Estimation and inference in discrete games with uncertain behavior

Andres Aradillas-Lopez*

June 6, 2025

Abstract

We revisit the econometric analysis of normal-form discrete games where, instead of assuming that all observations come from the same behavioral model (e.g, Nash equilibrium), we pre-specify a collection of *candidate* behavioral models (each one with potential multiple solutions) that could have generated each observation. Assuming the existence of an observable *instrument* Z that controls for the dependence between the behavior/solution selection mechanisms and payoff covariates, our global model becomes a convolution of the candidate behavioral models, with the convolution weights being nonparametric functionals of Z. We show conditions under which the parameters of the model are point-identified, and we propose conditional-GMM estimation and inference procedures. We evaluate the performance of our conditional-GMM estimator in Monte Carlo experiments. As an empirical illustration, we analyze geographic entry decisions by the two dominant firms in the home improvement retail industry in the United States (Lowe's and Home Depot). Assuming coalitional cooperation and Nash equilibrium as the two candidate behavioral models, we find evidence that both types of behavior are present across markets, and that the probability of cooperation decreases with market size.

Keywords: Estimation of games, uncertain behavior, multiple equilibria, semiparametric convolution.
JEL classification: C1, C14, C57.

1 Introduction

One of the main goals in the econometric analysis of games is the estimation of players' payoff functions. This requires assuming a behavioral theory (e.g, Nash equilibrium)

*Department of Economics, Pennsylvania State University, University Park, PA 16802, United States. Email: aaradill@psu.edu

that links the outcomes observed to the underlying game. Thus, the behavioral model assumed is key to the validity of the analysis, and having inferential methods that provide some degree of behavioral robustness is a goal we should pursue. Allowing for behavioral heterogeneity can be facilitated in experimental settings where the researcher has precise knowledge about payoff functions (see, e.g, Crawford, Costa-Gomes, and Iriberri (2013)), but it can be significantly more challenging with nonexperimental data where payoff functions themselves are unknown and must be estimated. As a result, the econometric analysis of games in nonexperimental settings typically relies on the assumption that every observation in the data was produced by the same behavioral model, which is assumed to be known (e.g, Nash equilibrium). This paper focuses on static discrete games where the parameters of players' normal-form payoff functions are the unknown object of interest. Instead of assuming one behavioral model, we propose a "global" model where we pre-specify a collection of *candidate* behavioral models, each one with potentially multiple solutions, that can produce each observation in the data. Unobservable behavior and solution-selection mechanisms determine which behavioral model and which corresponding solution are selected.

Assuming the existence of an observable *instrument* Z that controls for the dependence between the selection mechanisms and payoff covariates, the predictions of our global model can be written as a convolution of the predictions of the candidate solutions, with the convolution weights being nonparametric functionals of Z. If our candidate solutions are not observationally equivalent, the weights of the convolution can be semiparametrically identified and point-identification of the payoff parameters can be obtained if the latent selection mechanisms assign nonzero probability to behavioral models that have identification power. In the absence of such restrictions, a CS for the parameters can be constructed from our semiparametric convolution. In either case (estimation and inference), we propose conditional GMM procedures based on the semiparametric convolution produced by our global model. Our setup allows inference of the normal-form parameters as well as the convolution weights, which contain key information about the propensity of the selection mechanisms to select each behavioral model, and each solution within each behavioral model. While our framework depends on the availability of an instrument Z that satisfies key exclusion restrictions, the validity of any proposed instrument Z can be formally evaluated through consistent specification tests.

The paper proceeds as follows. Section 2 introduces our approach through a 2×2 game with two candidate behavioral models: complete-information pure-strategy Nash equilibrium and (coalitional) cooperation. Section 3 expands the model in Section 2 by allowing for mixed-strategies and by introducing incomplete-information Bayesian Nash equilibrium as a third candidate behavioral model. Section 4 goes beyond the 2×2 game and describes a general discrete game with $P \geq 2$ players and a collection of candidate behavioral models, each with potentially multiple solutions. In each section we discuss identification, estimation and inference for the normal-form parameters and for the properties of the underlying behavioral and solution-selection mechanisms. Section 5 conducts Monte Carlo studies of the properties of our proposed conditional-GMM estimation approach to recover payoff parameters and the identifiable properties of the underlying selection mechanisms. As an empirical illustration in Section 6, we apply our methodology to analyze geographic-market entry decisions by Lowe's and Home Depot using cooperation and Nash equilibrium as candidate behavioral models in each market. Our findings are consistent with a mixture of both behavioral models where the probability of cooperation decreases with market size. Section 7 concludes. Appendix A includes details of our main econometric results, and an Empirical Supplement includes additional results for our Monte Carlo experiments and our empirical illustration. An online Econometric Supplement with step-by-step econometric derivations of our results can be found at https://aaradill.github.io/econometric_supplement_uncertain_behavior.pdf

2 A 2×2 game

Consider the following simultaneous 2×2 normal-form game,

Table 1S-. A game of strategic substitutes

	$Y_2 = 1$	$Y_2 = 0$
$Y_1 = 1$	$t_1 - \Delta_1 - \varepsilon_1$, $t_2 - \Delta_2 - \varepsilon_2$	$t_1 - \varepsilon_1$, 0
$Y_1 = 0$	0 , $t_2 - \varepsilon_2$	0 , 0

Consider the case of strategic substitutes ($\Delta_p \geq 0$ for $p = 1, 2$) and complete-information, where both players observe $(t_p, \Delta_p, \varepsilon_p)_{p=1}^{2}$ before making their choices. We will expand our model to include mixed strategies, additional candidate behavioral models, more players and a richer action space in subsequent sections. Let $\mathcal{Y} \equiv \{0,1\} \times \{0,1\}$ denote the action space, with $y \equiv (y_1, y_2) \in \mathcal{Y}$ denoting a generic element of $\mathcal{Y}$, and $Y \equiv (Y_1, Y_2)$ denoting the actual outcome of the game. Suppose we consider two candidate behavioral models,

(i) **Noncooperative:** Players play a pure-strategy Nash equilibrium (PSNE).

(i) **Cooperative:** Players engage in cooperative (coalitional) bargaining and select the outcome $Y \equiv (Y_1, Y_2)$ that maximizes the sum of their payoffs (i.e, the total value of the coalition).

With transferable utility, maximization of joint payoffs is the prediction of Nash bargaining (see Myerson (1990), Thomson (1994), or Watson (2013, Appendix D)). Figure 1 fixes

$(t_p, \Delta_p)_{p=1}^2$ and splits $\mathbb{R}^2$ into regions of realizations of $(\varepsilon_1, \varepsilon_2)$ corresponding to each outcome $Y \in \mathcal{Y}$ that constitute a solution to each of our candidate behavioral models. Cooperation always yields a unique prediction for Y, while PSNE produces multiple predictions in some regions, which is a well-known result (see Bjorn and Vuong (1984), Bresnahan and Reiss (1990), Bresnahan and Reiss (1991b), Bresnahan and Reiss (1991a), Tamer (2003), Berry and Tamer (2007)). However, both models produce a unique prediction for $S \equiv Y_1 + Y_2$, as shown in Figure 2, so we will focus on this aggregate outcome.

2.1 A parameterization of normal-form payoff functions

We will assume that $\varepsilon \equiv (\varepsilon_1, \varepsilon_2)$ are unobserved by the econometrician, and we will model the remaining payoff shifters $(t_p, \Delta_p)_{p=1}^2$ as follows. Let $(X_1^{ns}, X_2^{ns}, X_1^s, X_2^s)$ be a collection of covariates that are observable to the econometrician. X_p^{ns} includes covariates that shift[1] player p's non-strategic payoff component t_p, while X_p^s denotes covariates that shift player p's strategic-interaction payoff component Δ_p. Let $(\beta_{10}, \beta_{20}, \Delta_{10}, \Delta_{20})$ be unknown parameters. We will parameterize $t_p = X_p^{ns\prime}\beta_{p0}$ and $\Delta_p \equiv X_p^{s\prime}\Delta_{p0}$ for $p = 1, 2$, so the normal-form representation of the game is given by,

Table 2S-. A parameterized game of strategic substitutes

	$Y_2 = 1$	$Y_2 = 0$
$Y_1 = 1$	$X_1^{ns\prime}\beta_{10} - X_1^{s\prime}\Delta_{10} - \varepsilon_1$, $X_2^{ns\prime}\beta_{20} - X_2^{s\prime}\Delta_{20} - \varepsilon_2$	$X_1^{ns\prime}\beta_{10} - \varepsilon_1$, 0
$Y_1 = 0$	0 , $X_2^{ns\prime}\beta_{20} - \varepsilon_2$	0 , 0

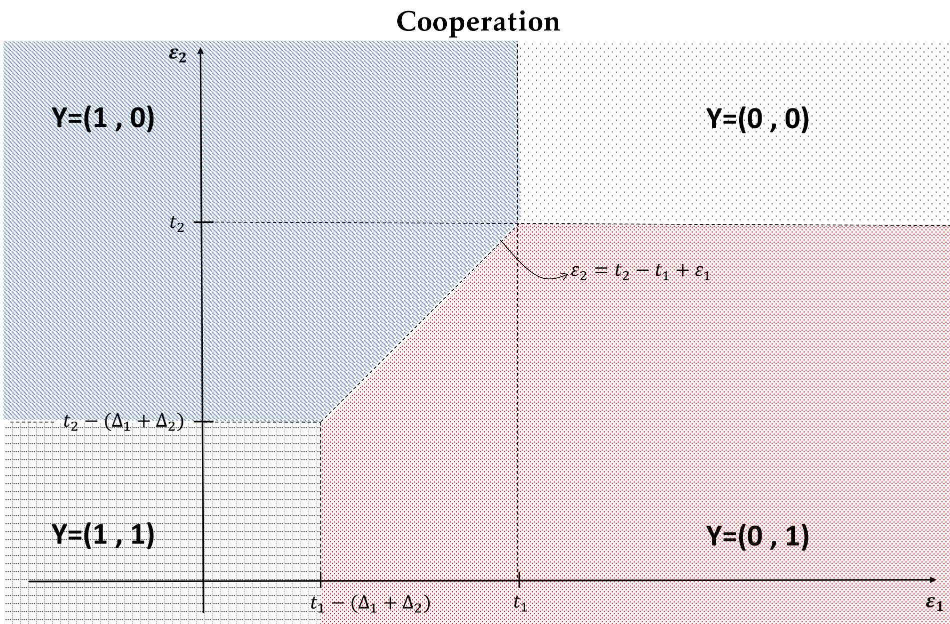
The values of $(\Delta_{10}, \Delta_{20})$, and the support of (X_1^s, X_2^s) are such that $X_p^{s\prime}\Delta_{p0} \geq 0$ w.p.1, so we have a strategic-substitutes game. Constant strategic interaction effects is a special case where $X_p^s \equiv 1$ and $X_p^{s\prime}\Delta_{p0} = \Delta_{p0}$. Nonstrategic X_p^{ns} and strategic X_p^s payoff shifters can have elements in common for each player and across players as long as the conditions we will describe below are satisfied. We will group $X \equiv (X_1^{ns} \cup X_2^{ns} \cup X_1^s \cup X_2^s)$ and we will maintain that players have complete information, so they observe the realization of (X, ε) and they know the true values of the payoff parameters.

2.2 A parameterization for the joint distribution of ε

We will consider a setting where the econometrician observes a sample of outcomes and payoff covariates from this game, but does not know which of the two candidate behavioral models produced each observation. Our approach will ultimately rely on a particular convolution of the parametric predictions of each candidate behavioral model. To obtain these predictions we will parameterize the joint distribution of the unobserved payoff shifters ε.

[1] We will use the terms "payoff covariates" and "payoff shifters" interchangeably.

Figure 1: Predicted outcomes for Y under cooperation and PSNE



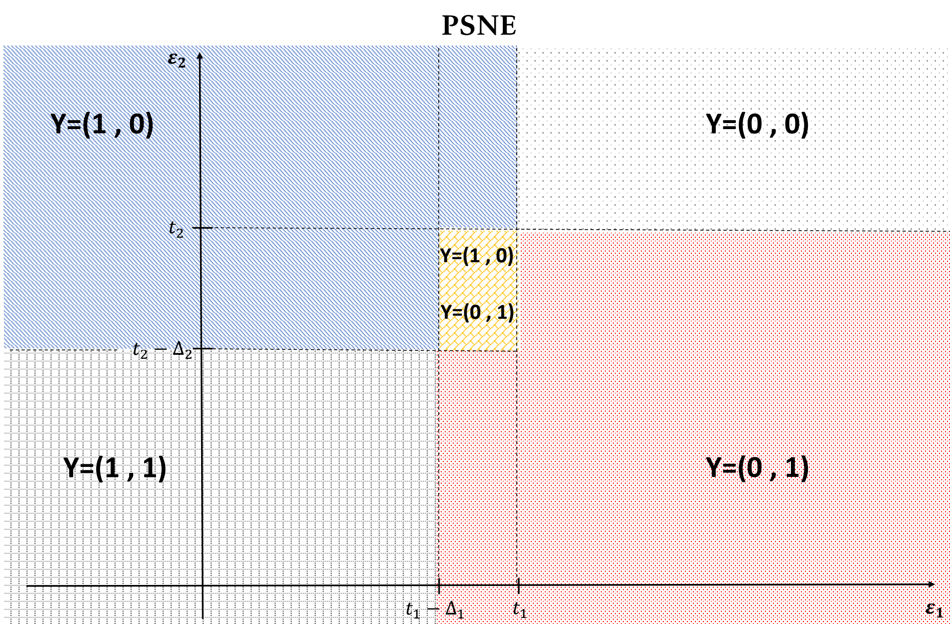
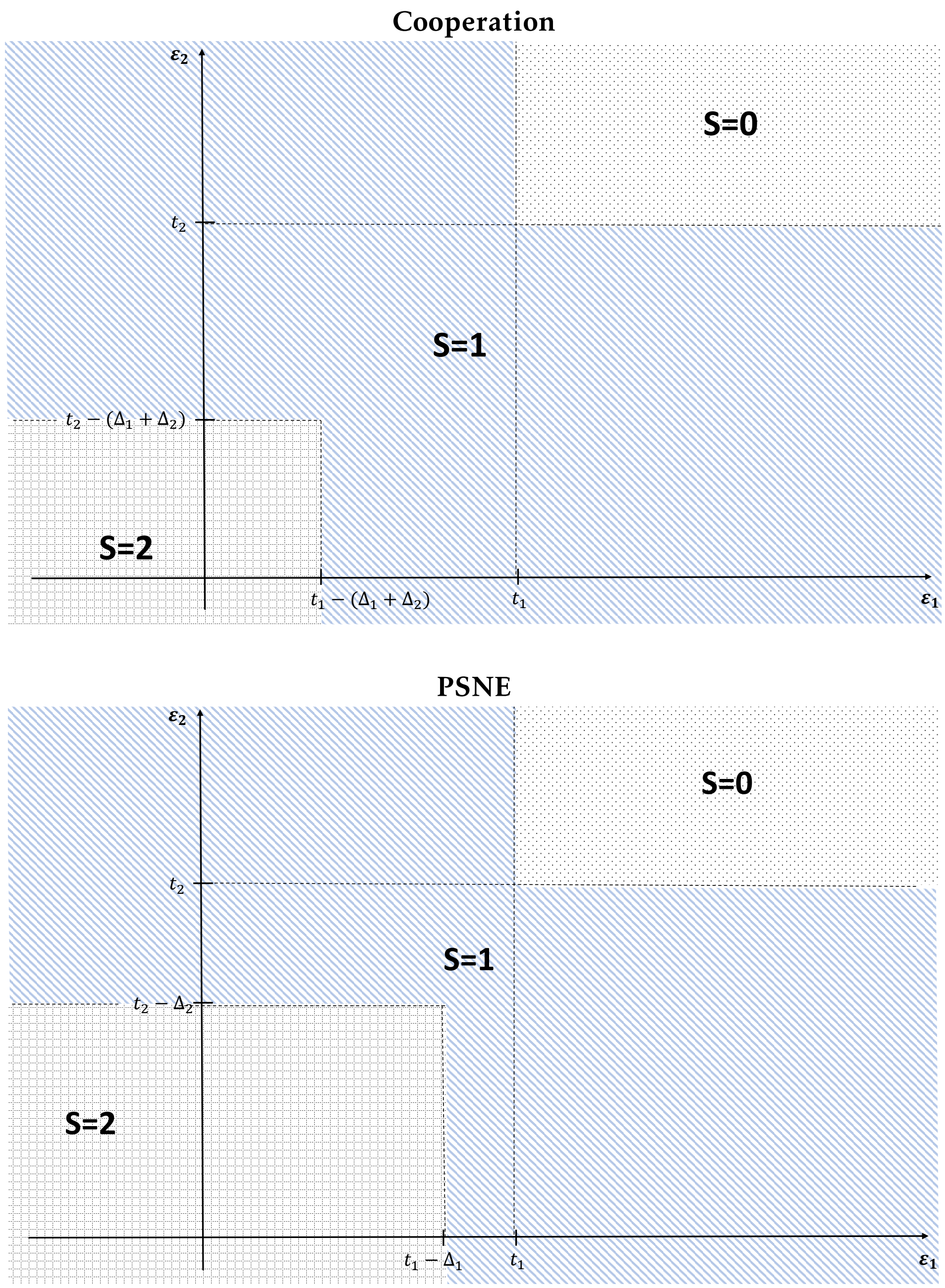


Figure 2: Predicted outcomes for $S \equiv Y_1 + Y_2$ under cooperation and PSNE



Assumption G1 ***(A parametric distribution for ε)***

(i) $\varepsilon \sim F_{1,2}(\cdot,\cdot|\rho_0)$, a parametric, jointly continuous distribution with unbounded support on $\mathbb{R}^2$ indexed by a scalar parameter ρ_0 that captures the interdependence between ε_1 and ε_2. We assume that ε are independent of X, although this can be relaxed if we parameterize the conditional distribution of $\varepsilon|X$. We denote the marginal distributions as $\varepsilon_p \sim F_p(\cdot)$.

