_N - x_N\beta_{Nx} - w\beta_{Nw})(-\beta_{ix1})$  and  $F_\varepsilon^{[i,1]}(-\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}, \dots, -\alpha_N - x_N\beta_{Nx} - w\beta_{Nw}) > 0$  for  $z \in \mathcal{Z}_0$ , a result that does not depend on the sign assumption, so the sign of  $\beta_{ix1}$  is point identified. So by “flipping” the sign of  $x_{i1}$  appropriately, by multiplying by  $-1$ , the sign assumption is without loss of generality. If  $N = 2$ , and  $\Delta_1$  and  $\Delta_2$  are nonnegative, then the sign can be identified using  $(0, 1)$  and  $(1, 0)$ : in that case,  $\frac{\partial P(y_i=0, y_{-i}=1 | z)}{\partial x_{i1}}$  has the opposite sign of  $\beta_{ix1}$ .

## 4. EXTENSIONS

## 4.1 Identification with discrete explanatory variables

The identification strategy with discrete explanatory variables is to partition the data based on the finitely many values of the discrete explanatory variables. (By *discrete*, this paper implicitly assumes finitely many support points.) Then, viewing the model as conditional on the discrete explanatory variables, it is possible to point identify the parameters of the model. This recovers the slope coefficients on the continuous explanatory variables, the interaction effects, and the intercept terms that absorb the contribution of the discrete explanatory variables. Further arguments show that it is possible to recover the slope coefficients on the discrete explanatory variables and the true intercept parameters  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$ .

The following assumption allows discrete explanatory variables.

**ASSUMPTION 4.1** (Continuous and Discrete Explanatory Variables). *There is a partitioning of  $z = (x_1, x_2, \dots, x_N, w)$  into  $z^{(d)} = (x_1^{(d)}, x_2^{(d)}, \dots, x_N^{(d)}, w^{(d)})$  and  $z^{(c)} = (x_1^{(c)}, x_2^{(c)}, \dots, x_N^{(c)}, w^{(c)})$  where the following statements hold:*

- (1) *The variable  $z^{(d)}$  has a discrete distribution with finite and discrete support,  $\mathcal{Z}^{(d)}$ .*
- (2) *For each value of  $z^{(d)} \in \mathcal{Z}^{(d)}$ , the conditional distribution of  $z^{(c)}|z^{(d)}$  is supported on a convex set with nonempty interior, and admits an ordinary continuous density with respect to Lebesgue measure that is strictly positive on the support, except possibly the boundary.*
- (3) *For each agent  $i$ ,  $x_{i1}^{(c)} = x_{i1}$ , so that the explanatory variable with scale normalization from Assumption 3.1 is continuous.*
- (4) *For each value of  $z^{(d)} \in \mathcal{Z}^{(d)}$ , the support of  $z^{(c)}|z^{(d)}$  includes a set  $\mathcal{Z}_{z^{(d)}}^{(c)}$  with  $P(z^{(c)} \in \mathcal{Z}_{z^{(d)}}^{(c)}|z^{(d)}) > 0$  such that, for  $z^{(c)} \in \mathcal{Z}_{z^{(d)}}^{(c)}$ ,  $\frac{\partial P(y=0_{1 \times N}|z^{(c)}, z^{(d)})}{\partial x_{i1}^{(c)}} \neq 0$  for all agents  $i$ .*

As with Assumption 3.4, this assumption is understood to apply to the variables that actually exist when there are, for example, no shared variables. The assumption requires at least one agent-specific continuous explanatory variable per agent. Also, as with Assumption 3.4,  $\mathcal{Z}_{z^{(d)}}^{(c)}$  is the entire support of  $z^{(c)}|z^{(d)}$  if  $\varepsilon$  has everywhere positive density, under the maintained assumption that  $\beta_{ix1} \neq 0$  for all agents  $i$ .

The slope coefficients are similarly partitioned as  $\beta_{ix} = (\beta_{ix}^{(d)}, \beta_{ix}^{(c)})$  and  $\beta_{iw} = (\beta_{iw}^{(d)}, \beta_{iw}^{(c)})$ . The dimension of  $x_i^{(d)}$  is  $D_i^{(d)}$ , the dimension of  $x_i^{(c)}$  is  $D_i^{(c)} \geq 1$ , the dimension of  $w^{(d)}$  is  $L^{(d)}$ , and the dimension of  $w^{(c)}$  is  $L^{(c)}$ . By Assumption 3.1,  $\beta_{ix1}^{(c)} = 1$ . For each  $z^{(d)} \in \mathcal{Z}^{(d)}$  such that  $z^{(d)} = (x_1^{(d)}, x_2^{(d)}, \dots, x_N^{(d)}, w^{(d)})$ , let

$$\alpha_{i,z^{(d)}} = \alpha_i + x_i^{(d)} \beta_{ix}^{(d)} + w^{(d)} \beta_{iw}^{(d)}.$$

So, then note that the utility functions from equation (1) can be written as

$$\begin{aligned} u_{im}(0, y_{-im}) &= 0 \quad \text{and} \\ u_{im}(1, y_{-im}) &= \alpha_{im,z^{(d)}} + x_{im}^{(c)} \beta_{ix}^{(c)} + w_m^{(c)} \beta_{iw}^{(c)} + \Delta_i \sum_{j \neq i} y_{jm} + \varepsilon_{im}. \end{aligned} \tag{2}$$

Consequently, conditional on the discrete explanatory variables, the model falls into the class of models addressed in Section 3.2, viewing  $\alpha_{i,z^{(d)}}$  as an intercept term that, although dependent on  $z^{(d)}$ , is fixed conditional on  $z^{(d)}$ . Since the remaining explanatory variables are continuous, the results from Section 3.2 can be used to identify the parameters of this utility function.

**REMARK 4.1** (Heterogeneity in Interaction Effects). If the interaction effects depended *nonparametrically* on the discrete explanatory variables, so that the utility functions were  $u_i(0, y_{-i}) = 0$  and  $u_i(1, y_{-i}) = \alpha_{i,z^{(d)}} + x_i^{(c)} \beta_{ix}^{(c)} + w^{(c)} \beta_{iw}^{(c)} + \Delta_{i,z^{(d)}} \sum_{j \neq i} y_j + \varepsilon_i$ , then the same identification strategy could be used to identify  $\Delta_{i,z^{(d)}}$ , just as  $\alpha_{i,z^{(d)}}$  is identified. See also, for example, [Aradillas-Lopez and Gandhi \(2014\)](#) or [Kline \(2015a\)](#). Also, the slope coefficients on the continuous explanatory variables could be allowed to depend on the discrete explanatory variables. If the interaction effects depended on continuous explanatory variables, then the same approach could be used as an approximation after discretization. Using a parametric specification for the interaction effect (e.g., the interaction effect for agent  $i$  is  $q\delta_i$  for relevant explanatory variables  $q$  and parameter  $\delta_i$ ) might result in more precise estimates, using a different estimation strategy. However, that would depend on  $q$ : if  $q$  were binary (and  $\delta_i$  included an intercept), then the linear specification is without loss of generality, so it would not be expected to improve precision.

It is necessary to replace Assumptions 3.5 and 3.6 to allow discrete explanatory variables. Essentially, the replacement versions of these assumptions require that Assumptions 3.5 and 3.6 hold conditionally on the discrete explanatory variables.

**ASSUMPTION 4.2** (Sufficient Variation of Explanatory Variables With Discrete Explanatory Variables). *For each value of  $z^{(d)} \in \mathcal{Z}^{(d)}$ , the density of  $(-x_1^{(c)} \beta_{1x}^{(c)} - w^{(c)} \beta_{1w}^{(c)}, -x_2^{(c)} \beta_{2x}^{(c)} - w^{(c)} \beta_{2w}^{(c)}, \dots, -x_N^{(c)} \beta_{Nx}^{(c)} - w^{(c)} \beta_{Nw}^{(c)})|z^{(d)}$  exists and is strictly positive on a convex open set that contains  $(\alpha_{1,z^{(d)}}, \alpha_{2,z^{(d)}}, \dots, \alpha_{N,z^{(d)}})$  and  $(\alpha_{1,z^{(d)}} + N_0 \Delta_1, \alpha_{2,z^{(d)}} + N_0 \Delta_2, \dots, \alpha_{N,z^{(d)}} + N_0 \Delta_N)$ .*

Assumption 4.2 is very similar to Assumption 3.5.<sup>16</sup>

**ASSUMPTION 4.3** (Full Rank Marginal Effects of Agent-Specific Explanatory Variables With Discrete Explanatory Variables). *If there is a shared explanatory variable (i.e.,  $w$  is not void), either of the following statements holds:*

<sup>16</sup>In particular, as a practical matter, this assumption means that the econometrician should assume that there is a bounded parameter space for the parameters, which would imply that the possible values of  $(\alpha_{1,z^{(d)}}, \alpha_{2,z^{(d)}}, \dots, \alpha_{N,z^{(d)}})$  and  $(\alpha_{1,z^{(d)}} + N_0 \Delta_1, \alpha_{2,z^{(d)}} + N_0 \Delta_2, \dots, \alpha_{N,z^{(d)}} + N_0 \Delta_N)$  must be in some bounded set  $\Gamma'$ , since  $z^{(d)}$  can take on only finitely many values by construction. Then Assumption 4.2 simply requires that the support of the (negative of the) linear combination of the *continuous* explanatory variables appearing in the utility function contains  $\Gamma'$ .

- (1) It holds that  $\beta_{iw}^{(c)} = \beta_w^{(c)}$  for all agents  $i$ .
- (2) For at least one value of  $z^{(d)} \in \mathcal{Z}^{(d)}$ ,  $E(P_x(z)|z^{(d)})$  exists and has full rank, where  $P_x(z)$  is the  $N \times N$  matrix with  $P_{x,ij}(z) = \frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{i1}^{(c)}} \frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{j1}^{(c)}}$ .

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

Directly applying the result of Theorem 3.1, conditional on any value of  $z^{(d)} \in \mathcal{Z}^{(d)}$ , will result in point identifying the slope coefficients on the continuous explanatory variables, the interaction effects, and  $\alpha_{i,z^{(d)}} = \alpha_i + x_i^{(d)} \beta_{ix}^{(d)} + w^{(d)} \beta_{iw}^{(d)}$ . This is sufficient to recover the utility function at any value of the explanatory variables in the support (restricting the discrete explanatory variables to  $\mathcal{Z}^{(d)}$ ), based on equation (2). However, it might be of interest to point identify  $\alpha_i$ ,  $\beta_{ix}^{(d)}$ , and  $\beta_{iw}^{(d)}$ , for example, to extrapolate to unseen values of the discrete explanatory variables. If so, it is necessary to add an assumption on the full rank of the discrete explanatory variables.

**ASSUMPTION 4.4 (Full Rank Discrete Explanatory Variables).** For each agent  $i$ , the discrete explanatory variables appended with a constant  $\tilde{z}_i^{(d)} = (1, x_i^{(d)}, w^{(d)})$  have full rank in the sense that  $E((\tilde{z}_i^{(d)})' \tilde{z}_i^{(d)})$  has full rank.

Let  $\eta_i = (\alpha_i, \beta_{ix}^{(d)}, \beta_{iw}^{(d)})'$ . Then, since  $\alpha_{i,z^{(d)}}$  is point identified for all  $z^{(d)} \in \mathcal{Z}^{(d)}$ , it follows that  $\eta_i$  is point identified by  $\eta_i = (E((\tilde{z}_i^{(d)})' \tilde{z}_i^{(d)}))^{-1} E((\tilde{z}_i^{(d)})' \alpha_{i,z^{(d)}})$ .

In the statement of the theorem, let  $\beta_{wl}^{(c)} = (\beta_{1wl}^{(c)}, \beta_{2wl}^{(c)}, \dots, \beta_{Nwl}^{(c)})$  for any  $l \in \{1, 2, \dots, L^{(c)}\}$ , let  $\alpha_{z^{(d)}} = (\alpha_{1,z^{(d)}}, \alpha_{2,z^{(d)}}, \dots, \alpha_{N,z^{(d)}})$  for any  $z^{(d)} \in \mathcal{Z}^{(d)}$ , let  $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_N)$ , and let  $P^{[1]}(z)$  be the  $1 \times N$  matrix whose  $i$ th entry is  $\frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{i1}^{(c)}}$ , so that  $E(P_x(z)|z^{(d)}) = E(P^{[1]}(z)' P^{[1]}(z)|z^{(d)})$ .

**THEOREM 4.1 (Identification With Discrete Explanatory Variables).** Suppose that  $\Delta_i \leq 0$  for all agents  $i$ . Under Assumptions 3.1, 3.2, 3.3, 4.1, 4.2, and 4.3,  $(\beta_{ix}^{(c)}, \beta_{iw}^{(c)}, \Delta_i)$  is point identified for each agent  $i$ . Further, for any  $z^{(d)} \in \mathcal{Z}^{(d)}$ ,  $(\alpha_{1,z^{(d)}}, \alpha_{2,z^{(d)}}, \dots, \alpha_{N,z^{(d)}})$  is point identified. The parameters can be expressed in terms of observables as follows.

For any  $z^{(d)} \in \mathcal{Z}^{(d)}$ , let  $\pi_{z^{(d)}}(\cdot)$  be any function of  $z^{(c)}$  that is strictly positive on the support of  $z^{(c)}|z^{(d)}$ , except possibly the boundary. Then, assuming that the expectations in these expressions exist (see Remark 3.3), for any  $z^{(d)} \in \mathcal{Z}^{(d)}$  and any  $i$  and any  $k \in \{1, 2, \dots, D_i^{(c)}\}$ ,

$$\beta_{ixk}^{(c)} = \frac{E\left(\pi_{z^{(d)}}(z) \frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{ik}^{(c)}} \Big| z^{(d)}\right)}{E\left(\pi_{z^{(d)}}(z) \frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{i1}^{(c)}} \Big| z^{(d)}\right)}.$$



If  $\beta_{iw}^{(c)} = \beta_w^{(c)}$  for all agents  $i$ , then for any  $z^{(d)} \in \mathcal{Z}^{(d)}$  and for any  $l \in \{1, 2, \dots, L^{(c)}\}$ ,

$$\beta_{wl}^{(c)} = \frac{E\left(\pi_{z^{(d)}}(z) \frac{\partial P(y = 0_{1 \times N} | z)}{\partial w_l^{(c)}} \Big| z^{(d)}\right)}{\sum_i^N E\left(\pi_{z^{(d)}}(z) \frac{\partial P(y = 0_{1 \times N} | z)}{\partial x_{i1}^{(c)}} \Big| z^{(d)}\right)}.$$

Alternatively, if  $z^{(d)} \in \mathcal{Z}^{(d)}$  is such that  $E(P_x(z) | z^{(d)})$  has full rank, then for any  $l \in \{1, 2, \dots, L^{(c)}\}$ ,

$$\beta_{wl}^{(c)} = (E(P^{[1]}(z)' P^{[1]}(z) | z^{(d)}))^{-1} E\left(P^{[1]}(z)' \frac{dP(y = 0_{1 \times N} | z)}{dw_l^{(c)}} \Big| z^{(d)}\right).$$

Further, set  $c_i^{(c)} \equiv -x_i^{(c)} \beta_{ix}^{(c)} - w^{(c)} \beta_{iw}^{(c)}$  and  $c^{(c)} = (c_1^{(c)}, c_2^{(c)}, \dots, c_N^{(c)})$ , which are point identified by the above statements. Then, for any  $z^{(d)} \in \mathcal{Z}^{(d)}$ ,

$$\alpha_{z^{(d)}} = \operatorname{argmax}_{a=(a_1, a_2, \dots, a_N)} \frac{\partial^N P(y = 0_{1 \times N} | c^{(c)}, z^{(d)})}{\partial c_1^{(c)} \partial c_2^{(c)} \dots \partial c_N^{(c)}} \Big|_a$$

and

$$\Delta = \frac{1}{N_0} \left( \operatorname{argmax}_{b=(b_1, b_2, \dots, b_N)} (-1)^N \frac{\partial^N P(y = 1_{1 \times N} | c^{(c)}, z^{(d)})}{\partial c_1^{(c)} \partial c_2^{(c)} \dots \partial c_N^{(c)}} \Big|_b \right. \\ \left. - \operatorname{argmax}_{a=(a_1, a_2, \dots, a_N)} \frac{\partial^N P(y = 0_{1 \times N} | c^{(c)}, z^{(d)})}{\partial c_1^{(c)} \partial c_2^{(c)} \dots \partial c_N^{(c)}} \Big|_a \right),$$

where the maximization is over the support of  $(c_1^{(c)}, c_2^{(c)}, \dots, c_N^{(c)}) | z^{(d)}$ .