(ii) Group the non-strategic parameters in our model as $\gamma_0 \equiv (\beta_0, \rho_0)$, with $\beta_0 \equiv (\beta_{10}, \beta_{20})$, and the strategic-interaction parameters as $\Delta_0 \equiv (\Delta_{10}, \Delta_{20})$. We will denote $\theta_0 \equiv (\gamma_0, \Delta_0) \in \Theta$, where the parameter space Θ is assumed to be compact. We will denote a generic element in Θ as $\theta \equiv (\gamma, \Delta)$, with $\gamma \equiv (\beta_1, \beta_2, \rho)$ and $\Delta \equiv (\Delta_1, \Delta_2)$.

(iii) Let $\nabla_p F_{1,2}(\epsilon_1, \epsilon_2|\rho) \equiv \frac{\partial F_{1,2}(\epsilon_1,\epsilon_2|\rho)}{\partial \epsilon_p}$. Then, $\left|\nabla_p F_{1,2}(\epsilon_1, \epsilon_2|\rho)\right| \leq \overline{D} < \infty \ \forall \ (\epsilon_1, \epsilon_2) \in \mathbb{R}^2$, $\rho \in \Theta$. ■

Since both of our candidate behavioral models are characterized by threshold-crossing decision rules, the scale of ε_p cannot be separately identified from the scale of the payoff parameters. For this reason, the variance of ε_p is fixed in $F_p(\cdot)$. In our strategic-substitutes game, we will restrict Θ to satisfy $X_1^{s\prime}\Delta_1 \geq 0$ and $X_2^{s\prime}\Delta_2 \geq 0$ w.p.1 $\forall \ \Delta \in \Theta$.

2.3 An expression for $E[S|X]$ in each behavioral model

For a given (θ, X), the regions of $\varepsilon \in \mathbb{R}^2$ that predict each possible value of $S \in \{0,1,2\}$ for our candidate behavioral models are (see Figure 2),

$$
\begin{aligned}
&\textbf{Cooperative-behavior regions:}\\
&\mathcal{R}_S^C(0|X,\beta) \equiv \Big\{(\varepsilon_1,\varepsilon_2) \in \mathbb{R}^2 : \varepsilon_1 \geq X_1^{ns\prime}\beta_1 \ , \varepsilon_2 \geq X_2^{ns\prime}\beta_2\Big\},\\
&\mathcal{R}_S^C(2|X,\beta,\Delta) \equiv \Big\{(\varepsilon_1,\varepsilon_2) \in \mathbb{R}^2 : \varepsilon_1 \leq X_1^{ns\prime}\beta_1 - \Big(X_1^{s\prime}\Delta_1 + X_2^{s\prime}\Delta_2\Big) \ , \varepsilon_2 \leq X_2^{ns\prime}\beta_2 - \Big(X_1^{s\prime}\Delta_1 + X_2^{s\prime}\Delta_2\Big)\Big\},\\
&\mathcal{R}_S^C(1|X,\beta,\Delta) \equiv \mathbb{R}^2 \setminus \Big(\mathcal{R}_S^C(0|X,\beta) \cup \mathcal{R}_S^C(2|X,\beta,\Delta)\Big).\\
&\textbf{Noncooperative-behavior regions:}\\
&\mathcal{R}_S^{NC}(0|X,\beta) \equiv \Big\{(\varepsilon_1,\varepsilon_2) \in \mathbb{R}^2 : \varepsilon_1 \geq X_1^{ns\prime}\beta_1 \ , \varepsilon_2 \geq X_2^{ns\prime}\beta_2\Big\},\\
&\mathcal{R}_S^{NC}(2|X,\beta,\Delta) \equiv \Big\{(\varepsilon_1,\varepsilon_2) \in \mathbb{R}^2 : \varepsilon_1 \leq X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1 \ , \varepsilon_2 \leq X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2\Big\},\\
&\mathcal{R}_S^{NC}(1|X,\beta,\Delta) \equiv \mathbb{R}^2 \setminus \Big(\mathcal{R}_S^{NC}(0|X,\beta) \cup \mathcal{R}_S^{NC}(2|X,\beta,\Delta)\Big).
\end{aligned}
\tag{1}
$$

Note that $\mathcal{R}_S^C(0|X,\beta) = \mathcal{R}_S^{NC}(0|X,\beta)$ for any β. Thus,

$$
\begin{aligned}
&\textbf{If players' behavior is cooperative:}\\
&S = 0\cdot\mathbb{1}\Big\{\varepsilon\in\mathcal{R}_S^C(0|X,\beta_0)\Big\} + 1\cdot\mathbb{1}\Big\{\varepsilon\in\mathcal{R}_S^C(1|X,\beta_0,\Delta_0)\Big\} + 2\cdot\mathbb{1}\Big\{\varepsilon\in\mathcal{R}_S^C(2|X,\beta_0,\Delta_0)\Big\}.\\
&\textbf{If players' behavior is noncooperative:}\\
&S = 0\cdot\mathbb{1}\Big\{\varepsilon\in\mathcal{R}_S^{NC}(0|X,\beta_0)\Big\} + 1\cdot\mathbb{1}\Big\{\varepsilon\in\mathcal{R}_S^{NC}(1|X,\beta_0,\Delta_0)\Big\} + 2\cdot\mathbb{1}\Big\{\varepsilon\in\mathcal{R}_S^{NC}(2|X,\beta_0,\Delta_0)\Big\}.
\end{aligned}
\tag{2}
$$

From here, let

$$
\begin{aligned}
\mu_S^C(X,\theta) &\equiv F_1(X_1^{ns\prime}\beta_1) + F_2(X_2^{ns\prime}\beta_2) - F_{1,2}(X_1^{ns\prime}\beta_1\ ,\ X_2^{ns\prime}\beta_2|\rho)\\
&\quad + F_{1,2}(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1 - X_2^{s\prime}\Delta_2\ ,\ X_2^{ns\prime}\beta_2 - X_1^{s\prime}\Delta_1 - X_2^{s\prime}\Delta_2|\rho),\\
\mu_S^{NC}(X,\theta) &\equiv F_1(X_1^{ns\prime}\beta_1) + F_2(X_2^{ns\prime}\beta_2) - F_{1,2}(X_1^{ns\prime}\beta_1\ ,\ X_2^{ns\prime}\beta_2|\rho)\\
&\quad + F_{1,2}(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1\ ,\ X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2|\rho).
\end{aligned}
\tag{3}
$$

Combining (2) and (3), we have $E[S|X] = \mu_S^C(X,\theta_0)$ if players' behavior is cooperative, and $E[S|X] = \mu_S^{NC}(X,\theta_0)$ if players' behavior is noncooperative.

2.4 A global behavioral model that nests cooperative and noncooperative behavior

We will embed both of the candidate behavioral models into a *global* model, where the realization of a behavioral selection mechanism ξ determines the behavioral model selected. We denote $\xi = 1$ if players' behavior is cooperative, and $\xi = 2$ if players' behavior is PSNE. The realization of ξ (and therefore the underlying behavioral model) is unobserved by the econometrician, but it is observed by both players prior to playing the game. Our global model allows for players to always cooperate or always play PSNE, but it accommodates the case where players cooperate in some realizations of the game and follow non-cooperative behavior in others. The precise way through which behavior is selected (e.g, whether players observe a signal from "nature" prior to deciding whether to cooperate, etc.) will be irrelevant as long as the mechanism ξ satisfies the exclusion restrictions in Assumption G2. From (2), in our global behavioral model we have,

$$
\begin{aligned}
S =& \mathbb{1}\{\xi=1\}\cdot\Big[\mathbb{1}\Big\{\varepsilon\in\mathcal{R}_S^C(1|X,\beta_0,\Delta_0)\Big\} + 2\cdot\mathbb{1}\Big\{\varepsilon\in\mathcal{R}_S^C(2|X,\beta_0,\Delta_0)\Big\}\Big]\\
&+\mathbb{1}\{\xi=2\}\cdot\Big[\mathbb{1}\Big\{\varepsilon\in\mathcal{R}_S^{NC}(1|X,\beta_0,\Delta_0)\Big\} + 2\cdot\mathbb{1}\Big\{\varepsilon\in\mathcal{R}_S^{NC}(2|X,\beta_0,\Delta_0)\Big\}\Big]
\end{aligned}
\tag{4}
$$

2.4.1 Introducing an instrument Z for the behavioral selection mechanism ξ

We will assume the existence of an observable Z that controls for the dependence between the behavioral mechanism ξ and players' payoff covariates.

Assumption G2 ***(An exclusion restriction for the behavioral selection mechanism)*** *$\xi \perp \varepsilon$, and there exists an observable Z such that $\xi|(X,Z) \sim \xi|Z$.* ■

We can weaken $\xi \perp \varepsilon$ to the *conditional* independence restriction $\xi \perp \varepsilon|(X,Z)$ if we assume a parameterization $F_{\varepsilon|X,Z}(\cdot,\cdot|X,Z,\rho_0)$ for the conditional distribution of $\varepsilon|(X,Z)$. Z can have elements in common with X, with $Z \subset X$ as a possibility. However, as we will make precise below, we rule out the case where $Z = X$. From now on, we will group $U \equiv X \cup Z$. Let $\pi(Z) \equiv Pr(\xi = 1|Z)$ (the probability of cooperation conditional on Z), which we will leave nonparametrically specified. By Assumption G2, $E[S|U] = \pi(Z) \cdot \mu_S^C(X,\theta_0) + (1-\pi(Z)) \cdot \mu_S^{NC}(X,\theta_0)$ in our global behavioral model. Thus, the predictions of our global model are a *convolution* of the parametric predictions of each candidate behavioral model, and the convolution weights are nonparametric functionals of Z. Let

$$\begin{aligned} m_S^C(X,\theta) &\equiv F_{1,2}(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1 - X_2^{s\prime}\Delta_2 \ , X_2^{ns\prime}\beta_2 - X_1^{s\prime}\Delta_1 - X_2^{s\prime}\Delta_2|\rho), \\ m_S^{NC}(X,\theta) &\equiv F_{1,2}(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1 \ , X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2|\rho), \\ \Xi_S(X,\theta) &\equiv m_S^C(X,\theta) - m_S^{NC}(X,\theta). \end{aligned} \tag{5}$$

From (3), our expression for $E[S|U]$ simplifies to,

$$E[S|U] = m_S^{NC}(X,\theta_0) + \pi(Z) \cdot \Xi_S(X,\theta_0). \tag{6}$$

Our goal is to do inference on θ_0 using the semiparametric convolution (6) when the econometrician observes[2] a random sample $(Y_i, X_i, Z_i)_{i=1}^n$ produced by our global model.

2.5 Identifiability of θ_0 in our global behavioral model

Group $m_S(X,\theta) \equiv (m_S^C(X,\theta) \ , m_S^{NC}(X,\theta))$. From (6), our model predicts the exclusion restriction $E[S|U] = E[S|m_S(X,\theta_0), Z]$. We could pursue inference based on this restriction, but (6) provides additional structure, which can facilitate identification and mitigate the curse of dimensionality vis-a-vis a method based solely on $E[S|U] = E[S|m_S(X,\theta_0), Z]$. To exploit the semiparametric convolution in (6), we will first construct an estimator for $\pi(Z)$. Identification of the behavioral weight $\pi(Z)$ will require ruling out that both candidate behavioral models are observationally equivalent. To this end, consider the following restriction.

[2] We only need to observe $(S_i, X_i, Z_i)_{i=1}^n$.

Assumption G3 ***(Presence of strategic-interaction effects)*** *For almost every (a.e) realization of Z, we have $Pr(X_1^{s\prime}\Delta_{10} + X_2^{s\prime}\Delta_{20} > 0 \mid Z) > 0$.* ■

Accordingly, from now on our parameter space Θ will be assumed to satisfy,

$$\left.\begin{array}{l} Pr(X_1^{s\prime}\Delta_1 \geq 0 \;\text{ and }\; X_2^{s\prime}\Delta_2 \geq 0) = 1 \\ Pr(X_1^{s\prime}\Delta_1 + X_2^{s\prime}\Delta_2 > 0 \mid Z) > 0, \;\text{ a.e } Z \end{array}\right\} \;\forall\; \Delta \in \Theta. \tag{7}$$

With constant strategic-interaction effects, (7) would be satisfied if $\Delta_p \geq \underline{c} > 0$ for $p = 1, 2$, $\forall\; \Delta \in \Theta$. Without Assumption G3, both candidate behavioral models could be observationally equivalent, making it impossible to identify and estimate $\pi(Z)$.

Notation: In what follows, we will often denote the dimension of a vector-valued object ξ as $\mathbb{R}^{d_\xi}$.

2.5.1 Identification and estimation of γ_0 from $Pr(S = 0|X)$

As we noted before, both of our candidate behavioral models produce identical predictions for the event $S = 0$. From Figure 2, the regions defined for $S = 0$ are,

$$\mathcal{R}_S^C(0|X,\beta) = \mathcal{R}_S^{NC}(0|X,\beta) \equiv \mathcal{R}_S(0|X,\beta) = \Big\{(\varepsilon_1,\varepsilon_2) \in \mathbb{R}^2 : \varepsilon_1 \geq X_1^{ns\prime}\beta_1 \;,\; \varepsilon_2 \geq X_2^{ns\prime}\beta_2\Big\}$$

Therefore, $Pr(S = 0|U) = Pr(S = 0|X) = Pr\left(\varepsilon \in \mathcal{R}_S(0|X,\beta_0)|X\right)$. For a given $\gamma \equiv (\beta_1, \beta_2, \rho)$ and X, let

$$\mathbb{P}_0(X,\gamma) \equiv 1 - F_1(X_1^{ns\prime}\beta_1) - F_2(X_2^{ns\prime}\beta_2) + F_{1,2}(X_1^{ns\prime}\beta_1, X_2^{ns\prime}\beta_2|\rho). \tag{8}$$

Our model yields $Pr(S = 0|X) = \mathbb{P}_0(X,\gamma_0)$. Letting $d_0 \equiv \mathbb{1}\{S = 0\}$, we can look at conditions under which γ_0 can be identified from the conditional likelihood of $d_0|X$, given by

$$f(d_0|X) = \mathbb{P}_0(X,\gamma_0)^{d_0} \cdot (1 - \mathbb{P}_0(X,\gamma_0))^{1-d_0}. \tag{9}$$

For $\ell = 1, 2$, let $\nabla_\ell F_{1,2}(t_1,t_2|\rho) \equiv \frac{\partial F_{1,2}(t_1,t_2|\rho)}{\partial t_\ell}$, and let $\nabla_\rho F_{1,2}(t_1,t_2|\rho) \equiv \frac{\partial F_{1,2}(t_1,t_2|\rho)}{\partial \rho}$. From (8),

$$\begin{aligned} \frac{\partial \mathbb{P}_0(X,\gamma)}{\partial \beta_1} &= X_1^{ns} \cdot \Big(\nabla_1 F_{12}(X_1^{ns\prime}\beta_1, X_2^{ns\prime}\beta_2|\rho) - f_1(X_1^{ns\prime}\beta_1)\Big), \\ \frac{\partial \mathbb{P}_0(X,\gamma)}{\partial \beta_2} &= X_2^{ns} \cdot \Big(\nabla_2 F_{12}(X_1^{ns\prime}\beta_1, X_2^{ns\prime}\beta_2|\rho) - f_2(X_2^{ns\prime}\beta_2)\Big), \\ \frac{\partial \mathbb{P}_0(X,\gamma)}{\partial \rho} &= \nabla_\rho F_{1,2}(X_1^{ns\prime}\beta_1, X_2^{ns\prime}\beta_2|\rho) \end{aligned}$$

And group

$$\frac{\partial \mathbb{P}_0(X,\gamma)}{\partial \gamma} = \left(\frac{\partial \mathbb{P}_0(X,\gamma)}{\partial \beta_1}', \frac{\partial \mathbb{P}_0(X,\gamma)}{\partial \beta_2}', \frac{\partial \mathbb{P}_0(X,\gamma)}{\partial \rho} \right)', \quad \frac{\partial \log \mathbb{P}_0(X,\gamma)}{\partial \gamma} = \frac{1}{\mathbb{P}_0(X,\gamma)} \cdot \frac{\partial \mathbb{P}_0(X,\gamma)}{\partial \gamma}.$$

Sufficient conditions for (local) identification of γ_0 follow from full-rank properties of the information matrix (see Rothenberg (1971)). We describe them next.

Assumption I1 ***(Existence of $2+\delta$ moments, an exclusion restriction for (X_1^{ns}, X_2^{ns}), and a full-rank condition)*** *We have $E[||X||^{2+\delta}] < \infty$ for some $\delta > 0$. The true parameter value θ_0 belongs in the interior of the parameter space Θ. For each $p = 1,2$, the support of X_p^{ns} is not contained in any proper linear subspace of $\mathbb{R}^{d_{X_p^{ns}}}$, where $d_{X_p^{ns}} \equiv dim(X_p^{ns})$. For $p,q = \{1,2\}$, there exists a component $X_{p,j}^{ns} \in W_p$ such that $X_{p,j}^{ns} \notin W_q$. The elements of the information matrix*

$$H_0(\gamma) = E\left[\frac{\partial \log \mathbb{P}_0(X,\gamma)}{\partial \gamma} \cdot \frac{\partial \log \mathbb{P}_0(X,\gamma)}{\partial \gamma}' \right]$$

exist and are continuous functions of γ everywhere on Θ. Furthermore, there exists an open neighborhood of γ_0 where $H_0(\gamma)$ has constant rank, and $H_0(\gamma_0)$ is nonsingular. ■

Local identification of γ_0 follows from Assumption I1 using standard arguments (see Rothenberg (1971, Theorem 1)). Without an exclusion restriction between X_1^{ns} and X_2^{ns}, the full-rank property in Assumption I1 may fail. $\widehat{\gamma}$ will henceforth denote the MLE estimator for γ_0 based on the likelihood function (9).

2.5.2 Identifiability of Δ_0 under the assumption that players display noncooperative behavior with strictly positive probability

We will maintain the restrictions in Assumption I1, so γ_0 is identified and we will treat it as known. We will show that $\Delta_0 \equiv (\Delta_{10}, \Delta_{20})$ is identifiable under the maintained assumption that players display noncooperative behavior with nonzero positive probability. That is, under the maintained assumption that $Pr(\pi(Z) < 1) > 0$. Recall that $t_1 \equiv X_1^{ns\prime}\beta_{10}$, and $t_2 \equiv X_2^{ns\prime}\beta_{20}$. From the definitions in (5), for any given $\Delta \in \Theta$,

$$\begin{aligned} m_S^C(X,\gamma_0,\Delta) &\equiv F_{1,2}(t_1 - X_1^{s\prime}\Delta_1 - X_2^{s\prime}\Delta_2\ ,\ t_2 - X_1^{s\prime}\Delta_1 - X_2^{s\prime}\Delta_2 | \rho_0), \\ m_S^{NC}(X,\gamma_0,\Delta) &\equiv F_{1,2}(t_1 - X_1^{s\prime}\Delta_1\ ,\ t_2 - X_2^{s\prime}\Delta_2 | \rho_0). \end{aligned} \tag{10}$$

Recall that $\Xi_S(X,\theta) \equiv m_S^C(X,\theta) - m_S^{NC}(X,\theta)$. From (6), we can express

$$S = m_S^{NC}(X,\theta_0) + \pi(Z) \cdot \Xi_S(X,\theta_0) + \varepsilon_S, \quad \text{where} \quad E[\varepsilon_S | U] = 0. \tag{11}$$

So, $E[\varepsilon_S \cdot \Xi_S(X,\theta_0)|Z] = 0$, and $\pi(Z) \cdot E[\Xi_S(X,\theta_0)^2|Z] = E[(S - m_S^{NC}(X,\theta_0)) \cdot \Xi_S(X,\theta_0)|\, Z]$ a.e Z. The restriction in Assumption G3 implies that $E[\Xi_S(X,\theta_0)^2|Z] > 0$ a.e Z. Therefore,

$$\pi(Z) = \frac{E[(S - m_S^{NC}(X,\theta_0)) \cdot \Xi_S(X,\theta_0)|\, Z]}{E[\Xi_S(X,\theta_0)^2|Z]} \text{ a.e } Z. \tag{12}$$

Note that $E[\Xi_S(X,\theta_0)^2|Z] = 0$ would imply that both candidate behavioral models produce the same prediction for a.e Z. The purpose of Assumption G3 is to rule this out. From (7), we have $E[\Xi_S(X,\theta)^2|Z] > 0 \; \forall \; \theta \in \Theta$, a.e Z. Thus, the following weights are well-defined $\forall$ $\theta \in \Theta$, a.e Z,

$$\pi(Z,\theta) \equiv \frac{E[(S - m_S^{NC}(X,\theta)) \cdot \Xi_S(X,\theta)|\, Z]}{E[\Xi_S(X,\theta)^2|Z]}. \tag{13}$$

From (12), we have $\pi(Z,\theta_0) = \pi(Z)$ for a.e Z. For each $\theta \in \Theta$, let

$$\mu_S(U,\theta) \equiv m_S^{NC}(X,\theta) + \pi(Z,\theta) \cdot \Xi_S(X,\theta). \tag{14}$$

Then,

$$E[S|U] = \mu_S(U,\theta_0). \tag{15}$$

Suppose players cooperate almost surely, so $\pi(Z) = 1$ w.p.1. In this case,

$$\mu_S(U,\theta_0) = m_S^C(X,\theta_0) = F_{1,2}(t_1 - X_1^{s\prime}\Delta_{10} - X_2^{s\prime}\Delta_{20}\ ,\ t_2 - X_1^{s\prime}\Delta_{10} - X_2^{s\prime}\Delta_{20}|\rho_0).$$

Δ_0 would not identifiable unless X_1^s and X_2^s have no elements in common (ruling out constant strategic interaction effects), or unless we reduce the dimensionality of Δ, for example, by assuming $\Delta_{10} = \Delta_{20}$. Let us focus for now on the case where players display noncooperative behavior with nonzero probability, so $Pr(\pi(Z) < 1) > 0$. We will come back the case where players are allowed to cooperate almost surely in Section 2.7.

Assumption I2 ***(Strictly positive probability of noncooperative behavior)*** $Pr\,(\pi(Z) < 1) >$ 0, *so players display noncooperative behavior with strictly positive probability.* ■

Taking γ_0 as identified and maintaining Assumption I2, we will study identification and inference for Δ_0 based on the conditional moment restriction (15). We will follow previous work (see, e.g, Bierens (1982), Bierens and Ploberger (1997), Chen and Fan (1999), Dominguez and Lobato (2004), Khan and Tamer (2009) and Andrews and Shi (2013)), and we will propose a conditional-GMM procedure that converts the conditional moments in (15) into an infinite number of unconditional moments, and then aggregates them into a CvM population statistic. In our construction we will consider the type of instrument-function space proposed by Dominguez and Lobato (2004). Group $V \equiv (S,U)$ and let

$\varphi_S(V,\theta) \equiv S - \mu_S(U,\theta)$. For any $(u,\theta) \in \mathbb{R}^{d_U} \times \Theta$ let,

$$\tau(u,\theta) \equiv E\left[\varphi_S(V,\theta) \cdot \mathbb{1}\{U \le u\}\right]. \tag{16}$$

From (15), using iterated expectations we must have

$$\tau(u,\theta_0) = 0 \quad \forall\ u \in \mathbb{R}^{d_U}. \tag{17}$$

Dominguez and Lobato (2004, equation 2) can be invoked to show that $E[\varphi_S(V,\theta)|U] = 0$ a.e $U \Leftrightarrow \tau(u,\theta) = 0$ for a.e $u \in \mathbb{R}^{d_U}$, a result that follows from Billingsley (1995, Theorem 16.10iii). We will use the space of instrument functions $\{\mathbb{1}\{U \le u\} : u \in \text{Supp}(U)\}$ to transform (15) into an infinite collection of unconditional moment restrictions, and we will aggregate them through the population statistic,

$$Q_S(\theta) \equiv \frac{1}{2}\int \tau(u,\theta)^2 dF_U(u) = \frac{1}{2} \cdot E\Big[\tau(U,\theta)^2\Big]. \tag{18}$$

$Q_S(\theta) \ge 0\ \forall\ \theta$, and θ_0 is a minimizer of $Q_S(\theta)$, satisfying $Q_S(\theta_0) = 0$. By the arguments in Dominguez and Lobato (2004), any θ such that $Q_S(\theta) = 0$ must satisfy $E[S|U] = \mu_S(U,\theta)$ a.e. U. With γ_0 being identified, we will focus on $Q_S(\gamma_0,\Delta) \equiv \frac{1}{2}E[\tau(U,\gamma_0,\Delta)^2]$, where $\tau(u,\gamma_0,\Delta) \equiv E[\varphi_S(V,\gamma_0,\Delta) \cdot \mathbb{1}\{U \le u\}]$ and,

$$\varphi_S(V,\gamma_0,\Delta) = S - \mu_S(U,\gamma_0,\Delta) = S - m_S^{NC}(X,\gamma_0,\Delta) - \pi(Z,\gamma_0,\Delta) \cdot \Xi_S(X,\gamma_0,\Delta) \tag{19}$$

Δ_0 is *identifiable* if $Q_S(\gamma_0,\Delta) > 0\ \forall\ \Delta \in \Theta : \Delta \ne \Delta_0$, and Δ_0 is *locally identifiable* if this holds $\forall$ $\Delta \in \mathcal{A}$, an open neighborhood of Δ_0. Conversely, $\Delta \ne \Delta_0$ is observationally equivalent to Δ_0 if $Q_S(\gamma_0,\Delta) = 0$. We maintain that θ_0 belongs in the interior of Θ, so $\frac{\partial Q_S(\theta_0)}{\partial \Delta} = 0$. Local identifiability of Δ_0 can follow from invertibility conditions of the Hessian $\frac{\partial^2 Q_S(\theta_0)}{\partial\Delta\partial\Delta'}$. For $p = 1,2$, denote $\nabla_p F_{1,2}(\epsilon_1,\epsilon_2|\rho) \equiv \frac{\partial F_{1,2}(\epsilon_1,\epsilon_2|\rho)}{\partial \epsilon_p}$ and let $H_p^{NC}(X,\theta) \equiv \nabla_p F_{1,2}(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1\ ,\ X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2|\rho)$, and $H^C(X,\theta) \equiv \sum_{q=1}^2 \nabla_q F_{1,2}(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1 - X_2^{s\prime}\Delta_2\ ,\ X_2^{ns\prime}\beta_2 - X_1^{s\prime}\Delta_1 - X_2^{s\prime}\Delta_2|\rho)$, and denote $J_p(U,\theta) \equiv \pi(Z,\theta) \cdot H^C(X,\theta) + (1-\pi(Z,\theta)) \cdot H_p^{NC}(X,\theta)$. From here, let,

$$\dot{\mu}_{S,\Delta}(U,\theta) \equiv \begin{pmatrix} -J_1(U,\theta) \cdot X_1^S + \frac{E\left[J_1(U,\theta)\cdot\Xi_S(X,\theta)\cdot X_1^S\ \middle|\ Z\right]}{E\left[\Xi_S(X,\theta)^2\ \middle|\ Z\right]} \cdot \Xi_S(X,\theta) \\ -J_2(U,\theta) \cdot X_2^S + \frac{E\left[J_2(U,\theta)\cdot\Xi_S(X,\theta)\cdot X_2^S\ \middle|\ Z\right]}{E\left[\Xi_S(X,\theta)^2\ \middle|\ Z\right]} \cdot \Xi_S(X,\theta) \end{pmatrix}$$

In Appendix A (Section 2.5.2), we show that,

$$\frac{\partial^2 Q_S(\theta_0)}{\partial\Delta\partial\Delta'} = \int E\left[\dot{\mu}_{S,\Delta}(U,\theta_0) \cdot \mathbb{1}\{U \le u\}\right] \cdot E\left[\dot{\mu}_{S,\Delta}(U,\theta_0)' \cdot \mathbb{1}\{U \le u\}\right] dF_U(u). \tag{20}$$

From here, we can describe sufficient conditions for invertibility of $\frac{\partial^2 Q_S(\theta_0)}{\partial\Delta\partial\Delta'}$.

Assumption I3 *(A full rank restriction leading to local identification)*

Maintain the conditions in Assumptions I1 and I2. In addition, for each $p = 1,2$, the support of X_p^s is not contained in any proper linear subspace of $\mathbb{R}^{d_{X_p^s}}$, and there exists an open neighborhood $\mathcal{N}$ of Δ_0 such that $\int E\left[\dot{\mu}_{S,\Delta}(U,\gamma_0,\Delta)\cdot \mathbb{1}\{U\leq u\}\right]\cdot E\left[\dot{\mu}_{S,\Delta}(U,\gamma_0,\Delta)'\cdot \mathbb{1}\{U\leq u\}\right]dF_U(u)$ is invertible for all $\Delta\in\mathcal{N}$. ■

Assumption I3 does not require that (X_1^s, X_2^s) have full rank, allowing for $X_1^s = X_2^s$ as a special case.

Remark 1 *Assumption I3 cannot be satisfied if players cooperate almost surely, but it can be satisfied if players display noncooperative behavior almost surely.*

If $\pi(Z) = 1$ w.p.1 and players cooperate a.s, then $J_p(U,\theta_0) = H^C(X,\theta_0)$ and,

$$\dot{\mu}_{S,\Delta}(U,\theta_0) \equiv \begin{pmatrix} -H^C(X,\theta_0)\cdot X_1^S + \frac{E\left[H^C(X,\theta_0)\cdot\Xi_S(X,\theta_0)\cdot X_1^S \,\middle|\, Z\right]}{E\left[\Xi_S(X,\theta_0)^2 \,\middle|\, Z\right]}\cdot \Xi_S(X,\theta_0) \\ -H^C(X,\theta_0)\cdot X_2^S + \frac{E\left[H^C(X,\theta_0)\cdot\Xi_S(X,\theta_0)\cdot X_2^S \,\middle|\, Z\right]}{E\left[\Xi_S(X,\theta_0)^2 \,\middle|\, Z\right]}\cdot \Xi_S(X,\theta_0). \end{pmatrix}$$

Satisfying Assumption I3 would require (X_1^s, X_2^s) to have full rank, which cannot be satisfied if X_1^s and X_2^s have elements in common, confirming our previous claims regarding almost sure cooperation. Now suppose players display noncooperative behavior almost surely. In this case, $\pi(Z) = 0$ and $J_p(U,\theta_0) = H_p^{NC}(X,\theta_0)$. Thus,

$$\dot{\mu}_{S,\Delta}(U,\theta_0) \equiv \begin{pmatrix} -H_1^{NC}(X,\theta_0)\cdot X_1^S + \frac{E\left[H_1^{NC}(X,\theta_0)\cdot\Xi_S(X,\theta_0)\cdot X_1^S \,\middle|\, Z\right]}{E\left[\Xi_S(X,\theta_0)^2 \,\middle|\, Z\right]}\cdot \Xi_S(X,\theta_0) \\ -H_2^{NC}(X,\theta_0)\cdot X_2^S + \frac{E\left[H_2^{NC}(X,\theta_0)\cdot\Xi_S(X,\theta_0)\cdot X_2^S \,\middle|\, Z\right]}{E\left[\Xi_S(X,\theta_0)^2 \,\middle|\, Z\right]}\cdot \Xi_S(X,\theta_0) \end{pmatrix}$$

Assumption I3 can be satisfied even if $X_1^S = X_2^S$ since $Pr(H_1^{NC}(X,\theta_0) \neq H_2^{NC}(X,\theta_0)) > 0$ can follow from our exclusion restrictions between X_1^{NS} and X_2^{NS}. Global identification requires that Δ_0 be the *unique* minimizer of $Q_S(\gamma_0,\Delta)$ over $\Delta\in\Theta$. This will occur if $Pr\left(\mu_S(U,\theta_0)\neq\mu_S(U,\gamma_0,\Delta)\right) > 0\ \forall\ \Delta\neq\Delta_0$. To this end, consider the following restriction.