Under the addition of Assumption 4.4,  $(\alpha_1, \alpha_2, \dots, \alpha_N, \beta_{1x}^{(d)}, \beta_{2x}^{(d)}, \dots, \beta_{Nx}^{(d)}, \beta_{1w}^{(d)}, \beta_{2w}^{(d)}, \dots, \beta_{Nw}^{(d)})$  is point identified and can be expressed in terms of the observables as follows.

Let  $\eta_i = (\alpha_i, \beta_{ix}^{(d)}, \beta_{iw}^{(d)})'$ . Then

$$\eta_i = (E((\tilde{z}_i^{(d)})' \tilde{z}_i^{(d)}))^{-1} E((\tilde{z}_i^{(d)})' \alpha_{i,z^{(d)}}).$$

Some of the identification results concern expectations conditional on a value of the discrete explanatory variables. Of course, then, essentially the same results would also hold after integrating over the distribution of the discrete explanatory variables, which might be attractive from an estimation perspective.

#### 4.2 Discrete action games

The identification strategy can be used with certain more general discrete action games. Suppose there are  $N \geq 2$  agents, and the actions available to each agent are  $S = \{0, 1, 2, \dots, H\}$ . Suppose the utility functions are

$$u_{im}(y_{im}, y_{-im}) = \left( \tilde{\alpha}_i + x_{im} \beta_{ix} + w_m \beta_{iw} + \Delta_i \sum_j y_{jm} + \varepsilon_{im} \right) y_{im}. \quad (3)$$

For example, in a quantity competition game (e.g., Cournot competition),  $\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i \sum_j y_{jm} + \varepsilon_{im}$  can represent the average profit agent  $i$  gets from producing  $y_{im}$  units when the other agents produce  $y_{-im}$  units, or, equivalently, the realized market price minus (constant) marginal cost. In particular,  $\Delta_i$  can represent the effect of production on realized market price (i.e., slope of the demand curve). Note that  $\Delta_i$  multiplies  $\sum_j y_{jm}$ , rather than  $\sum_{j \neq i} y_{jm}$ , reflecting the effect of the production of agent  $im$  on realized market price.<sup>17</sup>

Suppose that  $\Delta_i \leq 0$  for all agents  $i$ . Then  $0_{1 \times N}$  is the Nash equilibrium if and only if  $\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i + \varepsilon_{im} \leq 0$  for all agents  $i$ .<sup>18</sup> Also,  $H_{1 \times N} \equiv (H, H, \dots, H)$  is the Nash equilibrium if and only if  $\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i(NH + H - 1) + \varepsilon_{im} \geq 0$ .<sup>19</sup> Let  $\alpha_i = \tilde{\alpha}_i + \Delta_i$ . Consequently,  $P(y = 0_{1 \times N} | z) = P(\{\varepsilon_i \leq -\alpha_i - x_i\beta_{ix} - w\beta_{iw}\} \forall i)$ , which is the same as the expression for  $P(y = 0_{1 \times N} | z)$  for binary actions. Also,  $P(y = H_{1 \times N} | z) = P(\{\varepsilon_i \geq -\alpha_i - x_i\beta_{ix} - w\beta_{iw} - \Delta_i(NH + H - 2)\} \forall i)$ , which is almost the same as the expression for  $P(y = 1_{1 \times N} | z)$  for binary actions except that  $(NH + H - 2)$  multiplies  $\Delta_i$  rather than  $N_0$ .

Consequently, the same identification strategy can be used, substituting  $(NH + H - 2)$  for  $N_0$  and  $H_{1 \times N}$  for  $1_{1 \times N}$  in the assumptions and results. Other identification results for games with more than two actions includes Aradillas-Lopez and Rosen (2013) (complete information, mainly partial identification) and Aradillas-Lopez and Gandhi (2014) (incomplete information, partial identification).

## 5. IDENTIFICATION OF THE DIRECTION OF THE INTERACTION EFFECT

The main identification strategy (i.e., Theorem 3.1) requires either that  $N \geq 3$  and the assumption that  $\Delta_i \leq 0$  for all agents  $i$ , or that  $N = 2$  and the assumption that the econometrician knows either that  $\Delta_i \geq 0$  for all agents  $i$  or that  $\Delta_i \leq 0$  for all agents  $i$ . In the case of  $N = 2$  agents, it is possible to identify the sign of the interaction effect (i.e., the sign of  $\Delta$ ).

<sup>17</sup>In the game with  $S = \{0, 1\}$ ,  $\Delta_i$  multiplied  $\sum_{j \neq i} y_{jm}$ , but that is equivalent to a model in which  $\Delta_i$  multiplies  $\sum_j y_{jm}$ , since when  $S = \{0, 1\}$ , the entire  $\alpha_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i \sum_{j \neq i} y_{jm} + \varepsilon_{im}$  term only matters when  $y_{im} = 1$ , so essentially the intercept captures the total effect of  $y_{im} = 1$  on average profit (e.g., fixed costs of production and effect on realized market price) when  $S = \{0, 1\}$ .

<sup>18</sup>If  $0_{1 \times N}$  is the Nash equilibrium, then  $\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i + \varepsilon_{im} \leq 0$  for all agents  $i$ . Conversely, if  $\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i + \varepsilon_{im} < 0$ , then  $\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i + \Delta_i \sum_{j \neq i} y_{jm} + \varepsilon_{im} < 0$  for any  $y_{-im}$ , so  $y_{im} \geq 1$  results in negative utility, so  $y_{im} = 0$  is a strictly dominant strategy. The probability 0 event that  $\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i + \varepsilon_{im} = 0$  for some agent  $i$  can be ignored.

<sup>19</sup>If  $H_{1 \times N}$  is the Nash equilibrium, then  $(\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i NH + \varepsilon_{im})H \geq (\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i(N-1)H + \Delta_i(H-1) + \varepsilon_{im})(H-1)$  for all agents  $i$ , which is equivalent to the claimed condition, since all agents  $i$  must not find it to be a profitable deviation to take action  $H-1$ . Conversely, suppose  $\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i(NH + H - 1) + \varepsilon_{im} > 0$ . Then, for  $0 \leq y_{im} < y'_{im} \leq H$ ,  $u_{im}(y'_{im}, y_{-im}) - u_{im}(y_{im}, y_{-im}) = (\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i(\sum_{j \neq i} y_{jm} + y'_{im} + y_{im}) + \varepsilon_{im})(y'_{im} - y_{im}) \geq (\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i((N-1)H + H + H - 1) + \varepsilon_{im})(y'_{im} - y_{im}) = (\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i(NH + H - 1) + \varepsilon_{im})(y'_{im} - y_{im}) > 0$ . Therefore, for each agent  $i$ , utility is strictly increasing in own action, so agent  $i$  maximizes utility by taking action  $H$  regardless of the actions of the other agents. The probability 0 event that  $\tilde{\alpha}_i + x_{im}\beta_{ix} + w_m\beta_{iw} + \Delta_i(NH + H - 1) + \varepsilon_{im} = 0$  for some agent  $i$  can be ignored.

### 5.1 Sketch of the identification strategy

Consider two specifications of  $z = (x_1, x_2, w) = (x_{11}, x_{1(-1)}, x_2, w)$ . 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  $x'_{11} < x''_{11}$ .

Since  $\beta_{1x_1} = 1 > 0$ , for any realization of  $\varepsilon$ , the utility agent 1 gets from taking action 1 is greater at  $z''$  compared to  $z'$ . In a single-agent model, this would imply that the probability that agent 1 takes action 1 is greater at  $z''$  compared to  $z'$ . However, in a game, it could be that the equilibrium selection mechanism (over the region of multiple equilibria) chooses the outcome with  $y_1 = 1$  more often when the observables are  $z'$  compared to when the observables are  $z''$ . So a monotone selection mechanism assumption (formalized below) guarantees monotonicity of the probability of choosing action 1. Further, even without an assumption on the selection mechanism, if  $x''_{11} - x'_{11} > |\Delta_1|$ , then the probability that agent 1 takes action 1 must be greater at  $z''$  compared to  $z'$ , because if  $\varepsilon$  is such that agent 1 could possibly take action 1 in a Nash equilibrium at  $z'$  (either as a unique equilibrium or as a selection from the region of multiple equilibria), then for that  $\varepsilon$  it is a dominant strategy for agent 1 to take action 1 at  $z''$ .<sup>20</sup>

If there is a negative interaction effect, agent 2 will tend to get less utility from taking action 1 at  $z''$  compared to at  $z'$ , because agent 1 is more likely to take action 1 at  $z''$  compared to at  $z'$ , decreasing the utility agent 2 gets from taking action 1. So agent 2 should be less likely to take action 1 at  $z''$  compared to at  $z'$ . Similarly, if there is a positive interaction effect, agent 2 should be more likely to take action 1 at  $z''$  compared to at  $z'$ . So the effect of  $x_{11}$  on the probability that  $y_2 = 1$  should be equal to the sign of  $\Delta_2$ , and similarly the effect of  $x_{21}$  on the probability that  $y_1 = 1$  should be equal to the sign of  $\Delta_1$ .

### 5.2 Formal identification results

The identification strategy relies either on a monotone selection mechanism assumption or the condition that the support of  $x_{11}$  and  $x_{21}$  is large enough so that there are values  $x''_{11}$  and  $x'_{11}$  in the support with  $x''_{11} - x'_{11} > |\Delta_1|$  and values  $x''_{21}$  and  $x'_{21}$  in the support with  $x''_{21} - x'_{21} > |\Delta_2|$ . The monotone selection mechanism assumption is discussed in a separate subsection, which can be skipped if the support condition is satisfied.

**5.2.1 Monotone selection mechanism** It is necessary to introduce some notation to state the monotone selection mechanism assumption. Let  $P(\cdot|z, \varepsilon)$  be the distribution over the selected equilibrium (i.e., equilibrium selection mechanism), when observables

<sup>20</sup>If  $\varepsilon$  is such that agent 1 could take action 1 in a Nash equilibrium at  $z'$ , then it has to be that  $\alpha_1 + x'_{11} + x^*_{1(-1)}\beta_{1x(-1)} + w^*\beta_{1w} + \max\{\Delta_1, 0\} + \varepsilon_1 \geq 0$ . Because  $x''_{11} > x'_{11} + |\Delta_1|$ , it follows that  $\alpha_1 + x''_{11} + x^*_{1(-1)}\beta_{1x(-1)} + w^*\beta_{1w} + \min\{\Delta_1, 0\} + \varepsilon_1 > \alpha_1 + x'_{11} + x^*_{1(-1)}\beta_{1x(-1)} + w^*\beta_{1w} + |\Delta_1| + \min\{\Delta_1, 0\} + \varepsilon_1 = \alpha_1 + x'_{11} + x^*_{1(-1)}\beta_{1x(-1)} + w^*\beta_{1w} + \max\{\Delta_1, 0\} + \varepsilon_1 \geq 0$ , which implies the utility agent 1 gets from taking action 1 is positive at  $z''$  and that  $\varepsilon$ , regardless of what agent 2 does.

are  $z$  and unobservables are  $\varepsilon$ , that respects the maintained assumption of pure strategy Nash equilibrium. This depends on  $\varepsilon$ , so this distribution is not observed in the data. Since there are finitely many possible outcomes,  $P(\cdot|z, \varepsilon) = (P(y = (1, 1)|z, \varepsilon), P(y = (1, 0)|z, \varepsilon), P(y = (0, 1)|z, \varepsilon), P(y = (0, 0)|z, \varepsilon))$ .

Let  $\mathcal{R}^-(z, \theta) = \{(\varepsilon_1, \varepsilon_2) : -\alpha_1 - x_1\beta_{1x} - w\beta_{1w} \leq \varepsilon_1 \leq -\alpha_1 - x_1\beta_{1x} - w\beta_{1w} - \Delta_1, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} \leq \varepsilon_2 \leq -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2\}$  be the set of  $\varepsilon$  such that, for that specification of  $z$  and  $\theta$ , the game with a nonpositive interaction effect has multiple equilibria. Similarly, let  $\mathcal{R}^+(z, \theta) = \{(\varepsilon_1, \varepsilon_2) : -\alpha_1 - x_1\beta_{1x} - w\beta_{1w} - \Delta_1 \leq \varepsilon_1 \leq -\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2 \leq \varepsilon_2 \leq -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}\}$  be the set of  $\varepsilon$  such that, for that specification of  $z$  and  $\theta$ , the game with a nonnegative interaction effect has multiple equilibria. (These are the “boxes” of multiple equilibria.) Then let

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

By definition of Nash equilibrium, it must be that the equilibrium selection mechanism has the following form if there is a nonpositive interaction effect:

$$P(\cdot|z, \varepsilon) = \begin{cases} (1, 0, 0, 0), & \text{if } \varepsilon_1 > -\alpha_1 - x_1\beta_{1x} - w\beta_{1w} - \Delta_1, \varepsilon_2 > -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2, \\ (0, 0, 0, 1), & \text{if } \varepsilon_1 < -\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, \varepsilon_2 < -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}, \\ (0, 0, 1, 0), & \text{if } \varepsilon_1 < -\alpha_1 - x_1\beta_{1x} - w\beta_{1w} - \Delta_1, \varepsilon_2 > -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2, \\ (0, 0, 1, 0), & \text{if } \varepsilon_1 < -\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, \\ & -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} < \varepsilon_2 < -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2, \\ (0, 1, 0, 0), & \text{if } \varepsilon_1 > -\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, \varepsilon_2 < -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}, \\ (0, 1, 0, 0), & \text{if } \varepsilon_1 > -\alpha_1 - x_1\beta_{1x} - w\beta_{1w} - \Delta_1, \\ & -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} < \varepsilon_2 < -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2, \\ (0, p_{z, \varepsilon}, 1 - p_{z, \varepsilon}, 0), & \text{if } \varepsilon \in \mathcal{R}^-(z, \theta), \end{cases}$$

where  $p_{z, \varepsilon} \in [0, 1]$ , defined for  $(z, \varepsilon)$  such that  $\varepsilon \in \mathcal{R}^-(z, \theta)$ , characterizes the equilibrium selection mechanism in the region of multiple equilibria. In that case,  $p_{z, \varepsilon}$  is the probability of selecting the outcome  $(1, 0)$ .

Similarly, it must be that the equilibrium selection mechanism has the form, if there is a nonnegative interaction effect,

$$P(\cdot|z, \varepsilon) = \begin{cases} (0, 1, 0, 0), \\ \quad \text{if } \varepsilon_1 > -\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, \varepsilon_2 < -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2, \\ (0, 0, 1, 0), \\ \quad \text{if } \varepsilon_1 < -\alpha_1 - x_1\beta_{1x} - w\beta_{1w} - \Delta_1, \varepsilon_2 > -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}, \\ (1, 0, 0, 0), \\ \quad \text{if } \varepsilon_1 > -\alpha_1 - x_1\beta_{1x} - w\beta_{1w} - \Delta_1, \varepsilon_2 > -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}, \\ (1, 0, 0, 0), \\ \quad \text{if } \varepsilon_1 > -\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, \\ \quad \quad -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2 < \varepsilon_2 < -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}, \\ (0, 0, 0, 1), \\ \quad \text{if } \varepsilon_1 < -\alpha_1 - x_1\beta_{1x} - w\beta_{1w}, \varepsilon_2 < -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2, \\ (0, 0, 0, 1), \\ \quad \text{if } \varepsilon_1 < -\alpha_1 - x_1\beta_{1x} - w\beta_{1w} - \Delta_1, \\ \quad \quad -\alpha_2 - x_2\beta_{2x} - w\beta_{2w} - \Delta_2 < \varepsilon_2 < -\alpha_2 - x_2\beta_{2x} - w\beta_{2w}, \\ (p_{z,\varepsilon}, 0, 0, 1 - p_{z,\varepsilon}), \\ \quad \text{if } \varepsilon \in \mathcal{R}^+(z, \theta), \end{cases}$$

where  $p_{z,\varepsilon} \in [0, 1]$ , defined for  $(z, \varepsilon)$  such that  $\varepsilon \in \mathcal{R}^+(z, \theta)$ , characterizes the equilibrium selection mechanism in the region of multiple equilibria. In that case,  $p_{z,\varepsilon}$  is the probability of selecting the outcome (1, 1). For both nonpositive and nonnegative interaction effects,  $p_{z,\varepsilon}$  is the probability of the outcome with  $y_1 = 1$ . For nonnegative interaction effect,  $p_{z,\varepsilon}$  is also the probability of the outcome with  $y_2 = 1$ . For nonpositive interaction effect,  $1 - p_{z,\varepsilon}$  is the probability of the outcome with  $y_2 = 1$ .

These expressions do not address the equilibrium selection mechanisms when one or more of those strict inequalities holds as equality. Those are probability 0 events, considering the density of  $\varepsilon$ , and therefore the behavior of that part of the equilibrium selection mechanism has no relevant observable implications. It is implicitly assumed that  $p_{z,\varepsilon}$  for any given  $z$  is a measurable function of  $\varepsilon$ .

**ASSUMPTION 5.1 (Monotonic Selection Mechanism).** *The selection mechanisms are weakly increasing, in the following senses:*

- (1) *One of the following conditions holds:*
  - (a) *It holds that  $p_{z,\varepsilon} \in (0, 1)$  for all  $(z, \varepsilon)$  such that  $\varepsilon \in \mathcal{R}(z, \theta)$ .*
  - (b) *It holds that  $p_{z,\varepsilon} \equiv 0$  for all  $(z, \varepsilon)$  such that  $\varepsilon \in \mathcal{R}(z, \theta)$ .*
  - (c) *It holds that  $p_{z,\varepsilon} \equiv 1$  for all  $(z, \varepsilon)$  such that  $\varepsilon \in \mathcal{R}(z, \theta)$ .*
- (2) *For any  $z' = (x'_{11}, x'_{1(-1)}, x'_2, w^*)$  and  $z'' = (x''_{11}, x''_{1(-1)}, x''_2, w^*)$ , where  $x'_{11} \leq x''_{11}$ , and for any  $\varepsilon \in \mathcal{R}(z', \theta) \cap \mathcal{R}(z'', \theta)$ , it holds that  $p_{z',\varepsilon} \leq p_{z'',\varepsilon}$ .*

(3) For any  $z' = (x_1^*, x'_{21}, x_{2(-1)}^*, w^*)$  and  $z'' = (x_1^*, x''_{21}, x_{2(-1)}^*, w^*)$ , where  $x'_{21} \leq x''_{21}$ , and for any  $\varepsilon \in \mathcal{R}(z', \theta) \cap \mathcal{R}(z'', \theta)$ , it holds that  $p_{z', \varepsilon} \geq p_{z'', \varepsilon}$  if there is a nonpositive interaction effect or  $p_{z', \varepsilon} \leq p_{z'', \varepsilon}$  if there is a nonnegative interaction effect.<sup>21</sup>

The assumption requires that  $p_{z, \varepsilon}$  is weakly increasing in  $x_{11}$ , meaning that the probability of the outcome with  $y_1 = 1$  is weakly increasing in  $x_{11}$ . Similarly, if there is a nonnegative interaction effect, the assumption requires that  $p_{z, \varepsilon}$  is weakly increasing in  $x_{21}$ , meaning that the probability of the outcome with  $y_2 = 1$  is weakly increasing in  $x_{21}$ . And if there is a nonpositive interaction effect, the assumption requires that  $1 - p_{z, \varepsilon}$  is weakly increasing in  $x_{21}$ , meaning again that the probability of the outcome with  $y_2 = 1$  is weakly increasing in  $x_{21}$ . The first part of the assumption rules out this selection mechanism: for some  $(z, \varepsilon)$ , the selection mechanism selects the equilibrium  $y_1 = 1$  with probability 1 and for other  $(z, \varepsilon)$ , the selection mechanism selects the equilibrium  $y_1 = 1$  with probability 0.

In particular, Assumption 5.1 is satisfied if the selection mechanism randomizes over the multiple equilibria according to a fixed probability (i.e.,  $p_{z, \varepsilon} \equiv p \in [0, 1]$ ): when there are multiple equilibria, the equilibrium outcome has  $y_1 = 1$  with probability  $p$ . This concerns only the region of multiple equilibria, not the overall probability that the equilibrium outcome has  $y_1 = 1$ . Further, the assumption is satisfied if the equilibrium selection mechanism satisfies an exclusion restriction: If  $p_{z, \varepsilon} \in (0, 1)$  does not depend on  $x_{11}$  or  $x_{21}$ , then obviously the monotonicity assumption is satisfied. The use of exclusion restrictions in the identifiability of games has been addressed, for example, in Bajari, Hahn, Hong, and Ridder (2011).

There are evidently many specifications of the selection mechanism that satisfy Assumption 5.1 even without requiring an exclusion restriction. For example, the equilibrium selection mechanism satisfies the monotonicity condition if it relates to maximizing total utilitarian welfare in the following way. First, suppose that if the game has a nonnegative interaction effect, so that the selection mechanism over multiple equilibria selects among  $(1, 1)$  and  $(0, 0)$ , that in case of multiple equilibria the probability of selecting the  $(1, 1)$  outcome is increasing in  $\alpha_1 + x_1\beta_{1x} + w\beta_{1w} + \Delta_1 + \varepsilon_1 + \alpha_2 + x_2\beta_{2x} + w\beta_{2w} + \Delta_2 + \varepsilon_2$ , which is the sum of the utilities the two agents get from outcome  $(1, 1)$ , minus the sum of the utilities the two agents get from the outcome  $(0, 0)$  (which is zero). Alternatively, suppose that if the game has a nonpositive interaction effect, so that the selection mechanism over multiple equilibria selects among  $(1, 0)$  and  $(0, 1)$ , that in case of multiple equilibria the probability of selecting the  $(1, 0)$  outcome is increasing in  $(\alpha_1 + x_1\beta_{1x} + w\beta_{1w} + \varepsilon_1) - (\alpha_2 + x_2\beta_{2x} + w\beta_{2w} + \varepsilon_2)$ , which is the sum of the utilities the two agents get from the outcome  $(1, 0)$  minus the sum of the utilities the two agents get from the outcome  $(0, 1)$ .<sup>22</sup> By construction, the equilibrium selection mechanism satisfies the monotonicity condition.

<sup>21</sup>If there is both a nonpositive interaction effect and a nonnegative interaction effect, then there is a zero interaction effect, so no scope for multiple equilibria, so this assumption becomes vacuous.

<sup>22</sup>Also, it would mean that the probability of playing the outcome  $(1, 0)$  is increasing in the relative utility agent 1 gets from outcome  $(1, 0)$  compared to the utility agent 2 would get from outcome  $(0, 1)$ . So the “stronger” agent is more likely to take action 1.

### 5.2.2 Identification theorem

ASSUMPTION 5.2 (Strictly Positive Density). *The density of  $\varepsilon$  is everywhere positive.*

The proof of Theorem 5.1 shows that it is more than sufficient for the density of  $\varepsilon$  to be everywhere positive on a set that strictly contains the convex hull of  $\mathcal{R}(z', \theta)$  and  $\mathcal{R}(z'', \theta)$  for the  $z'$  and  $z''$  used in the statement of the theorem. Let  $\text{sgn}(\cdot)$  return 1 if the argument is positive, 0 if the argument is zero, and  $-1$  if the argument is negative.

THEOREM 5.1 (Identification of the Sign of the Interaction Effect). *Under Assumptions 3.1, 3.2, 3.3, and 5.2, the following results hold:*

- (1) *For any  $z' = (x_1^*, x'_{21}, x_{2(-1)}^*, w^*)$  and  $z'' = (x_1^*, x''_{21}, x_{2(-1)}^*, w^*)$ , with  $x'_{21} < x''_{21}$ ,*

$$\text{sgn}(\Delta_1) = \text{sgn}(P(y_1 = 1|z'') - P(y_1 = 1|z'))$$

*under either of the following two conditions:*

- (a) *Assumption 5.1 holds.*  
 (b) *The relationship  $x''_{21} - x'_{21} > |\Delta_2|$  holds.*  
 (2) *For any  $z' = (x'_{11}, x_{1(-1)}^*, x_2^*, w^*)$  and  $z'' = (x''_{11}, x_{1(-1)}^*, x_2^*, w^*)$ , with  $x'_{11} < x''_{11}$ ,*

$$\text{sgn}(\Delta_2) = \text{sgn}(P(y_2 = 1|z'') - P(y_2 = 1|z'))$$

*under either of the following two conditions:*

- (a) *Assumption 5.1 holds.*  
 (b) *The relationship  $x''_{11} - x'_{11} > |\Delta_1|$  holds.*

REMARK 5.1 (Comparison to de Paula and Tang (2012)). 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.<sup>23</sup> The identification result in Theorem 5.1 instead relies on the exclusion restrictions entailed by the existence of agent-specific explanatory variables. The Supplement shows how the de Paula and Tang (2012) identification strategy does carry over into the complete information game framework, in the sense that their “test statistic” is valid for both complete and incomplete information games. See also Aradillas-Lopez and Gandhi (2014) for more general incomplete information games with ordered action spaces.

<sup>23</sup>There are also papers that establish the identification of the sign of the parameter on the endogenous explanatory variable in triangular models, which shares some qualitative similarities to the problem of identification of the direction of the interaction effect. Papers in that literature include Abrevaya, Hausman, and Khan (2010) and Kline (2015b).

## 6. ESTIMATION

This section provides a three-step semiparametric estimator. The focus is on the second and third steps corresponding to estimating  $\beta$ ,  $\alpha$ , and  $\Delta$ . The proposed estimator is based on the analogy principle, given the constructive identification results. The results are stated for the case of continuous explanatory variables (i.e., Theorem 3.1), with the understanding that the results can be applied conditionally on values of discrete explanatory variables (i.e., Theorem 4.1). See Section 6.3.

The first step (when  $N = 2$ ) corresponding to estimating the sign of  $\Delta$  according to Theorem 5.1 is trivial, and so is omitted.<sup>24</sup> 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. Footnote 24 sketches the details, which are standard.

## 6.1 Estimation of slope coefficients

Theorem 3.1 suggests an estimator for  $\beta$  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  $\beta_{iw} \equiv \beta_w$  for all agents  $i$ , so as to use the first expression for the identification of  $\beta_w$  in terms of population quantities, but the properties of the estimator without that assumption would be derived similarly.<sup>25</sup>

The data consist of independent observations of the markets (i.e., plays of the game) indexed by  $m = 1, 2, \dots, M$ . Let  $\hat{\delta}_{M,ixk} = -\frac{2}{M(M-1)} \frac{1}{h_M^{d+1}} \sum_{m=1}^M \sum_{m' \neq m} 1[y_m = 0_{1 \times N}] K^{[1,ixk]}(\frac{z_m - z_{m'}}{h_M})$ , where  $D_i = \dim(x_i)$  and  $L = \dim(w)$ , and  $d = \sum_i D_i + L$ . 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  $z$ . Let  $\hat{\delta}_{M,wl} = -\frac{2}{M(M-1)} \frac{1}{h_M^{d+1}} \sum_{m=1}^M \sum_{m' \neq m} 1[y_m = 0_{1 \times N}] K^{[1,wl]}(\frac{z_m - z_{m'}}{h_M})$ , where  $K^{[1,wl]}(\cdot)$  means the derivative of  $K(\cdot)$  with respect to the  $w_l$  component of  $z$ .

Let  $\hat{\delta}_{M,ix} = (\hat{\delta}_{M,ix1}, \hat{\delta}_{M,ix2}, \dots, \hat{\delta}_{M,ixD_i})$  and  $\hat{\delta}_{M,w} = (\hat{\delta}_{M,w1}, \dots, \hat{\delta}_{M,wL})$ . Let  $\hat{\delta} = (\hat{\delta}_{M,1x}, \hat{\delta}_{M,2x}, \dots, \hat{\delta}_{M,Nx}, \hat{\delta}_{M,w})$ . Let  $\delta$  be the associated population quantities, where  $\delta_{ixk} = E(p(z) \frac{\partial P(y=0_{1 \times N}|z)}{\partial x_{ik}})$  and  $\delta_{wl} = E(p(z) \frac{\partial P(y=0_{1 \times N}|z)}{\partial w_l})$ .

The econometrician might assume  $\beta_i \equiv \beta$  for all agents  $i$ . If so, let  $\hat{\beta}_{M,xk} = \frac{\sum_i \hat{\delta}_{M,ixk}}{\sum_i \hat{\delta}_{M,ix1}}$  for  $k \geq 2$ , and let  $\hat{\beta}_{M,wl} = \frac{\hat{\delta}_{M,wl}}{\sum_i \hat{\delta}_{M,ix1}}$ . Let  $\hat{\beta}_{M,-1} = (\hat{\beta}_{M,x2}, \dots, \hat{\beta}_{M,xD}, \hat{\beta}_{M,w1}, \dots, \hat{\beta}_{M,wL})$ . Let  $\beta_{-1} = (\beta_{x2}, \dots, \beta_{xD}, \beta_{w1}, \dots, \beta_{wL})$ . (In this case,  $D_i = D$  for all agents  $i$ .)

<sup>24</sup>In short, the idea is that the estimator of  $E(P(y_i = 1|z'')) - E(P(y_i = 1|z'))$  in Theorem 5.1 can be estimated by standard nonparametric methods, and then the sign can be estimated: the estimator of the sign uses the fact that an estimator for the sign of  $\theta$  that is estimated by  $\hat{\theta}$  with  $\hat{\theta} - \theta = O_p(M^{-r})$  is  $\text{sgn}(1\{|M^{r/2}|\hat{\theta}| \geq 1\}\hat{\theta})$ , which converges arbitrarily fast. See also, for example, Andrews and Soares (2010) or Kline (2011) for other similar estimation procedures.

<sup>25</sup>The properties of the estimator of  $\beta_{wl}$  under the assumption that  $E(P_x(z))$  has full rank, which depends on the expectation of products of derivatives, can be derived similarly to 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 let  $\hat{\beta}_{M,wl} = \frac{\hat{\delta}_{M,wl}}{\sum_i \hat{\delta}_{M,ix1}}$ . Let  $\hat{\beta}_{M,-1} = (\hat{\beta}_{M,1x2}, \dots, \hat{\beta}_{M,1xD_1}, \hat{\beta}_{M,2x2}, \dots, \hat{\beta}_{M,2xD_2}, \dots, \hat{\beta}_{M,Nx2}, \dots, \hat{\beta}_{M,NxD_N}, \hat{\beta}_{M,w1}, \dots, \hat{\beta}_{M,wL})$ . Let  $\beta_{-1} = (\beta_{1x2}, \dots, \beta_{1xD_1}, \beta_{2x2}, \dots, \beta_{2xD_2}, \dots, \beta_{Nx2}, \dots, \beta_{NxD_N}, \beta_{w1}, \dots, \beta_{wL})$ .