Assumption I4 *$\forall\ \Delta\neq\Delta'$ in Θ and a.e Z, $Pr(m_S^{NC}(X,\gamma_0,\Delta)\neq m_S^{NC}(X,\gamma_0,\Delta')\,|Z) > 0$, and $Pr(m_S^C(X,\gamma_0,\Delta) - m_S^{NC}(X,\gamma_0,\Delta)\neq m_S^C(X,\gamma_0,\Delta') - m_S^{NC}(X,\gamma_0,\Delta')\,|Z) > 0$.* ■

From (13)-(14), Assumption I4 yields $Pr\left(\mu_S(U,\theta_0)\neq\mu_S(U,\gamma_0,\Delta)\right) > 0\ \forall\ \Delta\neq\Delta_0$ in Θ, which implies that Δ_0 is the unique minimizer of $Q_S(\gamma_0,\Delta)$ over $\Delta\in\Theta$.

2.6 An estimator for Δ_0 under the assumption that players display noncooperative behavior with strictly positive probability

Recall that $\widehat{\gamma} \equiv (\widehat{\beta}_1', \widehat{\beta}_2', \widehat{\rho})'$ denotes our MLE estimator for $\gamma_0 \equiv (\beta_{1i}', \beta_{20}' \rho_0')'$. Under the restrictions in Assumption I1, $\widehat{\gamma}$ satisfies the usual MLE linear representation,

$$\widehat{\gamma} = \gamma_0 + \frac{1}{n}\sum_{i=1}^{n} \psi_\gamma(V_i;\gamma_0) + o_p(n^{-1/2}), \tag{21}$$

where $\psi_\gamma(V_i;\gamma_0)$ is the MLE influence function. Maintaining Assumptions I2-I4, our proposal is to construct a conditional-GMM estimator for Δ_0 that minimizes a sample analog of $Q_S(\gamma_0,\Delta)$. We first construct a kernel-based estimator of the behavioral weight $\pi(Z,\gamma_0,\Delta)$ described in (13). Let $K : \mathbb{R}^{d_Z} \to \mathbb{R}$ and $h_n \to 0$ denote a kernel function and a bandwidth sequence respectively (whose properties we will describe below). Let $\widehat{t}_{1i} \equiv X_{1i}^{ns\prime}\widehat{\beta}_1$, $\widehat{t}_{2i} \equiv X_{2i}^{ns\prime}\widehat{\beta}_2$. Following the definitions in (10), let,

$$\begin{aligned}
m_S^C(X_i,\widehat{\gamma},\Delta) &\equiv F_{1,2}\left(\widehat{t}_{1i} - X_{1i}^{s}{}'\Delta_1 - X_{2i}^{s}{}'\Delta_2 \;, \widehat{t}_{2i} - X_{1i}^{s}{}'\Delta_1 - X_{2i}^{s}{}'\Delta_2 | \widehat{\rho}\right),\\
m_S^{NC}(X_i,\widehat{\gamma},\Delta) &\equiv F_{1,2}\left(\widehat{t}_{1i} - X_{1i}^{s}{}'\Delta_1 \;, \widehat{t}_{2i} - X_{2i}^{s}{}'\Delta_2 | \widehat{\rho}\right),\\
\Xi_S(X_i,\widehat{\gamma},\Delta) &\equiv m_S^C(X_i,\widehat{\gamma},\Delta) - m_S^{NC}(X_i,\widehat{\gamma},\Delta).
\end{aligned}$$

Define,

$$\widehat{\pi}(z,\theta) \equiv \frac{\sum_{i=1}^n \left(S_i - m_S^{NC}(X_i,\theta)\right)\cdot \Xi_S(X_i,\theta)\cdot K\left(\frac{Z_i - z}{h_n}\right)}{\sum_{i=1}^n \Xi_S(X_i,\theta)^2 \cdot K\left(\frac{Z_i - z}{h_n}\right)}. \tag{22}$$

From here, let

$$\widehat{\varphi}_S(V_i,\widehat{\gamma},\Delta) \equiv S_i - m_S^{NC}(X_i,\widehat{\gamma},\Delta) - \widehat{\pi}(Z_i,\widehat{\gamma},\Delta)\cdot \Xi_S(X_i,\widehat{\gamma},\Delta). \tag{23}$$

$\widehat{\varphi}_S(V_i,\widehat{\gamma},\Delta)$ is an estimator for $\varphi_S(V_i,\gamma_0,\Delta)$ as defined in (19). Let $\mathcal{Z} \subseteq \text{Supp}(Z)$ denote a pre-specified *inference range* for Z, whose choice is assumed to ensure (under restrictions we will described below) uniform asymptotic properties for $\widehat{\pi}(z,\theta)$ over $\mathcal{Z}\times\Theta$. For any $u \in \mathbb{R}^{d_U}$, let $\mathbb{1}_{\mathcal{Z}}\{U_i \le u\} \equiv \mathbb{1}\{U_i \le u \;, Z_i \in \mathcal{Z}\}$ and,

$$\widehat{\tau}_{\mathcal{Z}}(u,\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} \widehat{\varphi}_S(V_i,\theta)\cdot \mathbb{1}_{\mathcal{Z}}\{U_i \le u\}. \tag{24}$$

$\widehat{\tau}_{\mathcal{Z}}(u,\theta)$ is an estimator for $\tau_{\mathcal{Z}}(u,\theta) \equiv E\left[\varphi_S(V,\theta)\cdot \mathbb{1}_{\mathcal{Z}}\{U \le u\}\right]$, which is the version of $\tau(u,\theta)$ when we restrict $Z \in \mathcal{Z}$. As with $\tau(u,\theta)$, we have $\tau_Z(u,\theta_0) = 0 \;\forall\; u$. Now let $Q_{S,\mathcal{Z}}(\theta) \equiv \frac{1}{2}\cdot\int \tau_{\mathcal{Z}}(u,\theta)^2 dF_U(u) = \frac{1}{2}\cdot E\left[\tau_{\mathcal{Z}}(U,\theta)^2\right]$, which is a restricted version of $Q_S(\theta)$ when we integrate out Z over $\mathcal{Z}$. We will characterize Δ_0 as the minimizer of $Q_{S,\mathcal{Z}}(\gamma_0,\Delta)$. Accordingly, our sample objective function is,

$$\widehat{Q}_{S,\mathcal{Z}}(\widehat{\gamma},\Delta) \equiv \frac{1}{2}\cdot\frac{1}{n}\sum_{j=1}^{n}\widehat{\tau}_{\mathcal{Z}}(U_j,\widehat{\gamma},\Delta)^2 = \frac{1}{2}\cdot\frac{1}{n}\sum_{j=1}^{n}\left[\frac{1}{n}\sum_{i=1}^{n}\widehat{\varphi}_S(V_i,\widehat{\gamma},\Delta)\cdot\mathbb{1}_{\mathcal{Z}}\{U_i \leq U_j\}\right]^2 \qquad (25)$$

our estimator for Δ_0 is given by $\widehat{\Delta} = \underset{\Delta\in\Theta}{argmin}\ \widehat{Q}_{S,\mathcal{Z}}(\widehat{\gamma},\Delta)$. We will describe its asymptotic properties under the following set of restrictions.

Assumption E1 ***(Bandwidth and kernel restrictions)*** *Our nonnegative bandwidth sequence satisfies $n^{1/2}\cdot h_n^{2d_Z} \longrightarrow \infty$, and for an integer $M \geq 2d_Z + 1$, we have $n^{1/2}\cdot h_n^M \longrightarrow 0$. Our kernel $K : \mathbb{R}^{d_Z} \longrightarrow \mathbb{R}$ is bias-reducing of order M (for the integer M described above), and is symmetric around zero, with bounded support of the form $[-S,S]$. Our kernel is a function of bounded variation, satisfying $|K(\psi)| \leq \overline{K}\ \forall\ \psi \in \mathbb{R}^{d_Z}$, for some $\overline{K} < \infty$.* ■

Next, we decribe smoothness restrictions involving some functionals of Z. These are comparable to commonly maintained assumptions in semiparametric models with nonparametric components or "generated regressors".

Assumption E2 ***(Smoothness restrictions of some functionals of Z)***

(i) Let M be the integer described in Assumption E1. The following functionals are M−times differentiable with respect to z, with bounded derivatives for all $(z,\theta)\in\mathcal{Z}\times\Theta$: $f_Z(z)$, $\mu_I^{\Xi_S}(z,\theta) \equiv E[(S-m_S^{NC}(X,\theta))\cdot\Xi_S(X,\theta)\,|Z=z]$, $\mu_{II,2}^{\Xi_S}(z,\theta) \equiv E[\Xi_S(X,\theta)^2\,|Z=z]$, $\mu_{III,\theta_\ell}^{\Xi_S}(z,\theta) \equiv E\Big[\frac{\partial\Xi_S(X,\theta)}{\partial\theta_\ell}\cdot(S-m_S^{NC}(X,\theta))\,|Z=z\Big]$, $\mu_{IV,\theta_\ell}^{\Xi_S}(z,\theta) \equiv E\Big[\frac{\partial m_S^{NC}(X,\theta)}{\partial\theta_\ell}\cdot\Xi_S(X,\theta)\,|Z=z\Big]$, $\mu_{V,\theta_\ell}^{\Xi_S}(z,\theta) \equiv E\Big[\frac{\partial\Xi_S(X,\theta)}{\partial\theta_\ell}\cdot\Xi_S(X,\theta)\,|Z=z\Big]$, and $\mu_{VI,\theta_\ell}^{\Xi_S}(z,\theta) \equiv E\Big[\frac{\partial\Xi_S(X,\theta)}{\partial\theta_\ell}\,|Z=z\Big]$. These functionals are bounded above in absolute value by some $\overline{C} < \infty$. $f_Z(z)$ and $\mu_{II,2}^{\Xi_S}(z,\theta)$ are bounded below by some $\underline{C} > 0$.

(ii) The following functionals are continuously differentiable with respect to z, with bounded first derivative for all $(z,\theta)\in\mathcal{Z}\times\Theta$: $\delta_{I,\theta_\ell,\theta_j}(z,\theta) \equiv E\Big[\frac{\partial^2\Xi_S(X,\theta)}{\partial\theta_j\partial\theta_\ell}\cdot(S-m_S^{NC}(X,\theta))\,|Z=z\Big]$, $\delta_{II,\theta_\ell,\theta_j}(z,\theta) \equiv E\Big[\frac{\partial\Xi_S(X,\theta)}{\partial\theta_\ell}\cdot\frac{\partial m_S^{NC}(X,\theta)}{\partial\theta_j}\,|Z=z\Big]$, $\delta_{III,\theta_\ell,\theta_j}(z,\theta) \equiv E\Big[\frac{\partial^2 m_S^{NC}(X,\theta)}{\partial\theta_j\partial\theta_\ell}\cdot\Xi_S(X,\theta)\,|Z=z\Big]$, $\delta_{IV,\theta_\ell,\theta_j}(z,\theta) \equiv E\Big[\frac{\partial^2\Xi_S(X,\theta)}{\partial\theta_j\partial\theta_\ell}\cdot\Xi_S(X,\theta)\,|Z=z\Big]$, and $\delta_{V,\theta_\ell,\theta_j}(z,\theta) \equiv E\Big[\frac{\partial\Xi_S(X,\theta)}{\partial\theta_\ell}\cdot\frac{\partial\Xi_S(X,\theta)}{\partial\theta_j}\,|Z=z\Big]$.
■

Finally, we introduce some smoothness and regularity restrictions with respect to θ. These are comparable to commonly maintained assumptions for extremum estimators.

Assumption E3 ***(Lipschitz-restrictions with respect to*** $\boldsymbol{\theta}$***)***

Each of the functions below has the following type of Lipschitz property: $|g(u,\theta)-g(u,\theta')| \leq \overline{G}(u)\cdot\|\theta-\theta'\|\ \forall\ u \in Supp(U)$ *(unless noted otherwise), and* $\forall\ \theta,\theta' \in \Theta$*, where* $\overline{G}(\cdot)$ *is a nonnegative function that satisfies* $E[\overline{G}(U)^{2+\delta}] < \infty$ *for some* $\delta > 0$*.*

(i) The Lipschitz property holds for all $x \in Supp(X)$ *for the following functions:* $m_S^{NC}(x,\theta)$*,* $m_S^{C}(x,\theta)$*,* $\frac{\partial m_S^C(x,\theta)}{\partial\theta_\ell}$*,* $\frac{\partial m_S^C(x,\theta)}{\partial\theta_\ell}$*, with* $\overline{G}(X)$ *satisfying* $E[\overline{G}(X)^{2+\delta}] < \infty$ *for some* $\delta > 0$*.*

(ii) The Lipschitz property holds for all $z \in \mathcal{Z}$ *for the following functionals defined in Assumption E2:* $\mu_I^{\Xi_S}(z,\theta)$*,* $\mu_{II,2}^{\Xi_S}(z,\theta)$*,* $\mu_{III,\theta_\ell}^{\Xi_S}(z,\theta)$*,* $\mu_{IV,\theta_\ell}^{\Xi_S}(z,\theta)$*,* $\mu_{V,\theta_\ell}^{\Xi_S}(z,\theta)$*,* $\mu_{VI,\theta_\ell}^{\Xi_S}(z,\theta)$*. In each case,* $\overline{G}(Z)$ *satisfies* $E[\overline{G}(Z)^{2+\delta}] < \infty$ *for some* $\delta > 0$*.*

(iii) The Lipschitz property holds for all $z \in \mathcal{Z}$ *for the following functionals defined in Assumption E2:* $\delta_{I,\theta_\ell,\theta_j}(z,\theta)$*,* $\delta_{II,\theta_\ell,\theta_j}(z,\theta)$*,* $\delta_{III,\theta_\ell,\theta_j}(z,\theta)$*,* $\delta_{IV,\theta_\ell,\theta_j}(z,\theta)$*,* $\delta_{V,\theta_\ell,\theta_j}(z,\theta)$*. In each case,* $\overline{G}(Z)$ *satisfies* $E[\overline{G}(Z)^{2+\delta}] < \infty$ *for some* $\delta > 0$*.* ■

Proposition 1 *If the restrictions in Assumptions G1-G3, I1-I4 and E1-E3 hold, the estimator* $\widehat{\Delta}$ *satisfies* $\widehat{\Delta} \xrightarrow{p} \Delta_0$*, with an asymptotic linear representation of the form,*

$$\widehat{\Delta} - \Delta_0 = \frac{1}{n}\sum_{i=1}^{n}\psi_{\Delta,n}(V_i;\theta_0) + o_p\left(n^{-1/2}\right), \tag{26}$$

and $\sqrt{n}\cdot\left(\widehat{\Delta}-\Delta_0\right) \xrightarrow{d} \mathcal{N}(0,\Omega_\Delta)$*, where* $\Omega_\Delta = \lim_{n\to\infty} E\left[\psi_{\Delta,n}(V;\theta_0)\cdot\psi_{\Delta,n}(V;\theta_0)'\right]$*.*

Proof: The proof follows standard arguments in semiparametric models. A description of the influence function $\psi_{\Delta,n}(V_i;\theta_0)$ is included in Appendix A (Section A5). The step-by-step details are included in the online Econometric Supplement (Section S1). ■

Remark 2 *Proposition 1 only requires the exclusion restriction* $\xi|(X,Z) \sim \xi|Z$ *in Assumption G2 to be satisfied over our inference range* $\mathcal{Z}$*. That is, we only need,* $Z \in \mathcal{Z} \Rightarrow Pr(\xi=1|X,Z) = Pr(\xi=1|Z)$*.*

2.6.1 Estimating the behavioral weight $\pi(z)$

Equipped with $\widehat{\theta} \equiv (\widehat{\gamma},\widehat{\Delta})$, we can estimate $\pi(z)$, the probability of cooperation conditional on $Z = z$, with $\widehat{\pi}(z,\widehat{\theta})$, where $\widehat{\pi}(z,\theta)$ is as described in (22). Under the restrictions leading to Proposition 1, we can show that[3] $\sup_{\substack{\theta\in\Theta\\ z\in\mathcal{Z}}}|\widehat{\pi}(z,\theta)-\pi(z,\theta)| = o_p(1)$ and,

$$\widehat{\pi}(z,\theta)=\pi(z,\theta)+\frac{1}{n\cdot h_n^{d_z}}\sum_{i=1}^{n}\psi_n^{\pi}(V_i;z,\theta)+\vartheta_n^{\pi}(z,\theta), \quad \text{where} \quad \sup_{\substack{\theta\in\Theta\\ z\in\mathcal{Z}}}\left|\vartheta_n^{\pi}(z,\theta)\right| = o_p\left(n^{-1/2}\right), \tag{27}$$

The influence function $\psi_n^{\pi}(V_i;z,\theta)$ is described in the online Econometric Supplement (equation S1.11). Since $\pi(z)=\pi(z,\theta_0)$, Proposition 1 then yields $\sup_{z\in\mathcal{Z}}\left|\widehat{\pi}(z,\widehat{\theta})-\pi(z)\right| = o_p(1)$.