In either case,  $\hat{\beta}_M$  is a  $\sqrt{M}$ -consistent and asymptotically normally distributed estimator of  $\beta_{-1}$ . The dimension of the estimators depends on whether or not the econometrician assumes  $\beta_i \equiv \beta$  for all agents  $i$ . 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 6.1** (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 6.2** (Smooth Population Quantities). *The components of  $\frac{\partial P(y=0_{1 \times N}|z)}{\partial z}$  and  $\frac{\partial p(z)}{\partial z}(1[y = 0_{1 \times N}], z')$  have finite second moments. The  $E(1[y = 0_{1 \times N}] \partial^r p(z))$  exists for  $0 < r \leq S + 1$ . There is a function  $m(z)$  such that  $E((1 + 1[y = 0_{1 \times N}] + \|z\|)m(z))^2 < \infty$ ,  $\|\frac{\partial p(z+\psi)}{\partial z} - \frac{\partial p(z)}{\partial z}\| < m(z)\|\psi\|$ , and  $\|\frac{\partial p(z+\psi)P(y=0_{1 \times N}|z+\psi)}{\partial z} - \frac{\partial p(z)P(y=0_{1 \times N}|z)}{\partial z}\| < m(z)\|\psi\|$ .*

**ASSUMPTION 6.3** (Higher Order Kernel). *The function  $K(\cdot)$  is a kernel of order  $S$  that is symmetric about the origin, bounded, and differentiable.*

**ASSUMPTION 6.4** (Bandwidth Rate). *It holds that  $Mh_M^{2S} \rightarrow 0$  and  $Mh_M^{d+2} \rightarrow \infty$ .*

Finally, for the covariance matrix, let

$$B_i(\delta) = \underbrace{\begin{pmatrix} \frac{-\delta_{w1}}{\left(\sum_i \delta_{ix1}\right)^2} & 0 & \dots & 0 \\ \frac{-\delta_{w2}}{\left(\sum_i \delta_{ix1}\right)^2} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \frac{-\delta_{wL}}{\left(\sum_i \delta_{ix1}\right)^2} & 0 & \dots & 0 \end{pmatrix}}_{L \times D_i} \quad \text{and}$$

$$B(\delta) = \begin{pmatrix} B_1(\delta) & B_2(\delta) & \dots & B_N(\delta) & \frac{1}{\sum_i \delta_{ix1}} I_{L \times L} \end{pmatrix}.$$

In case it is assumed that  $\beta_i \equiv \beta$  for all agents  $i$ , let

$$A(\delta) = \left( \underbrace{(1 \ 1 \ \dots \ 1)}_{1 \times N} \otimes \underbrace{\begin{pmatrix} -\sum_i \delta_{ix2} & \frac{1}{\sum_i \delta_{ix1}} & 0 & \dots & 0 \\ \left(\sum_i \delta_{ix1}\right)^2 & \sum_i \delta_{ix1} & & & \\ -\sum_i \delta_{ix3} & & \frac{1}{\sum_i \delta_{ix1}} & \dots & 0 \\ \left(\sum_i \delta_{ix1}\right)^2 & 0 & \sum_i \delta_{ix1} & & \\ \dots & & & & \\ -\sum_i \delta_{ixD} & & & & \\ \left(\sum_i \delta_{ix1}\right)^2 & 0 & & \dots & \frac{1}{\sum_i \delta_{ix1}} \end{pmatrix}}_{(D-1) \times D} \right).$$

Because  $D_i = D$  in this case, the matrix  $A(\delta)$  is  $(D-1) \times ND$  and  $B(\delta)$  is  $L \times (ND + L)$ . The  $k$ th row of  $A(\delta)$  has nonzero entries in the columns: 1,  $k+1$ ,  $D+1$ ,  $D+k+1$ , up to  $(N-1)D+1$  and  $(N-1)D+k+1$ . The  $l$ th row of  $B(\delta)$  has nonzero entries in the columns: 1,  $D+1$ ,  $2D+1$ , up to  $(N-1)D+1$ , and  $ND+l$ . Finally, let  $C(\delta) = \begin{pmatrix} A(\delta) & 0 \\ B(\delta) & \end{pmatrix}$ . The matrix  $C(\delta)$  is  $((D-1)+L) \times (ND+L)$ .

If it is not assumed that  $\beta_i \equiv \beta$  for all agents  $i$ , let

$$A_i(\delta) = \underbrace{\begin{pmatrix} \frac{-\delta_{ix2}}{(\delta_{ix1})^2} & \frac{1}{\delta_{ix1}} & 0 & \dots & 0 \\ \frac{-\delta_{ix3}}{(\delta_{ix1})^2} & 0 & \frac{1}{\delta_{ix1}} & \dots & 0 \\ \dots & & & & \\ \frac{-\delta_{ixD_i}}{(\delta_{ix1})^2} & 0 & & \dots & \frac{1}{\delta_{ix1}} \end{pmatrix}}_{(D_i-1) \times D_i} \quad \text{and}$$

$$A(\delta) = \begin{pmatrix} A_1(\delta) & 0 & \dots & 0 \\ 0 & A_2(\delta) & 0 & 0 \\ 0 & & \dots & \\ 0 & 0 & \dots & A_N(\delta) \end{pmatrix}.$$

Let  $B(\delta)$  remain as defined as above. Finally, let  $C(\delta) = \begin{pmatrix} A(\delta) & 0 \\ B(\delta) & \end{pmatrix}$ . The matrix  $C(\delta)$  is  $(\sum_i (D_i - 1) + L) \times (\sum_i D_i + L)$ .

**THEOREM 6.1** (Estimation of  $\beta$ ). *Suppose that  $\beta_{iw} \equiv \beta_w$  and  $\Delta_i \leq 0$  for all agents  $i$ . Under Assumptions 3.1, 3.2, 3.3, 3.4, 6.1, 6.2, 6.3, and 6.4, it holds that*

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

where

$$\begin{aligned} V(\delta) &= 4E\left(\left(p(z)\frac{\partial P(y=0_{1 \times N}|z)}{\partial z} - (1[y=0_{1 \times N}] - P(y=0_{1 \times N}|z))\frac{\partial p(z)}{\partial z}\right)\right. \\ &\quad \times \left.\left(p(z)\frac{\partial P(y=0_{1 \times N}|z)}{\partial z} - (1[y=0_{1 \times N}] - P(y=0_{1 \times N}|z))\frac{\partial p(z)}{\partial z}\right)\right) \\ &\quad - 4\delta\delta'. \end{aligned}$$

## 6.2 Estimation of intercept parameters

Theorem 3.1 suggests an estimator for  $\alpha$  and  $\Delta$ : estimate  $\frac{\partial^N P(y=0_{1 \times N}|c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{a_1, a_2, \dots, a_N}$  and  $(-1)^N \frac{\partial^N P(y=1_{1 \times N}|c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{b_1, b_2, \dots, b_N}$  by nonparametric regressions of  $1[y_m = 0_{1 \times N}]$  and  $1[y_m = 1_{1 \times N}]$  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  $\beta$ ; then, estimate  $\alpha$  and  $\alpha + N_0\Delta$  by maximizing the first stage estimators and estimate  $\Delta$  by appropriately differencing these two estimates. So, let

$$\begin{aligned} \hat{Q}_M(\gamma) &= \left( \frac{\partial^N P_M(y=0_{1 \times N}|\hat{c}_1, \hat{c}_2, \dots, \hat{c}_N)}{\partial \hat{c}_1 \partial \hat{c}_2 \dots \partial \hat{c}_N} \Big|_{a_1, a_2, \dots, a_N} \right. \\ &\quad \left. (-1)^N \frac{\partial^N P_M(y=1_{1 \times N}|\hat{c}_1, \hat{c}_2, \dots, \hat{c}_N)}{\partial \hat{c}_1 \partial \hat{c}_2 \dots \partial \hat{c}_N} \Big|_{b_1, b_2, \dots, b_N} \right)' \end{aligned}$$

be a  $2 \times 1$  vector objective function, where  $\gamma = (a_1, a_2, \dots, a_N, b_1, b_2, \dots, b_N)$ , and

$$\begin{aligned} &\frac{\partial^N P_M(y=0_{1 \times N}|\hat{c}_1, \hat{c}_2, \dots, \hat{c}_N)}{\partial \hat{c}_1 \partial \hat{c}_2 \dots \partial \hat{c}_N} \Big|_{a_1, a_2, \dots, a_N} \quad \text{and} \\ &\frac{\partial^N P_M(y=1_{1 \times N}|\hat{c}_1, \hat{c}_2, \dots, \hat{c}_N)}{\partial \hat{c}_1 \partial \hat{c}_2 \dots \partial \hat{c}_N} \Big|_{b_1, b_2, \dots, b_N} \end{aligned}$$

are nonparametric estimators based on generated regressors  $\hat{c}_m = (\hat{c}_{m1}, \hat{c}_{m2}, \dots, \hat{c}_{mN})$ .

Let  $\gamma_0 = (\alpha_1, \alpha_2, \dots, \alpha_N, \alpha_1 + N_0\Delta_1, \alpha_2 + N_0\Delta_2, \dots, \alpha_N + N_0\Delta_N)$ .

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

$$\begin{aligned} \widehat{\Delta Q}_M(\gamma) &= \left( \frac{\partial^N P_M(y=0_{1 \times N}|c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{a_1, a_2, \dots, a_N} \right. \\ &\quad \left. (-1)^N \frac{\partial^N P_M(y=1_{1 \times N}|c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{b_1, b_2, \dots, b_N} \right)' \end{aligned}$$

$$- \left( \frac{\partial^N P_M(y = 0_{1 \times N} | \hat{c}_1, \hat{c}_2, \dots, \hat{c}_N)}{\partial \hat{c}_1 \partial \hat{c}_2 \dots \partial \hat{c}_N} \Big|_{a_1, a_2, \dots, a_N} \right. \\ \left. (-1)^N \frac{\partial^N P_M(y = 1_{1 \times N} | \hat{c}_1, \hat{c}_2, \dots, \hat{c}_N)}{\partial \hat{c}_1 \partial \hat{c}_2 \dots \partial \hat{c}_N} \Big|_{b_1, b_2, \dots, b_N} \right)',$$

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

**ASSUMPTION 6.5 (Compact Parameter Space).** *The parameter space  $\Gamma$  is compact.*

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

**ASSUMPTION 6.6 (Uniform Convergence of Regression Estimate).** *The infeasible  $Q_M(\gamma)$  converges in probability to*

$$Q(\gamma) = \left( \frac{\partial^N P_M(y = 0_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{a_1, a_2, \dots, a_N} \right. \\ \left. (-1)^N \frac{\partial^N P_M(y = 1_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{b_1, b_2, \dots, b_N} \right)'$$

*uniformly over the parameter space  $\Gamma$ .*

Assumption 6.6 requires the usual uniform convergence properties of a nonparametric regression estimator over a compact set, for the *infeasible* estimator that uses the true regressors  $c = (c_1, c_2, \dots, c_N)$ . Therefore, conditions under which it holds can be found in the literature on nonparametric regression. See Remark 6.3.

Now let parameter  $B$  be a particular specification of the  $\beta$  parameter (for example, an estimate  $\hat{\beta}$ ). Let  $c_{im}(B) = -x_{im}B_{ix} - w_mB_{iw}$  be the generated regressors at parameter  $B$ . Let  $\tilde{R}_{M,1}(a_1, a_2, \dots, a_N, B) = \frac{\partial^N P_M(y=0_{1 \times N} | c_1(B), c_2(B), \dots, c_N(B))}{\partial c_1(B) \partial c_2(B) \dots \partial c_N(B)} \Big|_{a_1, a_2, \dots, a_N}$  and  $\tilde{R}_{M,2}(b_1, b_2, \dots, b_N, B) = (-1)^N \frac{\partial^N P_M(y=1_{1 \times N} | c_1(B), c_2(B), \dots, c_N(B))}{\partial c_1(B) \partial c_2(B) \dots \partial c_N(B)} \Big|_{b_1, b_2, \dots, b_N}$ . So, evaluated at  $B = \hat{\beta}$ , these are the feasible estimates, and evaluated at  $B = \beta$ , these are the infeasible estimates.

**ASSUMPTION 6.7 (Lipschitz Properties of Feasible and Infeasible Estimators).** *The terms  $\tilde{R}_{M,1}(a_1, a_2, \dots, a_N, B)$  and  $\tilde{R}_{M,2}(b_1, b_2, \dots, b_N, B)$  have continuous derivatives with respect to  $B$ . If  $\hat{\beta} - \beta = O_p(M^{-1/2})$ , then  $\sup_{a_1, a_2, \dots, a_N, \|\hat{\beta} - \beta\| \leq \|\hat{\beta} - \beta\|} \left\| \frac{\partial \tilde{R}_{M,1}(a_1, a_2, \dots, a_N, B)}{\partial B} \Big|_{\hat{\beta}} \right\| = O_p(1)$  and  $\sup_{b_1, b_2, \dots, b_N, \|\hat{\beta} - \beta\| \leq \|\hat{\beta} - \beta\|} \left\| \frac{\partial \tilde{R}_{M,2}(b_1, b_2, \dots, b_N, B)}{\partial B} \Big|_{\hat{\beta}} \right\| = O_p(1)$ .*

Assumption 6.7 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 6.7 by a Taylor series approximation to  $\hat{Q}_M(\gamma)$ . (See the proof of Theorem 6.2 for the details.)

Since  $c_{im}(B)$  depends smoothly on  $B$ , most nonparametric estimators will satisfy the first part of the assumption: the existence of a continuous derivative with respect to  $B$ . The second part of the assumption can be established by application of a uniform law of large numbers. In particular, in the case that the estimation is by 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. These assumptions mainly require that there is not a parameter on the boundary problem, 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  $2N \times 1$  vector: element  $i \in \{1, \dots, N\}$  is

$$\frac{\partial}{\partial c_i} \frac{\partial^N P(y = 0_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{a_1, a_2, \dots, a_N},$$

and element  $N + i \in \{N + 1, \dots, 2N\}$  is

$$(-1)^N \frac{\partial}{\partial c_i} \frac{\partial^N P(y = 1_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{b_1, b_2, \dots, b_N}.$$

Let  $\frac{\partial^2 Q(\gamma)}{\partial \gamma \partial \gamma'}$  be the  $2N \times 2N$  matrix of derivatives of  $\frac{\partial Q(\gamma)}{\partial \gamma}$  with respect to  $\gamma'$ . Let  $\frac{\partial Q_M(\gamma)}{\partial \gamma}$  and  $\frac{\partial^2 Q_M(\gamma)}{\partial \gamma \partial \gamma'}$  be the infeasible nonparametric estimators based on  $c$ , and let  $\frac{\partial \hat{Q}_M(\gamma)}{\partial \gamma}$  and  $\frac{\partial^2 \hat{Q}_M(\gamma)}{\partial \gamma \partial \gamma'}$  be the feasible nonparametric estimators based on  $\hat{c}$ .