2.6.2 Testing the validity of Z and our overall model through a consistent specification tests

The validity of the instrument Z, and the restrictions leading to Proposition 1 can be consistently tested using existing methods for semiparametric models. The goal would be to test "$H_0 : E[S|U] = \mu_S(U,\theta_0)$ for a.e U" against "$H_1 : Pr(E[S|U] \neq \mu_S(U,\theta_0)) > 0$". Following Fan and Li (1996), Zheng (1996) and Aradillas-López (2012, Section 4), we can focus on testing $H_0 : E[\varepsilon_S E[\varepsilon_S|U]f_U(U)] = 0$, where $\varepsilon_S \equiv S-\mu_S(U,\theta_0)$ and $f_U(\cdot)$ is the density function of U. We can rewrite this as $H_0 : E[(E[\varepsilon_S|U])^2 f_U(U)] = 0$, which would hold iff $E[S|U]=\mu_S(U,\theta_0)$ for a.e U. Assuming that U is jointly continuously distributed, a test can be based on a analog test-statistic $\frac{1}{n\cdot(n-1)\cdot b_n^{d_U}}\sum_{i=1}^{n}\sum_{j\neq i}\widehat{\varepsilon_i}\cdot\widehat{\varepsilon_j}\cdot\mathcal{K}\left(\frac{U_i-U_j}{b_n}\right)$, where $\widehat{\varepsilon_i}\equiv S_i-\widehat{\mu_S}(U_i,\widehat{\theta})$, $b_n\to 0$ is a nonnegative bandwidth sequence, and $\mathcal{K}:\mathbb{R}^{d_U}\to\mathbb{R}$ is a kernel function. Its asymptotic properties can be derived following the steps in Aradillas-López (2012, Section 4), which studies semiparametric models with generated regressors that share the asymptotic properties of $\widehat{\pi}(z,\widehat{\theta})$ described in (27).

2.7 Inference for Δ_0 when we allow for the possibility that players cooperate almost surely

As we showed above, allowing for the possibility that $\pi(Z)=1$ w.p.1 implies that Δ_0 may not be identified without further restrictions. In this case, our suggestion is to use (6) to construct a *confidence set* (CS) for Δ_0. Once again, we maintain Assumption I1 so γ_0 is identified and estimable using MLE. While there are many ways to proceed, we illustrate an approach based on functionals similar to those we used for our conditional-GMM estimator in Section 2.6. Let $\mathcal{Z}$ and $\mathbb{1}_{\mathcal{Z}}\{U\leq u\}$ be as defined previously. Next, let $g:\mathbb{R}^{d_U}\to\mathbb{R}$ be a real-valued, pre-specified, positive function[4] of U. As before, let

[3] The step-by-step details can be found in the online Econometric Supplement (Section S1.1).

[4] Our results can be straightforwardly extended to a vector-valued instrument function g.

$\varphi_S(V,\theta) \equiv S - m_S^{NC}(X,\theta) - \pi(Z,\theta)\cdot \Xi_S(X,\theta)$. For a given $u \in \mathbb{R}^{d_U}$, let

$$\tau_g(u,\theta) \equiv E\left[\varphi_S(V,\theta)\cdot g(U)\cdot \mathbb{1}_{\mathcal{Z}}\{U \le u\}\right], \quad M_g(\theta) \equiv E[\tau_g(U,\theta)]. \tag{28}$$

By iterated expectations, $M_g(\theta_0) = 0$. Since γ_0 is identified, our proposal would be to use $M_g(\gamma_0,\Delta)$ to construct a CS for Δ. Let $\Theta_g^I \equiv \{\Delta \in \Theta : M_g(\gamma_0,\Delta) = 0\}$ be our target identified set for Δ based on the moment restriction $M_g(\theta_0) = 0$. Our sample statistic is

$$\widehat{M}_g(\widehat{\gamma},\Delta) = \frac{1}{n}\sum_{j=1}^{n}\widehat{\tau}_g(U_j,\widehat{\gamma},\Delta) = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\widehat{\varphi}_S(V_i,\widehat{\gamma},\Delta)\cdot g(U_i)\cdot \mathbb{1}_{\mathcal{Z}}\{U_i \le U_j\} \tag{29}$$

Proposition 2 *Suppose $E[g(U)^{2+\delta}] < \infty$ for some $\delta > 0$. If the restrictions in Assumptions G1-G3, I1, and E1-E3 hold, then*

$$\begin{aligned} &\widehat{M}_g(\widehat{\gamma},\Delta) = M_g(\gamma_0,\Delta) + \frac{1}{n}\sum_{i=1}^{n}\psi_{M_g,n}(V_i;\gamma_0,\Delta) + \varsigma_n^{M_g}(\Delta),\ \forall\ \Delta \in \Theta, \\ &\textit{where} \quad E[\psi_{M_g,n}(V;\gamma_0,\Delta)] = 0\ \forall\ \Delta \in \Theta, \quad \sup_{\Delta\in\Theta}\left|\varsigma_n^{M_g}(\Delta)\right| = o_p\left(n^{-1/2}\right) \end{aligned} \tag{30}$$

And over our target identified set Θ_g^I, the result in (S2.25) *simplifies to,*

$$\widehat{M}_g(\widehat{\gamma},\Delta) = \frac{1}{n}\sum_{i=1}^{n}\psi_{M_g,n}(V_i;\gamma_0,\Delta) + \varsigma_n^{M_g}(\Delta)\ \forall\ \Delta \in \Theta_g^I, \quad \textit{where} \quad \sup_{\Delta\in\Theta}\left|\varsigma_n^{M_g}(\Delta)\right| = o_p\left(n^{-1/2}\right). \tag{31}$$

Proof: The proof follows standard arguments in semiparametric models. A description of the influence function $\psi_{M_g,n}(V_i;\gamma_0,\Delta)$ is included in Appendix A (Section A6). The step-by-step details can be found in the online Econometric Supplement (Section S2). ■

Suppose the data-generating process belongs to a family of distributions $\mathcal{F}$. For each $F \in \mathcal{F}$ denote the corresponding identified set as $\Theta_{g,F}^I \equiv \{\Delta \in \Theta : M_{g,F}(\gamma_0,\Delta) = 0\}$. Note that $\Delta_0 \in \Theta_{g,F}^I$ for any F. A CS with uniform asymptotic coverage properties over $\mathcal{F}\times\Theta_{g,F}^I$ can be constructed based on Proposition 2 if we add the following restriction.

Assumption E4 *Suppose the DGP F belongs to a family of distributions $\mathcal{F}$ that satisfy all the restrictions in Assumptions G1-G3, I1, and E1-E3 and, in particular, suppose that the existence of $2+\delta$ moments in Assumptions I1 and E3, and the smoothness restrictions in Assumptions E2 and E3 hold uniformly over $\mathcal{F}$ (i.e, the bounds described for each one of those restrictions are common to every $F \in \mathcal{F}$). In addition, suppose that the constant δ described in the existence of $2+\delta$ moment restrictions satisfies $\delta \ge 1$, and that for some $\delta \ge 1$ and $\overline{C} < \infty$, the instrument* *function g also satisfies $E_F[|g(U)|^{2+\delta}] \leq \overline{C}$ for all $F \in \mathcal{F}$.* ∎

Let $\sigma^F_{M_g,n}(\Delta)^2 \equiv E_F\left[\psi^F_{M_g,n}(V;\gamma_0,\Delta)^2\right]$ and $\widehat{\sigma}_{M_g,n}(\Delta)^2 = \frac{1}{n}\sum_{i=1}^n \widehat{\psi}_{M_g,n}(V_i;\widehat{\gamma},\Delta)^2$, where $\widehat{\psi}_{M_g,n}(v;\gamma,\Delta)$ is an estimator of $\psi^F_{M_g,n}(v;\gamma,\Delta)$. Based on Proposition 2, a CS for Δ_0 with target asymptotic coverage probability $1-\alpha$ can be constructed as,

$$CS_n^{\Delta}(1-\alpha) = \left\{\Delta \in \Theta : \left|\frac{n^{1/2}\cdot \widehat{M}_g(\widehat{\gamma},\Delta)}{\widehat{\sigma}_{M_g,n}(\Delta)}\right| \leq \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right\} \tag{32}$$

In the online Econometric Supplement (Section S2.1) we show that, if Assumption E4 holds and if $\widehat{\sigma}_{M_g,n}(\Delta)^2 \xrightarrow{p} \sigma^F_{M_g,n}(\Delta)^2$ uniformly over $\mathcal{F}\times\Theta$, then

$$\lim_{n\to\infty} \sup_{\substack{(F,\Delta)\in\mathcal{F}\times\Theta: \\ \Delta\in\Theta^I_{g,F}}} \left|P_F\left(\Delta \in CS_n^{\Delta}(1-\alpha)\right) - (1-\alpha)\right| = 0,$$

so our CS would have correct asymptotic coverage probability for Δ_0. An empty CS would reject our model.

2.7.1 Inference for the behavioral weight $\pi(z)$

Fix $(z,\Delta) \in \text{Supp}(Z)\times\Theta$. Using (27), standard arguments imply that a $1-\alpha$ confidence interval (CI) for $\pi(z,\gamma_0,\Delta)$ can be constructed as $\widehat{\pi}(z,\widehat{\gamma},\Delta) \pm (n\cdot h_n^{d_Z})^{-1/2}\cdot\Phi^{-1}(1-\alpha/2)\cdot \widehat{\sigma}_{\pi}(z,\widehat{\gamma},\Delta)$, where $\widehat{\sigma}^2_{\pi}(z,\widehat{\gamma},\Delta)$ is an estimator of $E\left[\frac{1}{h_n^{d_Z}}\psi^{\pi}_n(V;z,\gamma_0,\Delta)^2\right]$. If Δ_0 were known, it could be plugged into the previous expression to obtain the desired CI for $\pi(z,\theta_0) = \pi(z)$. Since Δ_0 is unknown (and possibly nonidentifiable if $\pi(Z)=1$ w.p.1), we can use a Bonferroni bound together with $CS_n^{\Delta}(1-\alpha)$ to construct a valid $1-2\alpha$ CI for $\pi(z)$,

$$CS_n^{\pi(z)}(1-2\alpha) =$$
$$\left[\min_{\Delta\in CS_n^{\Delta}(1-\alpha)}\left(\widehat{\pi}(z,\widehat{\gamma},\Delta) - \frac{\Phi^{-1}(1-\alpha/2)}{\sqrt{n\cdot h_n^{d_Z}}}\cdot\widehat{\sigma}_{\pi}(z,\widehat{\gamma},\Delta)\right), \max_{\Delta\in CS_n^{\Delta}(1-\alpha)}\left(\widehat{\pi}(z,\widehat{\gamma},\Delta) + \frac{\Phi^{-1}(1-\alpha/2)}{\sqrt{n\cdot h_n^{d_Z}}}\cdot\widehat{\sigma}_{\pi}(z,\widehat{\gamma},\Delta)\right)\right].$$

3 An expanded model

Let us maintain our normal-form parameterization and consider the following *expanded* collection of candidate behavioral models: (1) Cooperation with complete information. (2) Complete-information Nash equilibrium behavior allowing for mixed strategies. (3) Incomplete-information Bayesian Nash equilibrium. We describe them next.

3.1 Cooperation

This behavioral model remains as described previously. The relevant regions for each $y \in \mathcal{Y}$ under cooperation are shown in the top panel of Figure 1. Let

$$\mathcal{R}^C(y|X,\theta_0) \equiv \Big\{(\varepsilon_1,\varepsilon_2) \in \mathbb{R}^2 : y \text{ is the optimal outcome under cooperation, given } (X,\theta_0)\Big\}.$$

These regions are depicted on the top panel of Figure 1.

3.2 Complete-information Nash equilibrium behavior with mixed strategies

Suppose we allow for mixed strategies in our complete-information Nash equilibrium (NE) model. For each $y \in \mathcal{Y}$ let

$$\mathcal{R}^{NE}(y|X,\theta_0) \equiv \Big\{(\varepsilon_1,\varepsilon_2) \in \mathbb{R}^2 : y \text{ is the unique NE of the game, given } (X,\theta_0)\Big\}$$

These regions can be obtained from Figure 1, where we can also obtain $\mathcal{M}_1^{NE}(X,\theta_0)$, the multiple NE region. When $\varepsilon \in \mathcal{M}_1^{NE}(X,\theta_0)$, the game has two pure-strategy NE (PSNE), $(1,0)$ and $(0,1)$, and one mixed-strategy NE (MSNE). We will index the multiple NE as

- NE #1: PSNE $(1,0)$, • NE #2: PSNE $(0,1)$, • NE #3: MSNE.

The possibility of selecting the MSNE means that $S \equiv Y_1 + Y_2$ is no longer uniquely predicted by our model. Thus, instead of S, we will focus directly on the predictions for Y made by each of our candidate behavioral models. Let $Y_{(\ell)}$ denote the (potential) outcome that would be produced by NE $\ell \in \{1,2,3\}$ in the multiple NE region $\mathcal{M}_1^{NE}(X,\theta_0)$. Denoting $Pr(Y_{(\ell)} = y|X,\varepsilon) \equiv \sigma^{\ell}_{\mathcal{M}_1,NE}(y|X,\varepsilon,\theta_0)$ and solving for the MSNE, we have

- For NE #1 (PSNE $(1,0)$): $\sigma^1_{\mathcal{M}_1,NE}(y|X,\varepsilon,\theta_0) = \mathbb{1}\{y = (1,0)\}$,
- For NE #2 (PSNE $(0,1)$): $\sigma^2_{\mathcal{M}_1,NE}(y|X,\varepsilon,\theta_0) = \mathbb{1}\{y = (0,1)\}$,
- For NE #3 (MSNE):

$$\sigma^3_{\mathcal{M}_1,NE}(y|X,\varepsilon,\theta_0) = \left(\frac{X_2^{ns\prime}\beta_{20}-\varepsilon_2}{X_2^{s\prime}\Delta_{20}}\right)^{y_1}\left(1-\frac{X_2^{ns\prime}\beta_{20}-\varepsilon_2}{X_2^{s\prime}\Delta_{20}}\right)^{1-y_1}\left(\frac{X_1^{ns\prime}\beta_{10}-\varepsilon_1}{X_1^{s\prime}\Delta_{10}}\right)^{y_2}\left(1-\frac{X_1^{ns\prime}\beta_{10}-\varepsilon_1}{X_1^{s\prime}\Delta_{10}}\right)^{1-y_2} \tag{33}$$

3.3 Incomplete-information Bayesian Nash equilibrium behavior

Let us expand our collection of candidate behavioral models to include noncooperative behavior with *incomplete information* where, prior to making their choices, players observe the realization of X, but the realization of ε_p is only privately observed by player p. Players also know the parameters of both players' payoff functions. The assumed solution concept in this case is *Bayesian Nash equilibrium* (BNE), where players maximize their expected payoff given their subjective beliefs, so $Y_1 = \mathbb{1}\{X_1^{ns\prime}\beta_{10} - X_1^{s\prime}\Delta_{10} \cdot \pi_2^e - \varepsilon_1 \geq 0\}$ and $Y_2 = \mathbb{1}\{X_2^{ns\prime}\beta_{20} - X_2^{s\prime}\Delta_{20} \cdot \pi_1^e - \varepsilon_2 \geq 0\}$, where π_1^e denotes player 2's subjective belief for $Pr(Y_1 = 1)$ and π_2^e denotes player 1's subjective belief for $Pr(Y_2 = 1)$. We assume that each player p conditions her beliefs on X, and on Y_p. Thus, π_2^e denotes player 1's subjective belief for $Pr(Y_2 = 1|X, Y_1 = 1)$ and π_1^e denotes player 2's subjective belief for $Pr(Y_1 = 1|X, Y_2 = 1)$. In this particular model we do not assume that players condition their beliefs directly on the realization of their own ε_p. This implies a behavioral model where players do not know (e.g, have not learned) the joint distribution of $(\varepsilon_1, \varepsilon_2)|X$, but know (e.g, have been able to learn) the joint distribution of equilibrium outcomes $(Y_1, Y_2)|X$.