**ASSUMPTION 6.8 (Parameter in the Interior).** *The true value  $\gamma_0$  is in the interior of  $\Gamma$ .*

**ASSUMPTION 6.9 (Asymptotic Distribution of Derivatives).** *The term  $\frac{\partial Q_M(\gamma)}{\partial \gamma} |_{\gamma_0}$  exists and  $r_M \frac{\partial Q_M(\gamma)}{\partial \gamma} |_{\gamma_0} \rightarrow^d N(0, \Omega_0)$  at the rate  $r_M$  with  $r_M M^{-1/2} \rightarrow 0$ .*

**ASSUMPTION 6.10 (Uniform Convergence of Derivatives).** *The term  $\frac{\partial^2 Q_M(\gamma)}{\partial \gamma \partial \gamma'}$  exists and is continuous on a neighborhood of  $\gamma_0$  and converges in probability to  $\frac{\partial^2 Q(\gamma)}{\partial \gamma \partial \gamma'}$  uniformly over a neighborhood of  $\gamma_0$ .*

**ASSUMPTION 6.11 (Lipschitz Properties of Feasible and Infeasible Estimators).** *The conditions in Assumption 6.7 hold for all components of  $\frac{\partial Q_M(\gamma)}{\partial \gamma}$  and  $\frac{\partial^2 Q_M(\gamma)}{\partial \gamma \partial \gamma'}$ .*

**ASSUMPTION 6.12 (Negative Definite Hessian).** *The Hessian of  $f_\varepsilon(\cdot)$  is negative definite on a neighborhood of  $0_{1 \times N}$ .*

**ASSUMPTION 6.13 (Smooth Feasible Objective Function).** *The components of  $\hat{Q}_M(\gamma)$  have continuous second derivatives with respect to  $\gamma$  in a neighborhood of  $\gamma_0$ .*

If the econometrician does not assume that  $\alpha_i \equiv \alpha$  and  $\Delta_i \equiv \Delta$  for all agents  $i$ , the estimator is  $\hat{\psi}_M = (\hat{\alpha}_{M,1}, \hat{\alpha}_{M,2}, \dots, \hat{\alpha}_{M,N}, \hat{\Delta}_{M,1}, \hat{\Delta}_{M,2}, \dots, \hat{\Delta}_{M,N})$ , where for all agents  $i$ ,  $\hat{\alpha}_{M,i} = \hat{\gamma}_{M,i}$  and  $\hat{\Delta}_{M,i} = \frac{1}{N_0}(\hat{\gamma}_{M,N+i} - \hat{\gamma}_{M,i})$ . The corresponding true value is  $\psi = (\alpha_1, \alpha_2, \dots, \alpha_N, \Delta_1, \Delta_2, \dots, \Delta_N) = (\alpha, \Delta)$ .

If the econometrician assumes that  $\alpha_i \equiv \alpha$  and  $\Delta_i \equiv \Delta$  for all agents  $i$ , by imposing that condition on the objective function, the constrained objective function is  $Q^r(\gamma) = Q(\underbrace{\gamma_1, \gamma_1, \dots, \gamma_1}_{N \text{ times}}, \underbrace{\gamma_2, \gamma_2, \dots, \gamma_2}_{N \text{ times}})$ , where  $\gamma = (a, b)$ . In that case, the estimator is  $\hat{\psi}_M = (\hat{\alpha}_M, \hat{\Delta}_M)$ , where  $\hat{\alpha}_M = \hat{\gamma}_{M,1}$  and  $\hat{\Delta}_M = \frac{1}{N_0}(\hat{\gamma}_{M,2} - \hat{\gamma}_{M,1})$ . The true value is  $\psi = (\alpha, \Delta)$ . Note that this involves a different definition of  $\alpha$  and  $\Delta$  compared to above, where  $\alpha$  and  $\Delta$  were vectors. The dimension of the estimators depends on whether or not the econometrician assumes  $\alpha_i \equiv \alpha$  and  $\Delta_i \equiv \Delta$  for all agents  $i$ .

In the statement of the theorem, let  $1_{1 \times N}$  be the  $1 \times N$  vector of 1s and let  $\frac{\partial^2 Q^r(\gamma)}{\partial \gamma \partial \gamma'} = \begin{pmatrix} \frac{\partial^2 Q^r(\gamma)}{\partial^2 a} & 0 \\ 0 & \frac{\partial^2 Q^r(\gamma)}{\partial^2 b} \end{pmatrix}$ , where  $\frac{\partial^2 Q^r(\gamma)}{\partial^2 a} = \sum_{i,j} \frac{\partial^2}{\partial c_i \partial c_j} \frac{\partial^N P(y=0_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} |_{a, a, \dots, a}$  and  $\frac{\partial^2 Q^r(\gamma)}{\partial^2 b} = \sum_{i,j} (-1)^N \frac{\partial^2}{\partial c_i \partial c_j} \frac{\partial^N P(y=1_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} |_{b, b, \dots, b}$ .

**THEOREM 6.2 (Estimation of  $\alpha$  and  $\Delta$ ).** *Suppose that  $\Delta_i \leq 0$  for all agents  $i$ . Suppose that  $\beta$  is point identified and that  $\hat{\beta}$  is an estimator of  $\beta$  such that  $\hat{\beta} - \beta = O_p(M^{-1/2})$ .*

*Under Assumptions 3.1, 3.2, 3.3, 3.5, 6.5, 6.6, and 6.7,*

$$\hat{\alpha}_M \rightarrow^P \alpha \quad \text{and} \quad \hat{\Delta}_M \rightarrow^P \Delta.$$

*Under the additional Assumptions 6.8, 6.9, 6.10, 6.11, 6.12, and 6.13,*

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

*If it is not assumed that  $\alpha_i \equiv \alpha$  and  $\Delta_i \equiv \Delta$  for all agents  $i$ , then*

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

$$\text{where } C = \begin{pmatrix} I_{N \times N} & 0 \\ -\frac{1}{N_0} I_{N \times N} & \frac{1}{N_0} I_{N \times N} \end{pmatrix}.$$

*If it is assumed that  $\alpha_i \equiv \alpha$  and  $\Delta_i \equiv \Delta$  for all agents  $i$ , then*

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

$$\text{where } C = \begin{pmatrix} 1 & 0 \\ -\frac{1}{N_0} & \frac{1}{N_0} \end{pmatrix} \text{ and } D = \begin{pmatrix} 1_{1 \times N} & 0 \\ 0 & 1_{1 \times N} \end{pmatrix}.$$

**REMARK 6.1 (Consistency).** The first order conditions are used to establish the asymptotic distribution, but not consistency. If the first order conditions were used to establish consistency, local but not global maxima of  $f_\varepsilon(\cdot)$  could cause inconsistency.

REMARK 6.2 (Rate of Convergence). The rate of convergence comes from Assumption 6.9, related to estimating an  $(N + 1)$ st derivative of a regression function with  $N$  explanatory variables. Per the identification result, that regression function concerns an integral of the density of  $\varepsilon$ , so the rate of convergence will depend on the assumed smoothness of the density of  $\varepsilon$ . Estimation is not subject to the curse of dimensionality as a function of the number of explanatory variables (i.e.,  $D_i$  or  $L$ ), since the regression concerns  $c \in \mathbb{R}^N$ . This reflects a dimension reduction strategy. However, estimation is subject to the curse of dimensionality as a function of the number of agents.

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  $\alpha$  under the same assumptions. (The interaction effect disappears in this submodel, requiring the arguments of Khan and Nekipelov (2012).)

REMARK 6.3 (Nonparametric 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,  $\Omega_0$  from Assumption 6.9 can be found in standard references, as it is the asymptotic covariance of estimates of the derivatives of a regression function.

### 6.3 Estimation with discrete explanatory variables

This section shows the extension of the estimator allowing discrete explanatory variables. Corresponding to the identification result (i.e., Theorem 4.1), the estimator for the slope coefficients on the continuous explanatory variables, the interaction effects, and  $\alpha_{i,z^{(d)}}$  follows from an application of the above estimation results to the subsets of the data defined by the value of the discrete explanatory variables.

The estimator of the slope coefficients on the discrete explanatory variables and the true intercept parameters  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$  is the sample analog of the constructive identification result in Theorem 4.1. The properties of that estimator are derived from the properties of the above estimator as follows.

Let  $p(z^{(d)})$  be the population probability that the discrete explanatory variables are equal to  $z^{(d)}$  and let  $p_M(z^{(d)})$  be the sample fraction of the data set with discrete explanatory variables equal to  $z^{(d)}$ . Also let  $E_M((\tilde{z}_i^{(d)})' \tilde{z}_i^{(d)})$  be the sample analog of  $E((\tilde{z}_i^{(d)})' \tilde{z}_i^{(d)})$ . The  $\tilde{z}_i^{(d)}$  notation is defined in Section 4.1.

Then, because  $z^{(d)}$  has a discrete distribution, the estimator for  $\eta_i$  is

$$\hat{\eta}_i = (E_M((\tilde{z}_i^{(d)})' \tilde{z}_i^{(d)}))^{-1} \left( \sum_{z^{(d)} \in \mathcal{Z}^{(d)}} p_M(z^{(d)}) (\tilde{z}_i^{(d)})' \hat{\alpha}_{i,z^{(d)}} \right).$$

Under regularity conditions, including Assumption 4.4,  $(E_M((\tilde{z}_i^{(d)})' \tilde{z}_i^{(d)}))^{-1}$  converges to  $(E((\tilde{z}_i^{(d)})' \tilde{z}_i^{(d)}))^{-1}$  at the parametric rate. Further, since  $\hat{\alpha}_{i,z^{(d)}}$  converges at slower than parametric rate for each  $z^{(d)} \in \mathcal{Z}^{(d)}$ , per Theorem 6.2, while  $p_M(z^{(d)})$  converges at the parametric rate under regularity conditions, it holds that  $r_M \sum_{z^{(d)} \in \mathcal{Z}^{(d)}} (p_M(z^{(d)}) - p(z^{(d)}))(\tilde{z}_i^{(d)})' \hat{\alpha}_{i,z^{(d)}}$  is asymptotically negligible since  $r_M(p_M(z^{(d)}) - p(z^{(d)})) \rightarrow^P 0$  for all  $z^{(d)} \in \mathcal{Z}^{(d)}$  because  $r_M$  is slower than the parametric rate by Assumption 6.9. And so the asymptotic distribution of  $r_M(\hat{\eta}_i - \eta_i)$  is the asymptotic distribution of

$$\begin{aligned} & r_M \left( (E_M((\tilde{z}_i^{(d)})' \tilde{z}_i^{(d)}))^{-1} \left( \sum_{z^{(d)} \in \mathcal{Z}^{(d)}} p_M(z^{(d)}) (\tilde{z}_i^{(d)})' \hat{\alpha}_{i,z^{(d)}} \right) \right. \\ & \quad \left. - (E((\tilde{z}_i^{(d)})' \tilde{z}_i^{(d)}))^{-1} \left( \sum_{z^{(d)} \in \mathcal{Z}^{(d)}} p(z^{(d)}) (\tilde{z}_i^{(d)})' \alpha_{i,z^{(d)}} \right) \right) \\ & = r_M \left( (E((\tilde{z}_i^{(d)})' \tilde{z}_i^{(d)}))^{-1} \right. \\ & \quad \left. \times \left( \sum_{z^{(d)} \in \mathcal{Z}^{(d)}} p_M(z^{(d)}) (\tilde{z}_i^{(d)})' \hat{\alpha}_{i,z^{(d)}} - \sum_{z^{(d)} \in \mathcal{Z}^{(d)}} p(z^{(d)}) (\tilde{z}_i^{(d)})' \alpha_{i,z^{(d)}} \right) \right) + o_p(1) \\ & = \left( (E((\tilde{z}_i^{(d)})' \tilde{z}_i^{(d)}))^{-1} \left( \sum_{z^{(d)} \in \mathcal{Z}^{(d)}} p(z^{(d)}) (\tilde{z}_i^{(d)})' r_M(\hat{\alpha}_{i,z^{(d)}} - \alpha_{i,z^{(d)}}) \right) \right) + o_p(1), \end{aligned}$$

which can be evaluated using the results of Theorem 6.2 to find the asymptotic distributions of  $r_M(\hat{\alpha}_{i,z^{(d)}} - \alpha_{i,z^{(d)}})$  for each  $z^{(d)} \in \mathcal{Z}^{(d)}$ . In particular, to derive the asymptotic distribution of that linear combination of  $(\hat{\alpha}_{i,z^{(d)}} - \alpha_{i,z^{(d)}})$  terms, note that the asymptotic joint distribution of  $(\hat{\alpha}_{i,z^{(d)}} - \alpha_{i,z^{(d)}})$  and  $(\hat{\alpha}_{i,z'^{(d)}} - \alpha_{i,z'^{(d)}})$  for  $z^{(d)} \neq z'^{(d)}$  is independent, since the estimators use disjoint partitions of the data.

## 7. MONTE CARLO EXPERIMENT

This section reports the results of a Monte Carlo experiment. Based on the normal form in Table 1, the specification of the true data generating process with  $N = 2$  is as follows:

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

The explanatory variables are a translation (by arctan) of positively correlated normal random variables. The arctan translation is used to generate explanatory variables that have bounded support. Because of the scaling by  $\frac{1}{\pi}$ , the support of each explanatory variable is the interval  $[-\frac{1}{2}, \frac{1}{2}]$ . The positive correlation reflects the notion that the



various observable components of profits are likely positively related.<sup>26</sup> The positive correlation of the unobservables reflects an unobserved market fixed effect.<sup>27</sup> The econometrician knows  $\Delta_1 \leq 0$  and  $\Delta_2 \leq 0$ .

The experiment is run twice: when the sample has  $M = 500$  and when the sample has  $M = 1000$ . For each sample size, 1000 such samples are generated, and the estimates recorded. The estimator imposes that the parameters are equal across player roles. Since only the unique potential outcomes are used, *any* selection mechanism would result in exactly the same numerical results.

Estimation requires two kernels: for estimating  $\beta$ , and for estimating  $\alpha$  and  $\Delta$  based on a kernel regression estimator. The kernel for estimation of  $\beta$  is the product of four fourth order Gaussian kernels, where the order is per Assumption 6.1. The kernel for estimation of  $\alpha$  and  $\Delta$  is the product of two fourth order Gaussian kernels. Similarly, the estimator requires two bandwidths.<sup>28</sup> Based on these bandwidths, the estimation theorems imply that the rate of convergence of the estimator for  $\beta$  is  $M^{1/2}$  and the rate of convergence of the estimator for  $\alpha$  and  $\Delta$  is  $M^{1/4}$ .

Table 2 reports: the mean, median, mean square error (MSE), interquartile range (IQR), and variance of the estimator in the experiment. Also, in the last column, Table 2 reports the empirical probability of rejecting the truth (Rej. Rate) 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 analogs. The results suggest the estimators and associated tests have good performance. The estimators seem approximately mean and me-

TABLE 2. Numerical results of Monte Carlo experiment.

Parameter	Mean	Median	MSE	IQR	Variance	Rej. Rate
$M = 500$						
$\beta = 0.750$	0.745	0.734	0.016	0.178	0.016	0.094
$\alpha = 0.100$	0.093	0.087	0.012	0.140	0.012	0.120
$\Delta = -0.200$	-0.193	-0.197	0.025	0.195	0.025	0.098
$M = 1000$						
$\beta = 0.750$	0.738	0.733	0.009	0.126	0.009	0.120
$\alpha = 0.100$	0.105	0.098	0.009	0.126	0.009	0.076
$\Delta = -0.200$	-0.216	-0.215	0.019	0.160	0.019	0.059

*Note:* MSE is mean square error, IQR is interquartile range, and Rej. Rate is the empirical rejection rate of the test at 0.10 significance level.

<sup>26</sup>The explanatory variables  $x_{ik}$  have approximately variances 0.0455 and covariances 0.0043.

<sup>27</sup>The unobservables have variance 0.04 and covariance 0.01 for a correlation of 0.25.

<sup>28</sup>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  $\beta$  is proportional to  $M^{-1/7}$  (as the optimal bandwidth for density-weighted average derivative estimation with four regressors) and the bandwidth for the estimator corresponding to  $\alpha$  and  $\Delta$  is proportional to  $M^{-1/16}$  (as the optimal bandwidth for kernel regression estimation of a third derivative with two regressors and a fourth order kernel). The bandwidth for the estimator corresponding to  $\beta$  is based on the plug-in estimator for density-weighted average derivatives in Powell and Stoker (1996).

TABLE 3. Numerical results of Monte Carlo experiment, empirical rates of convergence.

Parameter	Standard Deviation	Root MSE	Median Absolute Deviation
$\beta$	0.499	0.487	0.546
$\alpha$	0.206	0.206	0.200
$\Delta$	0.223	0.216	0.297

dian unbiased, and the tests reject the truth at roughly the nominal rate of  $p = 0.10$ . To check the empirical rate of convergence, the Monte Carlo experiment is repeated for sample sizes  $M = 500, 600, \dots, 1000$ , and for each sample size and each parameter, the following statistics are recorded: the standard deviation, the root mean square error, and the median absolute deviation (from the truth). Then the logarithm of those statistics is “regressed” against the logarithm of the sample size. The (negative of the) slope of that relationship is the empirical rate of convergence, displayed in Table 3.<sup>29</sup>

Consistent with the theory, empirically the rate of convergence for  $\beta$  is approximately  $\frac{1}{2}$ , and empirically the rate of convergence for  $\alpha$  and  $\Delta$  is approximately  $\frac{1}{4}$ , for each of standard deviation, root MSE, and median absolute deviation.

## 8. EMPIRICAL APPLICATION

This section presents the results of a very brief, stylized, empirical example to entry in airline markets, which has been previously studied, for example, in Berry (1992) and Ciliberto and Tamer (2009). The data come from the second quarter of the 2010 Airline Origin and Destination Survey (DB1B), as described in more detail in an empirical application of partial identification methods in Kline and Tamer (2015). The data contain 7882 markets, which are formally defined as trips between two airports irrespective of intermediate stops. The empirical question concerns the entry behavior of two kinds of firms: LCC (low cost carriers)<sup>30</sup> and OA (other airlines). A firm that is not an LCC is by definition an OA.

As in Kline and Tamer (2015), for the purposes of mapping the data to an entry game, the airlines are aggregated into two firms: LCC and OA. So, firm LCC enters the market if any low cost carrier serves that market, and similarly for firm OA entering the market. There are two explanatory variables: *market presence* and *market size*. Market presence is a market- and airline-specific variable: for each airline and for each airport, compute the number of markets that airline serves from that airport, divided by the total number of markets served from that airport by any airline. The market presence variable for a given market and airline is the average of these ratios (excluding the one market under consideration) at the two endpoints of the trip, providing some proxy for an airline’s presence in the airports associated with that market. (See Berry (1992) for more on

<sup>29</sup>If the statistic has form  $\frac{\sigma}{M^\alpha}$ , then the logarithm of the statistic has form  $\log(\sigma) - \alpha \log(M)$ .

<sup>30</sup>The low cost carriers are AirTran, Allegiant Air, Frontier, JetBlue, Midwest Air, Southwest, Spirit, Sun Country, USA3000, and Virgin America.

TABLE 4. Estimation results for a model of entry in airline markets.

Parameter	Estimate	90% Confidence Interval
$\alpha_{LCC}$	-0.0497	[-0.0708, -0.0286]
$\alpha_{OA}$	-0.3335	[-0.3722, -0.2947]
$\Delta_{LCC}$	-0.0192	[-0.0520, 0.0000]
$\Delta_{OA}$	-0.0557	[-0.0991, -0.0123]

this variable.) This variable is an *agent-specific* explanatory variable: the market presence for LCC enters only LCC's payoffs, and the market presence for OA enters only OA's payoffs. Since the airlines are aggregated into two firms (LCC and OA), the market presence variable must also be aggregated: the market presence for the LCC firm (resp., OA firm) is the maximum among the actual airlines in the LCC category (resp., OA category). Market size is a market-specific variable (but shared by all airlines in that market): the population at the endpoints of the trip.

The analysis proceeds conditionally on markets with a market size of less than 1,500,000. These are "small to moderate" markets. In larger markets, particularly considering the stylized nature of this analysis that considers only entry decisions of two firms, rather than the intensity of service provided by many firms, it is unrealistic to detect substantial competitive effects of entry. There are  $M = 5837$  markets with a market size less than 1,500,000.

Estimation proceeds as in Sections 6 and 7, and the results are displayed in Table 4. The slope coefficients on the market presence explanatory variable are normalized to be 1. Since market presence is in percentage points between 0 and 1, that normalization implies that 100 times the other parameters can be interpreted as the effects relative to the effect of 1 percentage point of market presence. The intercept parameter in this utility function is the payoff to a firm that enters a market with no competition, and no market presence, and an unobservable at 0. The results suggest that such payoff for a low-cost carrier is relatively larger compared to other airlines, relative to their effects of market presence. The competitive effects seem to be much more similar across types of carrier, equivalent to either 2 or 6 percentage points of market presence, although there is at least some evidence that the magnitude of the competitive effect of entry is greater on other airlines compared to low cost carriers.

## 9. CONCLUSIONS AND DISCUSSION

This paper shows that it is possible to point identify the parameters of the utility functions in a complete information game without a large support regressor, based on a non-standard but plausible assumption on the mode of the unobservables. The resulting estimator is consistent and asymptotically normal, but nonstandard in the sense that the estimators of the intercept and interaction effect parameters converge at slower than the parametric rate.

More broadly, it is interesting to note that there appears to be an “identification possibility frontier” between the solution concept (e.g., Nash equilibrium or rationalizability (i.e., [Bernheim \(1984\)](#) and [Pearce \(1984\)](#))) and the assumptions on the explanatory variables (e.g., large support or bounded support). This paper has established point identification under the standard assumption of Nash equilibrium, with bounded regressors; in contrast, [Kline \(2015a\)](#) establishes point identification with a large support regressor, under a weaker solution concept related to rationalizability.

Finally, it is worth remarking on identification of the distribution of the unobservables. Under only the assumptions in this paper, it is not possible to point identify the tails of the distribution of the unobservables. For example, suppose that  $\Delta_i \leq 0$  for all agents  $i$ , and let  $\bar{t} = (\min\{x_1\beta_{1x} + w\beta_{1w}\}, \min\{x_2\beta_{2x} + w\beta_{2w}\}, \dots, \min\{x_N\beta_{Nx} + w\beta_{Nw}\})$ , 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. Note that if  $\varepsilon_i > -\alpha_i - \bar{t}_i - N_0\Delta_i$  for all agents  $i$ , then for any realization of the explanatory variables in the support,  $\alpha_i + x_i\beta_{ix} + w\beta_{iw} + N_0\Delta_i + \varepsilon_i > 0$ , so  $1_{1 \times N}$  is necessarily the unique pure strategy Nash equilibrium for all such  $\varepsilon$ . Consequently, any rearrangement of the probability mass of the distribution of  $\varepsilon$  within the region where  $\varepsilon_i > -\alpha_i - \bar{t}_i - N_0\Delta_i$  for all agents  $i$  results in the same distribution over  $P(y|z)$ , and therefore the distribution of  $\varepsilon$  is not point identified.

Arguments similar to the proof of [Theorem 3.1](#) show that the values of the cumulative distribution function of  $\varepsilon$  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  $(\alpha_1 + x_1\beta_{1x} + w\beta_{1w}, \alpha_2 + x_2\beta_{2x} + w\beta_{2w}, \dots, \alpha_N + x_N\beta_{Nx} + w\beta_{Nw})$ . If a parametric family for the distribution of  $\varepsilon$  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  $\varepsilon$  is point identified. For example, this is true for the normal distribution with mean zero but unknown covariance  $\Sigma$ .<sup>31</sup>

## 10. PROOFS

**PROOF OF THEOREM 3.1.** *Proof of identification of  $\beta$ :*

By [Assumption 3.3](#),

$$F_\varepsilon(t_1, t_2, \dots, t_N) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} \dots \int_{-\infty}^{t_N} f_\varepsilon(e_1, e_2, \dots, e_N) de_N \dots de_2 de_1.$$

$F_\varepsilon^{[i,1]}(t_1, t_2, \dots, t_N)$  is the first derivative of  $F_\varepsilon(\cdot)$  with respect to  $t_i$ , evaluated at  $t = (t_1, t_2, \dots, t_N)$ . So,

$$F_\varepsilon^{[i=1,1]}(t_1, t_2, \dots, t_N) = \int_{-\infty}^{t_2} \dots \int_{-\infty}^{t_N} f_\varepsilon(t_1, e_2, e_3, \dots, e_N) de_N \dots de_2$$

<sup>31</sup>This uses the fact that if  $f_\varepsilon(t)$  is the density at  $t$ , and  $\frac{f_\varepsilon(t)}{dt}$  is the  $N \times 1$  vector of derivatives with respect to the arguments, then  $-\frac{f_\varepsilon(t)}{dt}/f_\varepsilon(t) = \Sigma^{-1}t$ , so  $\Sigma$  can be recovered by observing  $-\frac{f_\varepsilon(t)}{dt}/f_\varepsilon(t)$  at linearly independent  $t$ .

and

$$F_{\varepsilon}^{[i=2,1]}(t_1, t_2, \dots, t_N) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_3} \dots \int_{-\infty}^{t_N} f_{\varepsilon}(e_1, t_2, e_3, \dots, e_N) de_N \dots de_3 de_1,$$

and so forth. Since the density is nonnegative,  $0 \leq F_{\varepsilon}^{[i,1]}(t_1, t_2, \dots, t_N) \leq f_{\varepsilon_i}(t_i)$ . If the density were everywhere positive, then  $0 < F_{\varepsilon}^{[i,1]}(t_1, t_2, \dots, t_N)$ .

Suppose that  $\alpha_i + x_i \beta_{ix} + w \beta_{iw} + \varepsilon_i < 0$  for all  $i$ . Then the outcome  $0_{1 \times N}$  is a pure strategy Nash equilibrium. Moreover, since  $\Delta_i \leq 0$ , it is the *unique* pure strategy Nash equilibrium, since  $u_i(1, y_{(-i)}) \leq \alpha_i + x_i \beta_{ix} + w \beta_{iw} + \varepsilon_i < 0$ . Conversely, suppose that  $0_{1 \times N}$  is a pure strategy Nash equilibrium. Then it must be that  $\alpha_i + x_i \beta_{ix} + w \beta_{iw} + \varepsilon_i \leq 0$ . Therefore, since by Assumption 3.3 there is zero probability that  $\alpha_i + x_i \beta_{ix} + w \beta_{iw} + \varepsilon_i = 0$  conditional on any  $z$ , and using Assumption 3.2,  $P(y = 0_{1 \times N} | z) = P(\{\varepsilon_i \leq -\alpha_i - x_i \beta_{ix} - w \beta_{iw}\} \forall i)$ .

By Assumption 3.3,  $\frac{\partial P(y=0_{1 \times N} | z)}{\partial x_{ik}} = F_{\varepsilon}^{[i,1]}(-\alpha_1 - x_1 \beta_{1x} - w \beta_{1w}, -\alpha_2 - x_2 \beta_{2x} - w \beta_{2w}, \dots, -\alpha_N - x_N \beta_{Nx} - w \beta_{Nw})(-\beta_{ixk})$  and  $\frac{\partial P(y=0_{1 \times N} | z)}{\partial w_l} = \sum_i^N (F_{\varepsilon}^{[i,1]}(-\alpha_1 - x_1 \beta_{1x} - w \beta_{1w}, -\alpha_2 - x_2 \beta_{2x} - w \beta_{2w}, \dots, -\alpha_N - x_N \beta_{Nx} - w \beta_{Nw})(-\beta_{iwl}))$ . If the density of  $\varepsilon$  were everywhere positive, then  $\frac{\partial P(y=0_{1 \times N} | z)}{\partial x_{i1}} \neq 0$  for all  $z$  since  $\beta_{ix1} \neq 0$ .

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  $F_{\varepsilon}^{[i,1]}(z, \theta) \equiv F_{\varepsilon}^{[i,1]}(-\alpha_1 - x_1 \beta_{1x} - w \beta_{1w}, -\alpha_2 - x_2 \beta_{2x} - w \beta_{2w}, \dots, -\alpha_N - x_N \beta_{Nx} - w \beta_{Nw})$ . By Assumption 3.4, for any  $z \in \mathcal{Z}_0$ ,  $F_{\varepsilon}^{[i,1]}(z, \theta) > 0$ , since otherwise  $\frac{\partial P(y=0_{1 \times N} | z)}{\partial x_{i1}} = 0$ .

Then  $E(\pi(z) \frac{\partial P(y=0_{1 \times N} | z)}{\partial x_{ik}}) = E(\pi(z) F_{\varepsilon}^{[i,1]}(z, \theta)(-\beta_{ixk}))$ . Also, if  $\beta_{iw} = \beta_w$  for all agents  $i$ , then  $E(\pi(z) \frac{\partial P(y=0_{1 \times N} | z)}{\partial w_l}) = E(\pi(z) \sum_i^N F_{\varepsilon}^{[i,1]}(z, \theta)(-\beta_{iwl}))$ . Since  $\beta_{ix1} = 1$  by Assumption 3.1,  $\beta_{ixk} = E(\pi(z) \frac{\partial P(y=0_{1 \times N} | z)}{\partial x_{ik}}) / E(\pi(z) \frac{\partial P(y=0_{1 \times N} | z)}{\partial x_{i1}})$ . And if  $\beta_{iw} = \beta_w$  for all agents  $i$ , then  $\beta_{wl} = E(\pi(z) \frac{\partial P(y=0_{1 \times N} | z)}{\partial w_l}) / \sum_i^N E(\pi(z) \frac{\partial P(y=0_{1 \times N} | z)}{\partial x_{i1}})$ . Since  $F_{\varepsilon}^{[i,1]}(z, \theta) > 0$  for all  $z \in \mathcal{Z}_0$ , which has positive probability, and  $\pi(z)$  is strictly positive on the support of  $z$  (except possibly the boundary),  $E(\pi(z) \frac{\partial P(y=0_{1 \times N} | z)}{\partial x_{i1}}) < 0$ , so the divisions are justified.

Alternatively, if  $E(P_x(z))$  has full rank, let  $F_{\varepsilon}^{[1]}(z, \theta)$  be the  $1 \times N$  matrix whose  $i$ th entry is  $F_{\varepsilon}^{[i,1]}(z, \theta)$ . It follows that  $F_{\varepsilon}^{[1]}(z, \theta)' \frac{dP(y=0_{1 \times N} | z)}{dw_l} = F_{\varepsilon}^{[1]}(z, \theta)' F_{\varepsilon}^{[1]}(z, \theta)(-\beta_{wl})$ , where  $\beta_{wl}$  is the  $N \times 1$  matrix whose  $i$ th entry is  $\beta_{iwl}$ . So taking expectations and using Assumption 3.6,  $\beta_{wl} = -(E(F_{\varepsilon}^{[1]}(z, \theta)' F_{\varepsilon}^{[1]}(z, \theta)))^{-1} E(F_{\varepsilon}^{[1]}(z, \theta)' \frac{dP(y=0_{1 \times N} | z)}{dw_l})$ . Let  $P^{[1]}(z)$  be the  $1 \times N$  vector, with  $i$ th entry equal to  $\frac{\partial P(y=0_{1 \times N} | z)}{\partial x_{i1}}$ . Since  $\beta_{ix1} = 1$  by Assumption 3.1,  $F_{\varepsilon}^{[1]}(z, \theta) = -P^{[1]}(z)$ .