Let us describe BNE beliefs in this model. For a given $\rho \in \Theta$ and $c \in \mathbb{R}$, let $\overline{f}_1(\varepsilon_1; c|\rho) \equiv \frac{\int_{-\infty}^{c} f_{1,2}(\varepsilon_1, \epsilon_2|\rho)d\epsilon_2}{F_2(c)}$ $\overline{f}_2(\varepsilon_2; c|\rho) \equiv \frac{\int_{-\infty}^{c} f_{1,2}(\epsilon_1, \varepsilon_2|\rho)d\epsilon_1}{F_1(c)}$. Note that $\overline{f}_p(\varepsilon_p; c|\rho)$ is the density of ε_p conditional on the event that $\varepsilon_{-p} \leq c$. Next, for $(c_a, c_b) \in \mathbb{R}^2$, let $H_1(c_a; c_b|\rho) \equiv \int_{-\infty}^{c_a} \overline{f}_1(\epsilon_1; c_b|\rho)d\epsilon_1$, $H_2(c_a; c_b|\rho) \equiv \int_{-\infty}^{c_a} \overline{f}_2(\epsilon_2; c_b|\rho)d\epsilon_1$, and for a given (X, θ) and $\pi \equiv (\pi_1, \pi_2) \in [0,1]^2$ let

$$\mathcal{H}(\pi; X, \theta) \equiv \begin{pmatrix} \pi_1 - H_1(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1 \cdot \pi_2 \,;\, X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2 \cdot \pi_1|\rho) \\ \pi_2 - H_2(X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2 \cdot \pi_1 \,;\, X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1 \cdot \pi_2|\rho) \end{pmatrix}$$

BNE beliefs must solve $\mathcal{H}(\pi; X, \theta) = 0$ for π. Continuity implies existence of a solution by Brouwer's Fixed Point Theorem. Regularity and multiplicity can be studied through the properties of $\nabla_\pi \mathcal{H}(\pi; X, \theta)$, the Jacobian of $\mathcal{H}(\pi; X, \theta)$ wrt π. Maintain the following.

Assumption G4 ***(Regularity of BNE)*** *For a.e X and $\forall\ \theta \in \Theta$, the Jacobian $\nabla_\pi \mathcal{H}(\pi; X, \theta)$ is invertible at any π that solves $\mathcal{H}(\pi; X, \theta) = 0$.* ■

Assumption G4 guarantees that every BNE solution is *regular*, and that there is a finite number of BNE solutions[5]. Let $\pi_{(j)}^*(X, \theta) \equiv (\pi_{1(j)}^*(X, \theta), \pi_{2(j)}^*(X, \theta))$ denote the j^{th} solution to the BNE conditions $\mathcal{H}(\pi; X, \theta) = 0$, which we will refer to as the j^{th} BNE for (X, θ). We will let $J(X, \theta)$ denote the total number of BNE for (X, θ). Brouwer's Fixed Point Theorem guarantees that $J(X, \theta) \geq 1$, and regularity guarantees that $J(X, \theta) < \infty$ (see footnote 5). By Gale and Nikaido (1965, Theorem 7), BNE uniqueness can be obtained from properties of the principal minors of $\nabla_\pi \mathcal{H}(\pi; X, \theta)$. Fix (X, θ), then $\mathcal{H}(\pi : X, \theta)$ will be a univalent mapping (of π) from $[0,1]^2$ onto itself if the principal minors of $\nabla_\pi \mathcal{H}(\pi; X, \theta)$ do not vanish for any $\pi \in [0,1]^2$. The Gale Nikaido conditions for our model can be found in Appendix A (equation (A2)). In our BNE model, players' actions are described by,

[5]The Index Theorem (Mas-Colell, Whinston, and Green (1995, Proposition 17.D.2)) can be used to show that there exists an odd number of regular BNE solutions.

$Y_1 = \mathbb{1}\{X_1^{ns\prime}\beta_{10} - X_1^{s\prime}\Delta_{10}\cdot\pi_2^*(X,\theta_0) - \varepsilon_1 \geq 0\}$ and $Y_2 = \mathbb{1}\{X_2^{ns\prime}\beta_{20} - X_2^{s\prime}\Delta_{20}\cdot\pi_1^*(X,\theta_0) - \varepsilon_2 \geq 0\}$, where $(\pi_1^*(X,\theta_0), \pi_2^*(X,\theta_0)) \equiv \pi^*(X,\theta_0)$ is the BNE selected by players. Let $Y_{(j)}^{BNE}$ denote the outcome that would be produced by BNE j, and let $\mathbb{I}_{1(j)}(X,\varepsilon_1,\theta_0) \equiv \mathbb{1}\{X_1^{ns\prime}\beta_{10} - X_1^{s\prime}\Delta_{10}\cdot \pi_{2(j)}^*(X,\theta_0) - \varepsilon_1 \geq 0\}$, and $\mathbb{I}_{2(j)}(X,\varepsilon_2,\theta_0) \equiv \mathbb{1}\{X_2^{ns\prime}\beta_{20} - X_2^{s\prime}\Delta_{20}\cdot\pi_{1(j)}^*(X,\theta_0) - \varepsilon_2 \geq 0\}$. For each $y \equiv (y_1,y_2) \in \mathcal{Y}$, let $\mathbb{I}_{(j)}(y|X,\varepsilon,\theta_0) \equiv \prod_{p=1}^2 \mathbb{I}_{p(j)}(X,\varepsilon_p,\theta_0)^{y_p} \times (1-\mathbb{I}_{p(j)}(X,\varepsilon_p,\theta_0))^{1-y_p}$. Then $\mathbb{1}\{Y_{(j)}^{BNE} = y\} = \mathbb{I}_{(j)}(y|X,\varepsilon,\theta_0)$ in our BNE behavioral model.

3.4 Behavioral and equilibrium selection mechanisms

The behavioral selection mechanism is given by ξ. We denote $\xi = 1$ if players' behavior is cooperative, $\xi = 2$ if players play NE with complete information, and $\xi = 3$ if players play BNE with incomplete information. Whether the informational environment is an act of "nature" or a strategic choice made by players is irrelevant to us as long as the exclusion restrictions described below are satisfied. Note that two of our candidate behavioral models can have multiple solutions for Y. For this reason, we will introduce an equilibrium selection mechanism λ_2 that selects the NE in the multiple-equilibrium region when $\xi = 2$, and an equilibrium selection mechanism λ_3 that selects the BNE when $\xi = 3$. We have $\lambda_2 \in \{1,2,3\}$, where $\lambda_2 = \ell$ indicates that NE ℓ is selected when $\xi = 2$ and $\varepsilon \in \mathcal{M}_1^{NE}(X,\theta_0)$, and $\lambda_3 \in \{1,\ldots,J(X,\theta_0)\}$, where $\lambda_3 = j$ indicates the selection of the BNE $\pi_{(j)}^*(X,\theta_0)$ when $\xi = 3$. For each $y \in \mathcal{Y}$, our global behavioral model yields,

$$\begin{aligned}
\mathbb{1}\{Y = y\} = \mathbb{1}\{\xi = 1\}\cdot\mathbb{1}\{\varepsilon \in \mathcal{R}^C(y|X,\beta_0,\Delta_0)\} \\
+ \mathbb{1}\{\xi = 2\}\cdot\left(\mathbb{1}\{\varepsilon \in \mathcal{R}^{NE}(y|X,\beta_0,\Delta_0)\} + \mathbb{1}\{\varepsilon \in \mathcal{M}_1^{NE}(X,\beta_0,\Delta_0)\}\cdot\sum_{\ell=1}^{3}\mathbb{1}\{\lambda_2 = \ell\}\cdot\mathbb{1}\{Y_{(\ell)}^{NE} = y\}\right) \\
+ \mathbb{1}\{\xi = 3\}\cdot\sum_{j=1}^{J(X,\theta_0)}\mathbb{1}\{\lambda_3 = j\}\cdot\mathbb{I}_{(j)}(y|X,\varepsilon,\theta_0)
\end{aligned} \tag{34}$$

Next we extend our exclusion restriction to our expanded model.

Assumption G5 ***(Exclusion restrictions with cooperative, NE and BNE behavior)*** $\xi \perp \varepsilon$ *and there exists an observable Z such that $\xi|(X,Z) \sim \xi|Z$, with $Pr(\xi = 1|Z) \equiv \pi_1(Z)$, $Pr(\xi = 2|Z) \equiv \pi_2(Z)$ and $Pr(\xi = 3|Z) \equiv \pi_3(Z)$. In addition, Z is such that,*

$$Pr(\lambda_2 = \ell|\xi = 2,\varepsilon,X,Z) = Pr(\lambda_2 = \ell|\xi = 2,Z) \equiv \omega_2(\ell|Z) \ \ \text{for } \ell = 1,2,3$$
$$Pr(\lambda_3 = j|\xi = 3,\varepsilon,X,Z) = Pr(\lambda_3 = j|\xi = 3,Z) \equiv \omega_3(j|Z) \ \ \text{for } j = 1,\ldots,J(X,\theta_0)$$

We will leave $(\pi_b(Z))_{b=1}^3$ and $(\omega_2(\ell|Z))_{\ell=1}^3$ and $(\omega_3(j|Z))_{j=1}^{J(X,\theta_0)}$ nonparametrically specified. ■

The restriction $\xi \perp \varepsilon$ can be relaxed to $\xi \perp \varepsilon | (X, Z)$ if we parameterize the conditional distribution of $\varepsilon | (X, Z)$. Since $J(X, \theta_0)$, the number of BNE, can depend on X, our exclusion restriction for λ_3 implicitly requires that, either $J(X, \theta_0)$ be constant across X, *or* that λ_3 be degenerate and *not* randomize across BNE. Since the number of complete-information NE (three) in the multiple equilibrium region is independent of X, the exclusion restriction for λ_2 can be satisfied regardless of whether λ_2 randomizes across the multiple NE.

Remark 3 ***(Inference range $\mathcal{Z}$ and Assumption G5)*** *As we did previously, we will ultimately choose an inference range $\mathcal{Z}$ for Z (see Section 3.5.2 below), and the exclusion restrictions in Assumption G5 will only need to hold over $\mathcal{Z}$. Furthermore, since the Gale Nikaido conditions are directly verifiable for any (X, θ) (see Appendix A ,equation A2), $\mathcal{Z}$ can be potentially chosen so that the BNE is unique over our inference range, $\forall\ \theta \in \Theta$, eliminating the need to make assumptions about the BNE equilibrium selection mechanism λ_3.*

For our cooperative behavioral model let,

$$g_{\mathcal{R}}^{C}(y|X, \theta_0) \equiv \Pr\left(\varepsilon \in \mathcal{R}^{C}(y|X, \beta_0, \Delta_0)\middle|X\right) = \int \mathbb{1}\left\{\varepsilon \in \mathcal{R}^{C}(y|X, \beta_0, \Delta_0)\right\} f_{1,2}(\varepsilon|\rho_0) d\varepsilon, \tag{35}$$

Next, for our complete-information NE behavioral model denote,

$$\begin{aligned} g_{\mathcal{R}}^{NE}(y|X, \theta_0) &\equiv \Pr\left(\varepsilon \in \mathcal{R}^{NE}(y|X, \beta_0, \Delta_0)\middle|X\right) = \int \mathbb{1}\left\{\varepsilon \in \mathcal{R}^{NE}(y|X, \beta_0, \Delta_0)\right\} f_{1,2}(\varepsilon|\rho_0) d\varepsilon, \\ G_{\mathcal{M}_1}(X, \theta_0) &\equiv \Pr\left(\varepsilon \in \mathcal{M}_1^{NE}(X, \beta_0, \Delta_0)\middle|X\right) = \int \mathbb{1}\left\{\varepsilon \in \mathcal{M}_1^{NE}(X, \beta_0, \Delta_0)\right\} f_{1,2}(\varepsilon|\rho_0) d\varepsilon, \\ g_{\mathcal{M}_1,NE}^{\ell}(y|X, \theta_0) &\equiv \int \sigma_{\mathcal{M}_1,NE}^{\ell}(y|X, \varepsilon, \beta_0, \Delta_0) \cdot \mathbb{1}\left\{\varepsilon \in \mathcal{M}_1^{NE}(X, \beta_0, \Delta_0)\right\} f_{1,2}(\varepsilon|\rho_0) d\varepsilon. \end{aligned} \tag{36}$$

Finally, for our incomplete-information BNE behavioral model, let

$$g_{BNE}^{j}(y|X, \theta_0) \equiv E\left[\mathbb{I}_{(j)}(y|X, \varepsilon, \theta_0)|X\right] = \int \mathbb{I}_{(j)}(y|X, \varepsilon, \theta_0) f_{1,2}(\varepsilon|\rho_0) d\varepsilon. \tag{37}$$

Denote $\vartheta_2(\ell|Z) \equiv \pi_2(Z) \cdot \omega_2(\ell|Z)$ and $\vartheta_3(j|Z) \equiv \pi_3(Z) \cdot \omega_3(j|Z)$. Let $p_Y(y|U) \equiv Pr(Y = y|U)$. Using (34) and Assumption G5,

$$\begin{aligned} p_Y(y|U) = \pi_1(Z) \cdot g_{\mathcal{R}}^{C}(y|X, \theta_0) + \pi_2(Z) \cdot g_{\mathcal{R}}^{NE}(y|X, \theta_0) + \sum_{\ell=1}^{3} \vartheta_2(\ell|Z) \cdot g_{\mathcal{M}_1,NE}^{\ell}(y|X, \theta_0) \\ + \sum_{j=1}^{J(X,\theta_0)} \vartheta_3(j|Z) \cdot g_{BNE}^{j}(y|X, \theta_0) \end{aligned}$$

Since $\sum_{\ell=1}^{3}\omega_2(\ell|Z) = \sum_{j=1}^{J(X,\theta_0)}\omega_3(j|Z) = \sum_{b=1}^{3}\pi(Z) = 1$, the above expression becomes

$$p_Y(y|U) = m_1(y|X,\theta_0) + \delta(Z)'\Xi(y|X,\theta_0)\ \forall\ y \in \mathcal{Y}, \tag{38}$$

where

$$\begin{aligned}
\delta(Z) &\equiv \Big(\pi_1(Z)\ ,\ \pi_2(Z)\ ,\ (\vartheta_2(\ell|Z))_{\ell=2}^{3}\ ,\ (\vartheta_3(j|Z))_{j=2}^{J(X,\theta_0)}\Big)',\\
m_1(y|X,\theta_0) &\equiv g_{BNE}^{1}(y|X,\theta_0)\\
m_2(y|X,\theta_0) &\equiv \Big(g_{\mathcal{R}}^{C}(y|X,\theta_0) - g_{BNE}^{1}(y|X,\theta_0)\ ,\ g_{\mathcal{R}}^{NE}(y|X,\theta_0) + g_{\mathcal{M}_1,NE}^{1}(y|X,\theta_0) - g_{BNE}^{1}(y|X,\theta_0)\Big)'\\
m_3(y|X,\theta_0) &\equiv \Big(\Big(g_{\mathcal{M}_1,NE}^{\ell}(y|X,\theta_0) - g_{\mathcal{M}_1,NE}^{1}(y|X,\theta_0)\Big)_{\ell=2}^{3}\ ,\ \Big(g_{BNE}^{j}(y|X,\theta_0) - g_{BNE}^{1}(y|X,\theta_0)\Big)_{j=2}^{J(X,\theta_0)}\Big)'\\
\Xi(y|X,\theta_0) &\equiv \Big(m_2(y|X,\theta_0)'\ ,\ m_3(y|X,\theta_0)'\Big)'
\end{aligned} \tag{39}$$

Once again, the predictions of our global model are a semiparametric convolution of the predictions of each candidate behavioral model.

3.5 Identifiability and estimation of θ_0 in our expanded model

As in the simpler version of our global model, identifiability of θ_0 depends on whether we assume that players display noncooperative behavior with nonzero probability. If $\pi_1(Z) = 1$ w.p.1 and players cooperate almost surely, $p_Y(y|U) = g_{\mathcal{R}}^{C}(y|X,\theta)$, and our previous arguments show that Δ_0 is not identifiable unless we assume that X_1^s and X_2^s have no elements in common or unless we impose restrictions on $(\Delta_{10},\Delta_{20})$ such as symmetry. Let us focus first on the case where we maintain that players display noncooperative behavior with strictly positive probability.

3.5.1 Identifiability of θ_0 under the assumption that $Pr(\pi_1(Z) < 1) > 0$

The following restriction is the version of Assumption I2 for our expanded model.

Assumption I5 ***(Strictly positive probability of noncooperative behavior)*** $Pr(\pi_1(Z) < 1) > 0$, *so players display noncooperative behavior with strictly positive probability.*

Denote $D(y) \equiv \mathbb{1}\{Y = y\}$. From (38), we can express

$$D(y) = m_1(y|X,\theta_0) + \delta(Z)'\Xi(y|X,\theta_0) + \varepsilon(y), \quad \text{where} \quad E[\varepsilon(y)|U] = 0. \tag{40}$$

In particular, $E[\Xi(y|X,\theta_0)\cdot\varepsilon(y)|Z] = 0$ a.e Z. Thus,

$$E[\Xi(y|X,\theta_0)\cdot\Xi(y|X,\theta_0)'|Z]\cdot\delta(Z) = E[\Xi(y|X,\theta_0)\cdot(D(y) - m_1(y|X,\theta_0))|\,Z] \ \text{ a.e } Z. \tag{41}$$

Again, we propose to exploit the convolution structure in (40) by constructing first a semiparametric estimator for $\delta(Z)$. As in Section 2.5, identification of $\delta(Z)$ requires ruling out that any two of our candidate behavioral models are observationally equivalent. From (41), this boils down to the following full-rank restriction for $\Xi(y|X,\theta)$.