*Proof of identification of  $\alpha$  and  $\Delta$ :*

As above,  $P(y = 0_{1 \times N} | c) = P(\varepsilon_1 \leq -\alpha_1 + c_1, \varepsilon_2 \leq -\alpha_2 + c_2, \dots, \varepsilon_N \leq -\alpha_N + c_N)$  and  $P(y = 1_{1 \times N} | c) = P(\varepsilon_1 \geq -\alpha_1 - N_0 \Delta_1 + c_1, \varepsilon_2 \geq -\alpha_2 - N_0 \Delta_2 + c_2, \dots, \varepsilon_N \geq -\alpha_N - N_0 \Delta_N + c_N)$ . Thus,  $f_{\varepsilon}(-\alpha_1 + a_1, -\alpha_2 + a_2, \dots, -\alpha_N + a_N) = \frac{\partial^N P(y=0_{1 \times N} | c)}{\partial c_1 \partial c_2 \dots \partial c_N} |_{a_1, a_2, \dots, a_N}$  is point identified at points  $(a_1, a_2, \dots, a_N)$  such that the density of  $(c_1, c_2, \dots, c_N)$  is positive on a neighborhood of  $(a_1, a_2, \dots, a_N)$ . Therefore, by Assumption 3.5 this derivative is point

identified on an open set of  $(a_1, a_2, \dots, a_N)$  that contains  $(\alpha_1, \alpha_2, \dots, \alpha_N)$ . By Assumption 3.3, the mode of  $\varepsilon$  is  $0_{1 \times N}$ . Therefore,

$$(\alpha_1, \alpha_2, \dots, \alpha_N) = \operatorname{argmax}_{a_1, a_2, \dots, a_N} \frac{\partial^N P(y = 0_{1 \times N} | c)}{\partial c_1 \partial c_2 \cdots \partial c_N} \Big|_{a_1, a_2, \dots, a_N}.$$

Similarly,  $(-1)^N f_\varepsilon(-\alpha_1 - N_0 \Delta_1 + b_1, -\alpha_2 - N_0 \Delta_2 + b_2, \dots, -\alpha_N - N_0 \Delta_N + b_N) = \frac{\partial^N P(y = 1_{1 \times N} | c)}{\partial c_1 \partial c_2 \cdots \partial c_N} \Big|_{b_1, b_2, \dots, b_N}$ , so

$$\begin{aligned} & (\alpha_1 + N_0 \Delta_1, \alpha_2 + N_0 \Delta_2, \dots, \alpha_N + N_0 \Delta_N) \\ &= \operatorname{argmax}_{b_1, b_2, \dots, b_N} (-1)^N \frac{\partial^N P(y = 1_{1 \times N} | c)}{\partial c_1 \partial c_2 \cdots \partial c_N} \Big|_{b_1, b_2, \dots, b_N}. \end{aligned}$$

The maximization is over the support of  $(c_1, c_2, \dots, c_N)$ .  $\square$

**LEMMA 10.1.** *Under the same conditions as Theorem 3.1, if the densities of  $\varepsilon_i$  for each agent  $i$  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.** For  $E(\pi(z) \frac{\partial P(y=0_{1 \times N} | z)}{\partial x_{ik}})$ , by the arguments of Theorem 3.1,  $\pi(z) \frac{\partial P(y=0_{1 \times N} | z)}{\partial x_{ik}} = (\pi(z) F_\varepsilon^{[i,1]}(z, \theta))(-\beta_{ixk})$ . If there is an upper bound  $\bar{f}$  for the densities of  $\varepsilon_i$  for each agent  $i$ , then by the arguments of Theorem 3.1,  $F_\varepsilon^{[i,1]}(z, \theta) \leq \bar{f}$ . Consequently,  $(\pi(z) F_\varepsilon^{[i,1]}(z, \theta))(-\beta_{ixk})$  is bounded above by a constant multiple of the integrable  $\pi(z)$ . It is similar for  $E(\pi(z) \frac{\partial P(y=0_{1 \times N} | z)}{\partial w_l})$ . And for  $E(P_x(z))$  and  $E(P^{[1]}(z) \frac{dP(y=0_{1 \times N} | z)}{dw_l})$ , the integrand is bounded above by a constant if there is an upper bound  $\bar{f}$  for the densities of  $\varepsilon_i$  for all agents  $i$ , and therefore is integrable.  $\square$

**PROOF OF THEOREM 4.1.** After conditioning on the discrete explanatory variable, as noted in the text, the model falls into the class of models addressed in Section 3.2 with an intercept that depends on the value of the discrete explanatory variables. Therefore, by applying the statement of Theorem 3.1 to the data and model conditional on  $z^{(d)}$ , the identification results for  $(\beta_{ix}^{(c)}, \beta_{iw}^{(c)}, \Delta_i)$  and  $\alpha_{i, z^{(d)}}$  for each agent  $i$  follow. Then, under Assumption 4.4, it follows that  $\eta_i = (E((\tilde{z}_i^{(d)})' \tilde{z}_i^{(d)}))^{-1} E((\tilde{z}_i^{(d)})' \alpha_{i, z^{(d)}})$ .  $\square$

**PROOF OF THEOREM 5.1.** Suppose that  $\Delta_1 \leq 0$  and  $\Delta_2 \leq 0$ . By the condition for  $y_2 = 1$  to be part of a pure strategy Nash equilibrium, and using Assumptions 3.2 and 3.3,

$$\begin{aligned} P(y_2 = 1 | z) &= \int P(y_2 = 1 | z, \varepsilon) f_\varepsilon(\varepsilon) d\varepsilon \\ &= \int_{\varepsilon \in \mathcal{R}^-(z, \theta)} (1 - p_{z, \varepsilon}) f_\varepsilon(\varepsilon) d\varepsilon + P(\varepsilon_2 \geq -\alpha_2 - x_2 \beta_{2x} - w \beta_{2w} - \Delta_2) \\ &\quad + P(\varepsilon_1 \leq -\alpha_1 - x_1 \beta_{1x} - w \beta_{1w}, \\ &\quad -\alpha_2 - x_2 \beta_{2x} - w \beta_{2w} \leq \varepsilon_2 \leq -\alpha_2 - x_2 \beta_{2x} - w \beta_{2w} - \Delta_2). \end{aligned}$$

Let  $\mathcal{S}(z', z'', \theta) = \{\varepsilon : -\alpha_1 - x''_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w} \leq \varepsilon_1 \leq -\alpha_1 - x'_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w}, -\alpha_2 - x^*_2\beta_{2x} - w^*\beta_{2w} \leq \varepsilon_2 \leq -\alpha_2 - x^*_2\beta_{2x} - w^*\beta_{2w} - \Delta_2\}$ . It follows that

$$\begin{aligned}
& P(y_2 = 1|z'') - P(y_2 = 1|z') \\
&= -P(\varepsilon \in \mathcal{S}(z', z'', \theta)) + \int_{\varepsilon \in \mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)} (p_{z', \varepsilon} - p_{z'', \varepsilon}) f_\varepsilon(\varepsilon) d\varepsilon \\
&\quad + \int_{\varepsilon \in \mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C} (1 - p_{z'', \varepsilon}) f_\varepsilon(\varepsilon) d\varepsilon \\
&\quad - \int_{\varepsilon \in \mathcal{R}^-(z'', \theta)^C \cap \mathcal{R}^-(z', \theta)} (1 - p_{z', \varepsilon}) f_\varepsilon(\varepsilon) d\varepsilon \\
&= -P(\varepsilon \in \mathcal{S}(z', z'', \theta) \cap (\mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C)) \\
&\quad - P(\varepsilon \in \mathcal{S}(z', z'', \theta) \cap (\mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C)^C) \\
&\quad + \int_{\varepsilon \in \mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)} (p_{z', \varepsilon} - p_{z'', \varepsilon}) f_\varepsilon(\varepsilon) d\varepsilon \\
&\quad + \int_{\varepsilon \in \mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C \cap \mathcal{S}(z', z'', \theta)} (1 - p_{z'', \varepsilon}) f_\varepsilon(\varepsilon) d\varepsilon \\
&\quad + \int_{\varepsilon \in \mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C \cap \mathcal{S}(z', z'', \theta)^C} (1 - p_{z'', \varepsilon}) f_\varepsilon(\varepsilon) d\varepsilon \\
&\quad - \int_{\varepsilon \in \mathcal{R}^-(z'', \theta)^C \cap \mathcal{R}^-(z', \theta)} (1 - p_{z', \varepsilon}) f_\varepsilon(\varepsilon) d\varepsilon \\
&= - \int_{\varepsilon \in \mathcal{S}(z', z'', \theta) \cap (\mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C)^C} f_\varepsilon(\varepsilon) d\varepsilon \\
&\quad - \int_{\varepsilon \in \mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)} (p_{z'', \varepsilon} - p_{z', \varepsilon}) f_\varepsilon(\varepsilon) d\varepsilon \\
&\quad - \int_{\varepsilon \in \mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C \cap \mathcal{S}(z', z'', \theta)} p_{z'', \varepsilon} f_\varepsilon(\varepsilon) d\varepsilon \\
&\quad - \int_{\varepsilon \in \mathcal{R}^-(z'', \theta)^C \cap \mathcal{R}^-(z', \theta)} (1 - p_{z', \varepsilon}) f_\varepsilon(\varepsilon) d\varepsilon.
\end{aligned}$$

The last equality uses the fact that  $\mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C \cap \mathcal{S}(z', z'', \theta)^C = \emptyset$ , since  $\mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C \subseteq \mathcal{S}(z', z'', \theta)$ : if  $\varepsilon \in \mathcal{R}^-(z'', \theta)$ , then  $-\alpha_2 - x^*_2\beta_{2x} - w^*\beta_{2w} \leq \varepsilon_2 \leq -\alpha_2 - x^*_2\beta_{2x} - w^*\beta_{2w} - \Delta_2$  and  $-\alpha_1 - x''_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w} \leq \varepsilon_1 \leq -\alpha_1 - x'_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w} - \Delta_1$ . So, if further  $\varepsilon \in \mathcal{R}^-(z', \theta)^C$ , then  $\varepsilon_1 < -\alpha_1 - x'_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w}$  because otherwise  $-\alpha_1 - x'_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w} \leq \varepsilon_1 \leq -\alpha_1 - x'_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w} - \Delta_1$ , which would imply that  $\varepsilon \in \mathcal{R}^-(z', \theta)$ . Therefore, as claimed,  $\mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C \subseteq \mathcal{S}(z', z'', \theta)$ .

If  $\Delta_2 < 0$ , then since  $x''_{11} > x'_{11}$ ,  $\mathcal{S}(z', z'', \theta)$  has positive Lebesgue measure. Also,  $(\mathcal{S}(z', z'', \theta) \cap (\mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C)^C) \cup (\mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C \cap \mathcal{S}(z', z'', \theta)) =$

$\mathcal{S}(z', z'', \theta)$ . Also,  $(\mathcal{S}(z', z'', \theta) \cap (\mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C)^C) \cup (\mathcal{R}^-(z'', \theta)^C \cap \mathcal{R}^-(z', \theta)) = \{\varepsilon : -\alpha_1 - x''_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w} - \Delta_1 < \varepsilon_1 \leq -\alpha_1 - x'_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w} - \Delta_1, -\alpha_2 - x^*_2\beta_{2x} - w^*\beta_{2w} \leq \varepsilon_2 \leq -\alpha_2 - x^*_2\beta_{2x} - w^*\beta_{2w} - \Delta_2\}$  has positive Lebesgue measure if  $\Delta_2 < 0$ . And  $\mathcal{S}(z', z'', \theta) \cap (\mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C)^C = \{\varepsilon : -\alpha_1 - x''_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w} - \Delta_1 < \varepsilon_1 \leq -\alpha_1 - x'_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w}, -\alpha_2 - x^*_2\beta_{2x} - w^*\beta_{2w} \leq \varepsilon_2 \leq -\alpha_2 - x^*_2\beta_{2x} - w^*\beta_{2w} - \Delta_2\}$ .

*Case of Assumption 5.1:* Supposing  $p_{z'', \varepsilon} > 0$  by Assumption 5.1, which allows  $p_{z'', \varepsilon} \equiv 1$ , and since  $f_\varepsilon(\varepsilon) > 0$  by Assumption 5.2, the first and third terms of the expression derived above sum to a strictly negative number. Supposing  $p_{z'', \varepsilon} \equiv 0$  by Assumption 5.1, the first and fourth terms of the expression derived above sum to a strictly negative number. Since  $p_{z'', \varepsilon} \geq p_{z', \varepsilon}$  by Assumption 5.1, the second term is weakly negative, and since  $p_{z, \varepsilon} \in [0, 1]$ , the third and fourth terms are weakly negative. So this expression is strictly negative.

*Case of  $x''_{11} - x'_{11} > |\Delta_1|$ :* If  $\varepsilon \in \mathcal{R}^-(z'', \theta)$ , then  $-\alpha_1 - x''_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w} \leq \varepsilon_1 \leq -\alpha_1 - x'_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w} - \Delta_1 < -\alpha_1 - x'_{11} - x^*_{1(-1)}\beta_{1x(-1)} - w^*\beta_{1w}$ , where the third inequality uses  $x''_{11} - x'_{11} > |\Delta_1|$ . Therefore,  $\varepsilon \notin \mathcal{R}^-(z', \theta)$ , so  $\mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta) = \emptyset$ . So the second term of the expression derived above is zero. Moreover,  $\mathcal{S}(z', z'', \theta) \cap (\mathcal{R}^-(z'', \theta) \cap \mathcal{R}^-(z', \theta)^C)^C$  has positive Lebesgue measure since  $x''_{11} - x'_{11} > |\Delta_1|$ . So the first term of the expression derived above is strictly negative, since  $f_\varepsilon(\varepsilon) > 0$  by Assumption 5.2. The third and fourth terms are weakly negative. So this expression is strictly negative.

A symmetric result for  $P(y_1 = 1|z)$  and  $x_{21}$  obtains if  $\Delta_1 < 0$ .

If  $\Delta_1 \geq 0$  and  $\Delta_2 \geq 0$ , then symmetric results also obtain by symmetric arguments.

Finally, suppose  $\Delta_2 = 0$ . Then  $P(y_2 = 1|z) = P(\varepsilon_2 \geq -\alpha_2 - x_2\beta_{2x} - w\beta_{2w})$  does not depend on  $x_{11}$ . Similarly, if  $\Delta_1 = 0$ , then  $P(y_1 = 1|z)$  does not depend on  $x_{21}$ .  $\square$

**PROOF OF THEOREM 6.1.** By Powell, Stock, and Stoker (1989),  $\sqrt{M}(\hat{\delta}_M - \delta) \rightarrow^d N(0, V(\delta))$ . So, by the delta method, the asymptotic covariance is  $C(\delta)V(\delta)C(\delta)'$ .  $\square$

**PROOF OF THEOREM 6.2.** By a Taylor series approximation, which exists by Assumption 6.7,  $\widehat{\Delta Q}_M(\gamma) = (\frac{\partial \tilde{R}_{M,1}(a_1, a_2, \dots, a_N, B)}{\partial B} \Big|_{\tilde{\beta}_1} (\beta - \hat{\beta}) \quad \frac{\partial \tilde{R}_{M,2}(b_1, b_2, \dots, b_N, B)}{\partial B} \Big|_{\tilde{\beta}_2} (\beta - \hat{\beta}))'$ , where  $\|\tilde{\beta}_1 - \beta\| \leq \|\hat{\beta} - \beta\|$  and  $\|\tilde{\beta}_2 - \beta\| \leq \|\hat{\beta} - \beta\|$ . So

$$\begin{aligned} \sup_{\gamma} \|\widehat{\Delta Q}_M(\gamma)\| &\lesssim \max \left( \sup_{\gamma, \|\tilde{\beta}_1 - \beta\| \leq \|\hat{\beta} - \beta\|} \left\| \frac{\partial \tilde{R}_{M,1}(a_1, a_2, \dots, a_N, B)}{\partial B} \Big|_{\tilde{\beta}_1} \right\| \|\beta - \hat{\beta}\|, \right. \\ &\quad \left. \sup_{\gamma, \|\tilde{\beta}_2 - \beta\| \leq \|\hat{\beta} - \beta\|} \left\| \frac{\partial \tilde{R}_{M,2}(b_1, b_2, \dots, b_N, B)}{\partial B} \Big|_{\tilde{\beta}_2} \right\| \|\beta - \hat{\beta}\| \right) \\ &= O_p(M^{-1/2}) \end{aligned}$$

by Assumption 6.7, where  $\lesssim$  means less than or equal up to a positive constant. Therefore, by Assumption 6.6,  $\hat{Q}_M(\gamma)$  converges in probability to  $Q(\gamma)$  uniformly over  $\Gamma$ . Consequently, because of Assumptions 3.3 and 6.5 and Theorem 3.1, Newey and McFadden



(1994, Theorem 2.1) applies, so  $\hat{\gamma}_M \rightarrow^p \gamma_0$ . (The notation that  $\gamma_0$  is the true value distinguishes between the use of  $\gamma$  as the argument of a function.)

The asymptotic distribution is derived for the case where it is assumed that  $\alpha_i \equiv \alpha$  and  $\Delta_i \equiv \Delta$  for all agents  $i$ , which implicitly entails deriving the asymptotic distribution without that assumption.

Since  $Q^r(\gamma) = Q(\underbrace{\gamma_1, \gamma_1, \dots, \gamma_1}_{N \text{ times}}, \underbrace{\gamma_2, \gamma_2, \dots, \gamma_2}_{N \text{ times}})$ , where  $\gamma = (a, b)$ ,

$$\frac{\partial \hat{Q}_M^r(\gamma)}{\partial \gamma} = \begin{pmatrix} \sum_i \frac{\partial}{\partial \hat{c}_i} \frac{\partial^N P_M(y = 0_{1 \times N} | \hat{c}_1, \hat{c}_2, \dots, \hat{c}_N)}{\partial \hat{c}_1 \partial \hat{c}_2 \dots \partial \hat{c}_N} \Big|_{a, a, \dots, a} \\ (-1)^N \sum_i \frac{\partial}{\partial \hat{c}_i} \frac{\partial^N P_M(y = 1_{1 \times N} | \hat{c}_1, \hat{c}_2, \dots, \hat{c}_N)}{\partial \hat{c}_1 \partial \hat{c}_2 \dots \partial \hat{c}_N} \Big|_{b, b, \dots, b} \end{pmatrix}$$

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

$$\frac{\partial^2 \hat{Q}_M^r(\gamma)}{\partial^2 a} = \sum_{i,j} \frac{\partial^2}{\partial \hat{c}_i \partial \hat{c}_j} \frac{\partial^N P_M(y = 0_{1 \times N} | \hat{c}_1, \hat{c}_2, \dots, \hat{c}_N)}{\partial \hat{c}_1 \partial \hat{c}_2 \dots \partial \hat{c}_N} \Big|_{a, a, \dots, a}$$

and

$$\frac{\partial^2 \hat{Q}_M^r(\gamma)}{\partial^2 b} = (-1)^N \sum_{i,j} \frac{\partial^2}{\partial \hat{c}_i \partial \hat{c}_j} \frac{\partial^N P_M(y = 1_{1 \times N} | \hat{c}_1, \hat{c}_2, \dots, \hat{c}_N)}{\partial \hat{c}_1 \partial \hat{c}_2 \dots \partial \hat{c}_N} \Big|_{b, b, \dots, b}.$$

By a Taylor series approximation, which exists by Assumption 6.13, and since  $\hat{\gamma}_M$  solves the first order condition by Assumption 6.8, 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  $r_M((-\frac{\partial \hat{Q}_M^r(\gamma)}{\partial \gamma} \Big|_{\gamma_0} + \frac{\partial Q_M^r(\gamma)}{\partial \gamma} \Big|_{\gamma_0}) - \frac{\partial Q_M^r(\gamma)}{\partial \gamma} \Big|_{\gamma_0} = r_M((\frac{\partial^2 \hat{Q}_M^r(\gamma)}{\partial \gamma \partial \gamma'} \Big|_{\tilde{\gamma}_M} - \frac{\partial^2 Q_M^r(\gamma)}{\partial \gamma \partial \gamma'} \Big|_{\tilde{\gamma}_M})(\hat{\gamma}_M - \gamma_0) + \frac{\partial^2 Q_M^r(\gamma)}{\partial \gamma \partial \gamma'} \Big|_{\tilde{\gamma}_M} (\hat{\gamma}_M - \gamma_0)$ . By Assumption 6.11, and arguments similar to the above,  $(-\frac{\partial \hat{Q}_M^r(\gamma)}{\partial \gamma} \Big|_{\gamma_0} + \frac{\partial Q_M^r(\gamma)}{\partial \gamma} \Big|_{\gamma_0}) = O_p(M^{-1/2})$  and  $(\frac{\partial^2 \hat{Q}_M^r(\gamma)}{\partial \gamma \partial \gamma'} \Big|_{\tilde{\gamma}_M} - \frac{\partial^2 Q_M^r(\gamma)}{\partial \gamma \partial \gamma'} \Big|_{\tilde{\gamma}_M}) = O_p(M^{-1/2})$ .

By Assumption 6.10,  $\frac{\partial^2 \hat{Q}_M^r(\gamma)}{\partial \gamma \partial \gamma'} \Big|_{\tilde{\gamma}_M} \rightarrow^p \frac{\partial^2 Q^r(\gamma)}{\partial \gamma \partial \gamma'} \Big|_{\gamma_0}$ , where  $\frac{\partial^2 Q^r(\tau)}{\partial \gamma \partial \gamma'} = \begin{pmatrix} \frac{\partial^2 Q^r(\tau)}{\partial^2 a} & 0 \\ 0 & \frac{\partial^2 Q^r(\tau)}{\partial^2 b} \end{pmatrix}$ , where

$$\frac{\partial^2 Q^r(\gamma)}{\partial^2 a} = \sum_{i,j} \frac{\partial^2}{\partial c_i \partial c_j} \frac{\partial^N P(y = 0_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{a, a, \dots, a}$$

and

$$\frac{\partial^2 Q^r(\gamma)}{\partial^2 b} = (-1)^N \sum_{i,j} \frac{\partial^2}{\partial c_i \partial c_j} \frac{\partial^N P(y = 1_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{b, b, \dots, b}.$$

The Hessian of the density of  $\varepsilon + \alpha$  evaluated at  $\alpha$  is the  $N \times N$  matrix with  $(i, j)$  element equal to  $\frac{\partial^2}{\partial c_i \partial c_j} \frac{\partial^N P(y=0_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{\alpha, \alpha, \dots, \alpha}$ . Therefore, by multiplying the Hessian by the vector of 1s, Assumption 6.12 implies that

$$\sum_{i,j} \frac{\partial^2}{\partial c_i \partial c_j} \frac{\partial^N P(y=0_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{\alpha, \alpha, \dots, \alpha} < 0.$$

By similar arguments based on the Hessian of the density of  $\varepsilon + \alpha + N_0 \Delta$  evaluated at  $\alpha + N_0 \Delta$ , it follows that

$$(-1)^N \sum_{i,j} \frac{\partial^2}{\partial c_i \partial c_j} \frac{\partial^N P(y=1_{1 \times N} | c_1, c_2, \dots, c_N)}{\partial c_1 \partial c_2 \dots \partial c_N} \Big|_{\alpha + N_0 \Delta, \alpha + N_0 \Delta, \dots, \alpha + N_0 \Delta} < 0.$$

Therefore, with probability approaching 1,  $\frac{\partial^2 Q'_M(\gamma)}{\partial \gamma \partial \gamma'} \Big|_{\hat{\gamma}_M}$  is invertible.

Thus,  $r_M(\hat{\gamma}_M - \gamma_0) \rightarrow^d N(0, (\frac{\partial^2 Q'_M(\gamma)}{\partial \gamma \partial \gamma'} \Big|_{\gamma_0})^{-1} D \Omega_0 D' (\frac{\partial^2 Q'_M(\gamma)}{\partial \gamma \partial \gamma'} \Big|_{\gamma_0})^{-1})$  by Assumption 6.9, where  $D = \begin{pmatrix} 1_{1 \times N} & 0 \\ 0 & 1_{1 \times N} \end{pmatrix}$ . And by the delta method,  $r_M(\hat{\psi}_M - \psi) \rightarrow^d N(0, C (\frac{\partial^2 Q'_M(\gamma)}{\partial \gamma \partial \gamma'} \Big|_{\gamma_0})^{-1} \times D \Omega_0 D' (\frac{\partial^2 Q'_M(\gamma)}{\partial \gamma \partial \gamma'} \Big|_{\gamma_0})^{-1} C')$ , where  $C = \begin{pmatrix} 1 & 0 \\ \frac{1}{N_0} & \frac{1}{N_0} \end{pmatrix}$ .  $\square$

#### REFERENCES

- Abrevaya, J., J. A. Hausman, and S. Khan (2010), "Testing for causal effects in a generalized regression model with endogenous regressors." *Econometrica*, 78 (6), 2043–2061. [59]
- Andrews, D. W. K. (1995), "Nonparametric kernel estimation for semiparametric models." *Econometric Theory*, 11 (3), 560–596. [67]
- Andrews, D. W. K. and M. M. Schafgans (1998), "Semiparametric estimation of the intercept of a sample selection model." *Review of Economic Studies*, 65 (3), 497–517. [42]
- Andrews, D. W. K. and G. Soares (2010), "Inference for parameters defined by moment inequalities using generalized moment selection." *Econometrica*, 78 (1), 119–157. [60]
- Aradillas-Lopez, A. and A. Gandhi (2014), "Estimation of games with ordered actions: An application to chain-store entry." Working paper. [51, 54, 59]
- Aradillas-Lopez, A. and A. M. Rosen (2013), "Inference in ordered response games with complete information." Working paper. [38, 48, 54]
- Aradillas-Lopez, A. and E. Tamer (2008), "The identification power of equilibrium in simple games." *Journal of Business & Economic Statistics*, 26 (3), 261–283. [48]
- Bajari, P., J. Hahn, H. Hong, and G. Ridder (2011), "A note on semiparametric estimation of finite mixtures of discrete choice models with application to game theoretic models." *International Economic Review*, 52 (3), 807–824. [58]
- Bajari, P., H. Hong, and S. P. Ryan (2010), "Identification and estimation of a discrete game of complete information." *Econometrica*, 78 (5), 1529–1568. [37]

- Beresteanu, A., I. Molchanov, and F. Molinari (2011), “Sharp identification regions in models with convex moment predictions.” *Econometrica*, 79 (6), 1785–1821. [48]
- Bernheim, B. D. (1984), “Rationalizable strategic behavior.” *Econometrica*, 52 (4), 1007–1028. [72]
- Berry, S. T. (1992), “Estimation of a model of entry in the airline industry.” *Econometrica*, 60 (4), 889–917. [70]
- Bierens, H. J. (1987), “Kernel estimators of regression functions.” In *Advances in Econometrics: Fifth World Congress*, Vol. I (T. F. Bewley, ed.), 99–144, Cambridge University Press, Cambridge. [67]
- Bresnahan, T. F. and P. C. Reiss (1990), “Entry in monopoly market.” *Review of Economic Studies*, 57 (4), 531–553. [37, 41]
- Bresnahan, T. F. and P. C. Reiss (1991a), “Empirical models of discrete games.” *Journal of Econometrics*, 48 (1), 57–81. [37, 41]
- Bresnahan, T. F. and P. C. Reiss (1991b), “Entry and competition in concentrated markets.” *Journal of Political Economy*, 99 (5), 977–1009. [37]
- Bulow, J. I., J. D. Geanakoplos, and P. D. Klemperer (1985), “Multimarket oligopoly: Strategic substitutes and complements.” *Journal of Political Economy*, 93 (3), 488–511. [40]
- Chamberlain, G. (1986), “Asymptotic efficiency in semi-parametric models with censoring.” *Journal of Econometrics*, 32 (2), 189–218. [67]
- Ciliberto, F. and E. Tamer (2009), “Market structure and multiple equilibria in airline markets.” *Econometrica*, 77 (6), 1791–1828. [41, 48, 70]
- Cosslett, S. R. (1983), “Distribution-free maximum likelihood estimator of the binary choice model.” *Econometrica*, 51 (3), 765–782. [44]
- Cosslett, S. R. (1987), “Efficiency bounds for distribution-free estimators of the binary choice and the censored regression models.” *Econometrica*, 55 (3), 559–585. [38, 67]
- Dudley, R. M. (1999), *Uniform Central Limit Theorems*. Cambridge Studies in Advanced Mathematics, Vol. 63. Cambridge University Press, Cambridge. [45]
- Dunker, F., S. Hoderlein, and H. Kaido (2013), “Random coefficients in static games of complete information.” Working paper. [37]
- Fox, J. T. and N. Lazzati (2013), “Identification of discrete choice models for bundles and binary games.” Working paper. [37]
- Fox, J. T. and N. Lazzati (2014), “Identification of discrete choice models for bundles and binary games.” Working paper. [38]
- Galichon, A. and M. Henry (2011), “Set identification in models with multiple equilibria.” *Review of Economic Studies*, 78 (4), 1264–1298. [48]

Horowitz, J. L. (2009), *Semiparametric and Nonparametric Methods in Econometrics*. Springer Series in Statistics. Springer, New York. [45, 61, 65]

Khan, S. and D. Nekipelov (2012), “Information structure and statistical information in discrete response models.” Working paper. [67]

Khan, S. and E. Tamer (2010), “Irregular identification, support conditions, and inverse weight estimation.” *Econometrica*, 78 (6), 2021–2042. [44, 48]

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. [60]

Kline, B. (2015a), “Identification of complete information games.” Working paper. [37, 44, 51, 72]

Kline, B. (2015b), “Identification of the direction of a causal effect by instrumental variables.” Working paper. [59]

Kline, B. and E. Tamer (2012), “Bounds for best response functions in binary games.” *Journal of Econometrics*, 166 (1), 92–105. [48]

Kline, B. and E. Tamer (2015), “Bayesian inference in a class of partially identified models.” Working paper. [48, 70]

Lee, M.-j. (1989), “Mode regression.” *Journal of Econometrics*, 42 (3), 337–349. [43]

Lewbel, A. (2000), “Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables.” *Journal of Econometrics*, 97 (1), 145–177. [44]

Magnac, T. and E. Maurin (2007), “Identification and information in monotone binary models.” *Journal of Econometrics*, 139 (1), 76–104. [44]

Manski, C. F. (1988), “Identification of binary response models.” *Journal of the American Statistical Association*, 83 (403), 729–738. [44]

Newey, W. K. and D. McFadden (1994), “Large sample estimation and hypothesis testing.” In *Handbook of Econometrics*, Vol. IV (R. F. Engle and D. L. McFadden, eds.), 2111–2245, Chapter 36, Elsevier Science B.V., Amsterdam. [76, 77]

Pagan, A. and A. Ullah (1999), *Nonparametric Econometrics*. Cambridge University Press. [67, 69]

de Paula, Á. and X. Tang (2012), “Inference of signs of interaction effects in simultaneous games with incomplete information.” *Econometrica*, 80 (1), 143–172. [59]

Pearce, D. G. (1984), “Rationalizable strategic behavior and the problem of perfection.” *Econometrica*, 52 (4), 1029–1050. [72]

Powell, J. L., J. H. Stock, and T. M. Stoker (1989), “Semiparametric estimation of index coefficients.” *Econometrica*, 57 (6), 1403–1430. [60, 61, 76]

Powell, J. L. and T. M. Stoker (1996), “Optimal bandwidth choice for density-weighted averages.” *Journal of Econometrics*, 75 (2), 291–316. [69]

Samarov, A. M. (1993), “Exploring regression structure using nonparametric functional estimation.” *Journal of the American Statistical Association*, 88 (423), 836–847. [60]

Tamer, E. (2003), “Incomplete simultaneous discrete response model with multiple equilibria.” *Review of Economic Studies*, 70 (1), 147–165. [37]

Vives, X. (2005), “Games with strategic complementarities: New applications to industrial organization.” *International Journal of Industrial Organization*, 23 (7), 625–637. [40]

Zhou, Y. (2013), “Identification and estimation of entry games under symmetry of unobservables.” Working paper. [38]

---

Co-editor Rosa L. Matzkin handled this manuscript.

Submitted March, 2014. Final version accepted June, 2015.