Assumption I6 *The presence of strategic-interaction effects restriction in Assumption G3 is maintained for the parameter space* Θ*, and there exists at least one* $y \in \mathcal{Y}$ *such that* $E[\Xi(y|X,\theta)\cdot\Xi(y|X,\theta)'|Z]$ *is invertible for every* $\theta \in \Theta$ *and a.e* Z*.*

Assumption I6 requires the existence of an outcome $y \in \mathcal{Y}$ such that, for any value θ_0 can take over Θ, each of our candidate behavioral models produce different predictions for $Pr(Y = y|X)$. In Appendix A (Section A3) we show that this is satisfied by $y \in \{(1,0)\ ,(0,1)\}$ in our setting. Let $\overline{y} \in \{(1,0),(0,1)\}$ satisfy the full-rank restriction in Assumption I6. We can express, $\delta(Z) = E[\Xi(\overline{y}|X,\theta_0)\cdot \Xi(\overline{y}|X,\theta_0)'|Z]^{-1}\cdot E[\Xi(\overline{y}|X,\theta_0)\cdot(D(\overline{y}) - m_1(\overline{y}|X,\theta_0))|\,Z]$. For each $\theta \in \Theta$, let

$$\delta(Z,\theta) \equiv E[\Xi(\overline{y}|X,\theta)\cdot\Xi(\overline{y}|X,\theta)'|Z]^{-1}\cdot E[\Xi(\overline{y}|X,\theta)\cdot(D(\overline{y}) - m_1(\overline{y}|X,\theta))|\,Z] \tag{42}$$

Note that $\delta(Z,\theta_0) = \delta(Z)$. Let,

$$\mathbb{P}(y|U,\theta) = m(y|X,\theta) + \delta(Z,\theta)'\Xi(y|X,\theta). \tag{43}$$

From (38),

$$p_Y(y|U) = \mathbb{P}(y|U,\theta_0)\ \forall\ y \in \mathcal{Y},\ \text{ a.e } U. \tag{44}$$

Since $\sum_{y\in\mathcal{Y}} p_Y(y|U) = 1$ and $\sum_{y\in\mathcal{Y}} \mathbb{P}(y|U,\theta) = 1\ \forall\ \theta$, we focus on a subset of three outcomes in $\mathcal{Y}$. Take $\mathcal{Y}^* \equiv \{(1,0),(0,1),(1,1)\}$. θ is observationally equivalent to θ_0 if $\mathbb{P}(y|U,\theta_0) = \mathbb{P}(y|U,\theta)$ w.p.1 for each $y \in \mathcal{Y}^*$. Therefore, θ_0 is identifiable from (44) if, for some $y \in \mathcal{Y}^*$, $Pr(\mathbb{P}(y|U,\theta) \neq \mathbb{P}(y|U,\theta_0)) > 0\ \forall\ \theta \in \Theta$: $\theta \neq \theta_0$, and θ_0 is locally identifiable from (44) if there exists a neighborhood $\mathcal{A} \subseteq \Theta$ of θ_0 such that this condition holds $\forall\ \theta \in \mathcal{A}$: $\theta \neq \theta_0$. As we did previously, we can apply a conditional GMM approach based on Dominguez and Lobato (2004). As before, group $V \equiv (Y,U)$. For each $(y,\theta,u) \in \mathcal{Y}^* \times \Theta \times \mathbb{R}^{d_U}$, let

$$\varphi(y|V,\theta) \equiv D(y) - \mathbb{P}(y|U,\theta) \quad \text{and} \quad T(y|u,\theta) \equiv E[\varphi(y|V,\theta)\cdot \mathbb{1}\{U \leq u\}]. \tag{45}$$

From (40) and iterated expectations, $T(y|u,\theta_0) = 0\ \forall\ u \in \mathbb{R}^{d_U}$, $y \in \mathcal{Y}^*$. And, by Dominguez and Lobato (2004, equation 2), $E[\varphi(y|V,\theta)|U] = 0$ a.e $U \Leftrightarrow T(y|u,\theta) = 0$ for F_U–a.e $u \in \mathbb{R}^{d_U}$ (a result that follows from Billingsley (1995, Theorem 16.10iii)). Let

$$Q_Y(y|\theta) \equiv \frac{1}{2}\int T(y|u,\theta)^2 dF_U(u) = \frac{1}{2}E[T(y|U,\theta)^2].$$

$Q_Y(y|\theta) \geq 0\ \forall\ \theta$ and, from the previous arguments, $Q_Y(y|\theta) = 0 \Leftrightarrow T(y|u,\theta) = 0\ F_U$–a.e $u \in \mathbb{R}^{d_U}$. We can aggregate over $y \in \mathcal{Y}^*$ through the population statistic,

$$Q_Y(\theta) \equiv \sum_{y\in\mathcal{Y}^*} Q_Y(y|\theta). \tag{46}$$

We can then characterize θ_0 as a minimizer of $Q_Y(\theta)$ over $\theta \in \Theta$. The following restrictions are sufficient for local identification of θ_0.

Assumption I7 *The true parameter value θ_0 belongs in the interior of the parameter space Θ. In addition, the following conditions hold.*

(i) For each $p = 1,2$, the support of X_p^{ns} is not contained in any proper linear subspace of $\mathbb{R}^{d_{X_p^{ns}}}$, where $d_{X_p^{ns}} \equiv dim(X_p^{ns})$. For $p,q = \{1,2\}$, there exists a component $X_{p,j}^{ns} \in W_p$ such that $X_{p,j}^{ns} \notin W_q$.

(ii) For each $p = 1,2$, the support of X_p^{s} is not contained in any proper linear subspace of $\mathbb{R}^{d_{X_p^{s}}}$, where $d_{X_p^{s}} \equiv dim(X_p^{s})$.

(iii) There exists an open neighborhood $\mathcal{A} \subseteq \Theta$ of θ_0 such that $\frac{\partial^2 Q_Y(\theta)}{\partial\theta\partial\theta'}$ is invertible $\forall\ \theta \in \mathcal{A}$.

Under the conditions of Assumption I7, θ_0 is locally identifiable from (44) since θ_0 is a locally unique minimizer of $Q_Y(\theta)$.

3.5.2 An estimator for θ_0 under the assumption that players display noncooperative behavior with strictly positive probability

As before, we propose an estimator that minimizes a sample analog of $Q_Y(\theta)$ using a semiparametric estimator for $\delta(Z,\theta)$. For each (z,θ) define,

$$\widehat{\delta}(z,\theta) \equiv \left(\sum_{i=1}^{n} \Xi(\overline{y}|X_i,\theta)\cdot\Xi(\overline{y}|X_i,\theta)' K\left(\frac{Z_i - z}{h_n}\right)\right)^{-1} \cdot \sum_{i=1}^{n} \Xi(\overline{y}|X_i,\theta)\cdot\left(D_i(\overline{y}) - m_1(\overline{y}|X_i,\theta)\right) K\left(\frac{Z_i - z}{h_n}\right),$$
$$\widehat{\mathbb{P}}(y|U,\theta) \equiv m(y|X,\theta) + \widehat{\delta}(Z,\theta)'\Xi(y|X,\theta), \quad \widehat{\varphi}(y|V,\theta) \equiv D(y) - \widehat{\mathbb{P}}(y|U,\theta). \tag{47}$$

As before, we will pre-specify an inference range $\mathcal{Z} \subseteq \text{Supp}(Z)$ where our weights satisfy uniform asymptotic properties. Since the Gale-Nikaido conditions (see Appendix A, equation (A2)) can be directly verifiable for any (X,θ), the inference range $\mathcal{Z}$ can potentially be chosen in a way that guarantees a unique BNE for every $\theta \in \Theta$ everywhere on our inference range, allowing us to bypass any assumptions involving the BNE selection mechanism λ_3.

An estimator for θ_0

As before, for any $u \in \mathbb{R}^{d_U}$, let $\mathbb{1}_{\mathcal{Z}}\{U \leq u\} \equiv \mathbb{1}\{U \leq u\ ,\ Z \in \mathcal{Z}\}$. We will define $T_{\mathcal{Z}}(y|u,\theta) \equiv E[\varphi(y|V,\theta)\cdot \mathbb{1}_{\mathcal{Z}}\{U \leq u\}]$, the version of $T(y|u,\theta)$ where Z is restricted to $\mathcal{Z}$. Next, let

$$Q_{Y,\mathcal{Z}}(y|\theta) \equiv \frac{1}{2}\int T_{\mathcal{Z}}(y|u,\theta)^2 dF_U(u) = \frac{1}{2}E[T_{\mathcal{Z}}(y|U,\theta)^2], \quad \text{and} \quad Q_{Y,\mathcal{Z}}(\theta) \equiv \sum_{y\in\mathcal{Y}^*} Q_{Y,\mathcal{Z}}(y|\theta). \tag{48}$$

$Q_{Y,\mathcal{Z}}(\theta)$ is our population statistic restricted to $\mathcal{Z}$. Our sample objective function is,

$$\widehat{Q}_{Y,\mathcal{Z}}(\theta) \equiv \sum_{y \in \mathcal{Y}^*} \widehat{Q}_{Y,\mathcal{Z}}(y|\theta), \quad \text{where}$$

$$\widehat{Q}_{Y,\mathcal{Z}}(y|\theta) \equiv \frac{1}{2} \cdot \frac{1}{n} \sum_{j=1}^{n} \widehat{T}_{\mathcal{Z}}(y|U_j,\theta)^2 = \frac{1}{2} \cdot \frac{1}{n} \sum_{j=1}^{n} \left[\frac{1}{n} \sum_{i=1}^{n} \widehat{\varphi}(y|V_i,\theta) \cdot \mathbb{1}_{\mathcal{Z}}\{U_i \le U_j\} \right]^2, \tag{49}$$

and $\widehat{\varphi}(y|V_i,\theta) \equiv D_i(y) - \widehat{\mathbb{P}}(y|U_i,\theta)$. Our proposed estimator for θ_0 is $\widehat{\theta} = \underset{\theta \in \Theta}{argmin}\, \widehat{Q}_{Y,\mathcal{Z}}(\theta)$. Using standard arguments, the restrictions in Assumptions E1, G4 -G5, I5 -I7, and the appropriate modification of Assumptions E2 and E3, our estimator $\widehat{\theta}$ can be shown to be $\sqrt{n}$–consistent and asymptotically normal. The details of its asymptotic distribution can be found in the online Econometric Supplement (Section S3).

3.5.3 Estimation of behavioral and equilibrium selection probabilities $\delta(z)$

Under the previous restrictions, we can estimate $\delta(z)$ through $\widehat{\delta}(z,\widehat{\theta})$, with asymptotic properties analogous to those described in Section 2.6.1 for $\widehat{\pi}(z,\widehat{\theta})$. Equipped with $\widehat{\delta}(z,\widehat{\theta})$ we can learn about the propensity of behavior selection as well as equilibrium selection within each behavioral model.

3.6 A consistent specification test

Equipped with $\widehat{\theta}$ we can perform a specification test for the validity of any proposed instrument Z, and the assumptions of our model through a consistent test for the null hypothesis that (38) is satisfied w.p.1 for all $y \in \mathcal{Y}$. This can be done by adapting the procedure outlined in Section 2.6.2.

3.7 Inference for θ_0 when we allow for the possibility that players cooperate almost surely

If we are unwilling to assume that players behave noncooperatively with nonzero probability, our recommendation is to construct a CS for θ_0 based on (44). While there are multiple ways to proceed, we outline an approach that follows on the steps of Section 2.7. Take any $y \in \mathcal{Y}$ and let $g_y : \mathbb{R}^{d_U} \to \mathbb{R}$ be a real-valued, pre-specified function of U. For a given $u \in \mathbb{R}^{d_U}$, let

$$m_g(y|u,\theta) \equiv E\Big[\varphi(y|V,\theta) \cdot g_y(U) \cdot \mathbb{1}_{\mathcal{Z}}\{U \le u\}\Big] \quad \text{and} \quad M_g(y|\theta) \equiv E[m_g(y|U,\theta)]. \tag{50}$$

(44) implies $M_g(y|\theta_0)$ $\forall$ $y \in \mathcal{Y}^*$ and a CS for θ_0 can be constructed from here. Next, let $\Theta_g^I \equiv \Big\{\theta \in \Theta : M_g(y|\theta) = 0\ \forall\ y \in \mathcal{Y}^*\Big\}$ be our target identified set for θ based on (44). Let

$$\widehat{M}_g(y|\theta) = \frac{1}{n} \sum_{j=1}^{n} \widehat{m}_g(y|U_j,\theta) = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{i=1}^{n} \widehat{\varphi}(y|V_i,\theta) \cdot g_y(U_i) \cdot \mathbb{1}_{\mathcal{Z}}\{U_i \le U_j\} \tag{51}$$

Following standard arguments, in the online Econometric Supplement (Section S4) we show that, under Assumptions E1, G4 -G5 and the appropriate modification of Assumptions E2 and E3, the statistic $\widehat{M}_g(y|\theta)$ satisfies a linear representation result analogous to Proposition 2, so for each $y \in \mathcal{Y}$,

$$\widehat{M}_g(y|\theta) = M_g(y|\theta) + \frac{1}{n}\sum_{i=1}^{n} \psi_{M_g,n}(y|V_i,\theta) + \varsigma_n^{M_g}(y|\theta),\ \forall\ \theta \in \Theta,$$
$$\text{where} \quad E[\psi_{M_g,n}(y|V_i,\theta)] = 0\ \forall\ \theta \in \Theta, \quad \sup_{\theta\in\Theta}\left|\varsigma_n^{M_g}(y|\theta)\right| = o_p\left(n^{-1/2}\right)$$

Using this result, we show that a CS for θ_0 with pre-specified asymptotic target coverage probability can be constructed based on a Wald-type statistic that aggregates $\widehat{M}_g(y|\theta)$ across $y \in \mathcal{Y}^*$.

Inference for the behavioral and equilibrium selection probabilities $\delta(z)$

We can construct a CS for $\delta(z)$ by relying on the asymptotic properties of $\widehat{\delta}(z,\theta)$ and projecting our CS for θ_0 on to $\widehat{\delta}(z,\theta)$, proceeding along the lines described in Section 2.7.1.

4 A general model

Here we extend our setup beyond a 2×2 game and consider a general discrete game with a collection of candidate behavioral models, each with possible multiple solutions.

4.1 A parameterized normal-form game

We have a collection of players, $p = 1,\ldots,\mathcal{P}$, where each p has a discrete and finite action space $\mathcal{Y}_p$ and a (parametric) payoff function $u_p(y;X_p,\varepsilon_p,\beta_{p_0})$, where $y \equiv (y_1,\ldots,y_{\mathcal{P}})$ is a particular action profile, $X_p \in \mathbb{R}^{d_{X_p}}$ and $\varepsilon_p \in \mathbb{R}$ denote observable and unobservable (to the econometrician) payoff shifters and where β_{p_0} is a finite-dimensional parameter. We denote the outcome of the game as $Y \equiv (Y_1,\ldots,Y_{\mathcal{P}}) \in \mathcal{Y}$, where $\mathcal{Y} \equiv \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_{\mathcal{P}}$. Group $X \equiv \cup_{p=1}^{\mathcal{P}} X_p \in \mathbb{R}^{d_X}$ and $\varepsilon \equiv (\varepsilon_1,\ldots,\varepsilon_{\mathcal{P}}) \in \mathbb{R}^{\mathcal{P}}$. We maintain that $\varepsilon \perp X$ and $\varepsilon \sim F_\varepsilon(\cdot|\rho_0)$, a parametric joint distribution[6] indexed by a finite dimensional parameter ρ_0. The parameters of the model are $\theta_0 \equiv (\beta_{1_0},\ldots,\beta_{\mathcal{P}_0}, \rho_0) \in \Theta$.

4.2 A collection of candidate behavioral models

We have a collection of $\mathcal{B}$ *candidate behavioral models*, labeled $b = 1,\ldots,\mathcal{B}$. The outcome of each observation of the game is the realization of a *solution* to one of the candidate behavioral models. The econometrician does not know the behavioral model that produced each observation, and different observations in the data could have been produced by different behavioral models.

4.2.1 Solutions for each behavioral model

For any given realization of X, each behavioral model b satisfies the following.

• $\mathbb{R}^{\mathcal{P}}$ can be partitioned into $\overline{R}_b(X,\theta_0)$ regions of realizations of ε. These regions are mutually exclusive and their union covers $\mathbb{R}^{\mathcal{P}}$. We will denote each region as $\mathcal{R}_{b,r_b}(X,\theta_0)$, with $r_b = 1,\ldots,\overline{R}_b(X,\theta_0)$.

[6] We can relax the assumption that $\varepsilon \perp X$ and parameterize the conditional distribution of $\varepsilon|X$.

Each region $\mathcal{R}_{b,r_b}(X,\theta_0)$ can be characterized parametrically given our normal-form parameterization.

- If $\varepsilon \in \mathcal{R}_{b,r_b}(X,\theta_0)$, the behavioral model has $\overline{S}_{b,r_b}(X,\theta_0) \geq 1$ solutions, which we will index as $s_{r_b} = 1,\ldots,\overline{S}_{b,r_b}(X,\theta_0)$. Let $Y_{(s_{r_b})}$ denote the (potential) outcome of solution s_{r_b}. We will let $\sigma^{s_{r_b}}_{b,r_b}(Y_{(s_{r_b})}|X,\varepsilon,\theta_0)$ denote the distribution of $Y_{(s_{r_b})}$ conditional on (X,ε). Each $\sigma^{s_{r_b}}_{b,r_b}(\cdot|X,\varepsilon,\theta_0)$ can be characterized parametrically given our normal-form parameterization.

4.3 Behavior-selection and solution-selection mechanisms

We construct a global model by including a behavior selection mechanism ξ and a solution selection mechanism. We have $\xi \in \{1,\ldots,\mathcal{B}\}$, where $\xi = b$ indicates that behavioral model b has been selected. If b is selected as the behavioral model, and $\varepsilon \in \mathcal{R}_{b,r_b}(X,\theta_0)$, a solution selection mechanism λ_{b,r_b} selects one among the existing solutions inside region $\mathcal{R}_{b,r_b}(X,\theta_0)$. Thus, $\lambda_{b,r_b} \in \{1,\ldots,\overline{S}_{b,r_b}(X,\theta_0)\}$, and $\lambda_{b,r_b} = s_{r_b}$ indicates that solution s_{r_b} has been selected. The precise way through which behavior and solutions are selected (e.g, whether "nature" is involved) is left unspecified as long as the exclusion restrictions described below are satisfied.

4.3.1 Exclusion restrictions

We assume that $\varepsilon \perp \xi$ and that $\exists\ Z$ (observable) such that[7], $\forall\ b,\ s_{r_b}$: $Pr(\xi = b|X,Z) = Pr(\xi = b|Z) \equiv \pi_b(Z)$ and, $Pr(\lambda_{b,r_b} = s_{r_b}|\xi = b,\varepsilon,X,Z) = Pr(\lambda_{b,r_b} = s_{r_b}|\xi = b,Z) \equiv \omega_{b,r_b}(s_{r_b}|Z)$.

4.4 An expression for $Pr(Y = y|U)$

Grouping $U \equiv X \cup Z$ and letting $p_Y(y|U) \equiv Pr(Y = y|U)$, we show in Appendix A (Section A4) that, under the conditions described, our general model yields,

$$p_Y(Y = y|U) = m_1(y|X,\theta_0) + \delta(Z)'\Xi(y|X,\theta_0), \tag{52}$$

where $m_1(y|X,\theta_0)$ and $\Xi(y|X,\theta_0)$ are parametric functions and $\delta(Z)$ are nonparametric weights, all of which are described in Appendix A. This expression generalizes the semiparametric behavioral convolution properties of our previous models.

4.4.1 Estimation and inference for θ_0

Identification and inference can be based on the semiparametric convolution in (52). Identification of the weights $\delta(Z)$ requires ruling out that any pair of our candidate behavioral models are observationally equivalent to each other. More precisely, it requires the existence of an action profile $y \in \mathcal{Y}$ such that each one of our behavioral models produces a different prediction for $Pr(Y = y|X)$ for any possible value that θ_0 can take in Θ. As before, this boils down to a full-rank condition, where $\exists\ y \in \mathcal{Y}$ such that $E[\Xi(y|X,\theta) \cdot \Xi(y|X,\theta)'|Z]$ is invertible for every $\theta \in \Theta$ and a.e Z. From

[7] We can relax the restriction $\varepsilon \perp \xi$ to $\varepsilon \perp \xi|(X,Z)$ if we parameterize the conditional distribution of $\varepsilon|(X,Z)$.

here we can characterize a functional $\delta(Z,\theta)$ such that that $\delta(Z) = \delta(Z,\theta_0)$, and a conditional moment restriction would then follow from (52). Identifiability of θ_0 can follow from the maintained assumption that the underlying selection mechanisms assign nonzero weight to at least one behavioral model that is capable of identifying θ_0 (in our previous examples, this amounted to assigning nonzero probability to noncooperative behavior). A conditional GMM estimator for θ_0 can then be constructed using the approach we described before. From here, estimation of the selection weights $\delta(Z)$, and consistent-specification tests of our model can then be pursued along the lines described in our previous sections. If we are unwilling to impose restrictions on $\delta(Z)$ that can lead to point-identification of θ_0, a CS for θ_0 can be based on the restriction in (52) in the manner outlined in Sections 2.7 and 3.7.

5 A Monte Carlo study

Here we study the performance of the estimator proposed in Section 2.6 in a Monte Carlo experiment. Take three observable, scalar payoff covariates, (W_1, W_2, Z), distributed as $W_1, W_2, Z \overset{i.i.d}{\sim} \mathcal{N}(0,1)$. We parameterize payoffs as described in Tables 1S and 2S, using $t_1 = \beta_{10}^C + \beta_{10}^W W_1 + \beta_{10}^Z Z$ and $t_2 = \beta_{20}^C + \beta_{20}^W W_2 + \beta_{20}^Z Z$, and constant strategic-interaction effects $(\Delta_{10}, \Delta_{20})$. The unobserved payoff shifters $(\varepsilon_1, \varepsilon_2)$ are bivariate standard Normal with correlation coefficient ρ_0, independent of (W_1, W_2, Z). We set $\beta_{10}^C = \beta_{20}^C = 0.5$, $\beta_{10}^W = \beta_{20}^W = -1$, $\beta_{10}^Z = \beta_{20}^Z = 0.5$, and $\rho_0 = 0.5$. The strategic-interaction parameters are $\Delta_{10} = 0.5$ and $\Delta_{20} = 1$. As in Section 2.6, players have two types of possible behavior: cooperative and PSNE. Players cooperate iff $Z + \eta \leq c$, where $\eta \sim \mathcal{N}(0,1)$ is a "signal" that is independent of all other payoff shifters, and c is a threshold. Our behavioral selection mechanism satisfies the exclusion restrictions in Assumption G2, with $Pr(\text{Cooperation}|Z) \equiv \pi(Z) = E[\mathbb{1}\{\eta \leq c - Z\}|Z] = \Phi(c - Z)$, which is *decreasing* in Z. We set $c = \sqrt{2} \cdot \Phi^{-1}(1/3)$, so the aggregate (i.e, unconditional) probability of cooperation is $Pr(Cooperation) \equiv \pi = E[\mathbb{1}\{Z + \eta \leq c\}] = \Phi(\sqrt{2} \cdot \Phi^{-1}(1/3)/\sqrt{2}) = 1/3$.

5.1 Estimation results for $\widehat{\theta}$

We applied the estimation procedure in Section 2.6 for 2,000 samples of sizes $n = 500$, $n = 1{,}000$ and $n = 2{,}000$. The non-strategic parameters $\gamma \equiv (\beta_1^C, \beta_1^W, \beta_1^Z, \beta_2^C, \beta_2^W, \beta_2^Z, \rho)$ were estimated using the MLE procedure in Section 2.5.1, and the strategic-interaction parameters $\Delta \equiv (\Delta_1, \Delta_2)$ were estimated using the conditional GMM procedure described in Section 2.6, using a Gaussian kernel for Z and a bandwidth of the form $h_n = c_h \cdot \widehat{\sigma}(Z) \cdot n^{-1/5}$. Let $\widehat{f_Z}(z)$ denote the kernel-estimator for $f_Z(z)$ and let $\widehat{f}_{Z,\alpha}$ and $\widehat{\tau}_{Z,\alpha}$ denote the α^{th} sample quantiles of $(\widehat{f_Z}(Z_i))_{i=1}^n$ and $(Z_i)_{i=1}^n$, respectively. Our conditional-GMM inference range was $\mathcal{Z} = \left\{z \in \mathbb{R} : \widehat{\tau}_{Z,0.005} \leq z \leq \widehat{\tau}_{Z,0.995}\ ,\ \widehat{f_Z}(z) \geq \widehat{f}_{Z,0.005}\right\}$. Table 1 summarizes our estimation results for $c_h = 1$, which is close to the so-called "rule of thumb" choice. Our estimators perform well, and they estimate both the non-strategic and the strategic parameters with reasonable precision for all sample sizes analyzed. The Empirical Supplement includes estimation results for $c_h = 0.80$ and $c_h = 1.40$. Even though we do not present a general theory of bandwidth selection, our estimation results were robust across our bandwidth choices.

Table 1: Monte Carlo results for θ.

$n = 500$							
Parameter	**True value**	**Element-wise quantiles of $\widehat{\theta}$ across our simulations**					
		0.15^{th} quantile	**0.25^{th} quantile**	**median**	**0.75^{th} quantile**	**0.85^{th} quantile**	**median bias**
β_1^c	0.5	0.237	0.346	0.525	0.674	0.741	0.165
β_1^w	-1	-1.326	-1.196	-1.026	-0.876	-0.813	0.155
β_1^z	0.5	0.343	0.401	0.515	0.634	0.708	0.116
β_2^c	0.5	0.226	0.351	0.530	0.669	0.724	0.160
β_2^w	-1	-1.346	-1.212	-1.024	-0.876	-0.812	0.161
β_2^z	0.5	0.343	0.408	0.518	0.635	0.701	0.113
ρ	0.5	0.114	0.221	0.561	0.837	0.947	0.310
Δ_1	0.5	0.218	0.337	0.621	0.922	1.073	0.283
Δ_2	1	0.437	0.682	1.048	1.345	1.499	0.334

$n = 1,000$							
Parameter	**True value**	**Element-wise quantiles of $\widehat{\theta}$ across our simulations**					
		0.15^{th} quantile	**0.25^{th} quantile**	**median**	**0.75^{th} quantile**	**0.85^{th} quantile**	**median bias**
β_1^c	0.5	0.311	0.393	0.510	0.617	0.672	0.112
β_1^w	-1	-1.200	-1.129	-1.007	-0.905	-0.855	0.111
β_1^z	0.5	0.384	0.429	0.509	0.589	0.640	0.080
β_2^c	0.5	0.298	0.380	0.500	0.604	0.659	0.110
β_2^w	-1	-1.211	-1.135	-1.017	-0.908	-0.857	0.111
β_2^z	0.5	0.377	0.422	0.501	0.581	0.627	0.079
ρ	0.5	0.165	0.279	0.507	0.700	0.795	0.207
Δ_1	0.5	0.213	0.333	0.556	0.801	0.934	0.231
Δ_2	1	0.611	0.795	1.049	1.271	1.388	0.245

$n = 2,000$							
Parameter	**True value**	**Element-wise quantiles of $\widehat{\theta}$ across our simulations**					
		0.15^{th} quantile	**0.25^{th} quantile**	**median**	**0.75^{th} quantile**	**0.85^{th} quantile**	**median bias**
β_1^c	0.5	0.368	0.421	0.503	0.578	0.622	0.079
β_1^w	-1	-1.120	-1.083	-1.002	-0.933	-0.899	0.074
β_1^z	0.5	0.415	0.448	0.506	0.559	0.593	0.055
β_2^c	0.5	0.377	0.424	0.507	0.578	0.617	0.078
β_2^w	-1	-1.135	-1.087	-1.009	-0.936	-0.899	0.075
β_2^z	0.5	0.419	0.447	0.502	0.556	0.589	0.054
ρ	0.5	0.280	0.364	0.509	0.643	0.711	0.141
Δ_1	0.5	0.257	0.347	0.557	0.745	0.850	0.198
Δ_2	1	0.771	0.888	1.059	1.208	1.291	0.172

• 2,000 simulations, $c_h = 1.0$, Gaussian kernel

5.1.1 Testing the null hypothesis of asymmetric strategic-interaction effects

Next we evaluate the inferential performance of our approach by testing for asymmetries in interaction effects. First, we test $H_0^a : \Delta_{20} \geq \Delta_{10}$ against $H_1^a : \Delta_{20} < \Delta_{10}$, and then we test $H_0^b : \Delta_{20} \leq \Delta_{10}$ against $H_1^b : \Delta_{20} > \Delta_{10}$. In both cases we construct a test-statistic using Proposition 1. We estimate the asymptotic variance matrix of $\sqrt{n} \cdot (\widehat{\Delta} - \Delta_0)$ as $\widehat{\Omega}_\Delta \equiv \frac{1}{n} \sum_{i=1}^n \widehat{\psi}_{\Delta,n}(V_i;\widehat{\theta}) \cdot \widehat{\psi}_{\Delta,n}(V_i;\widehat{\theta})'$, where the influence function $\widehat{\psi}_{\Delta,n}(V_i;\widehat{\theta})$ is constructed in the manner described in Appendix A (Section A5.1). From here, we estimate $\widehat{\sigma}^2_{\widehat{\Delta}_2 - \widehat{\Delta}_1}$, the asymptotic variance of $\sqrt{n} \cdot ((\widehat{\Delta}_2 - \widehat{\Delta}_1) - (\Delta_{20} - \Delta_{10}))$ and construct the test-statistic $\widehat{t}_{\widehat{\Delta}_2 - \widehat{\Delta}_1} \equiv \frac{\sqrt{n} \cdot (\widehat{\Delta}_2 - \widehat{\Delta}_1)}{\widehat{\sigma}_{\widehat{\Delta}_2 - \widehat{\Delta}_1}}$, which is asymptotically $\mathcal{N}(0,1)$ if $\Delta_{20} - \Delta_{10} = 0$. We proceed to test H_0^a and H_0^b comparing $\widehat{t}_{\widehat{\Delta}_2 - \widehat{\Delta}_1}$ against the corresponding $\mathcal{N}(0,1)$ critical values using a target significance level of 5%. Since $\Delta_{20} > \Delta_{10}$, we should reject H_0^b with frequency approaching one as n grows, while we should reject H_0^a with a frequency that is close to the target significance level and decreasing[8] with n. These predictions are supported by the findings summarized in Table 2, which correspond to $c_h = 1$. The Empirical Supplement includes the results for $c_h = 0.80$ and $c_h = 1.40$, and they are all in line with our findings for $c_h = 1$, suggesting once again that our results are robust across our bandwidth choices.

Table 2: Monte Carlo results for strategic-effect hypotheses tests.

Monte Carlo rejection frequencies of null hypothesis		
	$H_0^a : \Delta_{20} \geq \Delta_{10}$ vs. $H_1^a : \Delta_{20} < \Delta_{10}$	$H_0^b : \Delta_{10} \geq \Delta_{20}$ vs. $H_1^b : \Delta_{10} < \Delta_{20}$
$n = 500$	0.074	0.347
$n = 1{,}000$	0.063	0.580
$n = 2{,}000$	0.045	0.760

- True parameters values: $\Delta_{10} = 0.5$ and $\Delta_{20} = 1$
- 1,000 simulations, $c_h = 1.0$, Gaussian kernel
- Target significance level 5%

5.2 Estimation results for $\pi(Z)$, the probability of cooperation

Next we study the performance of our approach to estimate the probability of cooperation. As we did previously, let $\pi(Z)$ denote the probability of cooperation conditional on Z, and let π denote the aggregate (i.e, unconditional) probability of cooperation, so[9] $\pi = E_Z[\pi(Z)]$. For each one of our simulated samples, we estimated $\pi(z)$ for a collection of prespecified values of z, and we also constructed an estimator for π. Our estimator for $\pi(z)$ is $\widehat{\pi}(z, \widehat{\theta})$, as described in (22). We estimate π as $\widehat{\pi} = \frac{1}{n} \sum_{i=1}^n \widehat{\pi}(Z_i, \widehat{\theta})$. It can be shown that, even though each $\widehat{\pi}(Z_i, \widehat{\theta})$ converges at a nonparametric rate, our aggregate estimator $\widehat{\pi}$ converges at a parametric rate, a property that appears to be supported by the results in Table 3, where π is estimated with reasonable precision for all sample sizes analyzed.

[8] Since $\Delta_{20} > \Delta_{10}$, the asymptotic probability of rejecting $H_0^a : \Delta_{20} \geq \Delta_{10}$ against $H_1^a : \Delta_{20} < \Delta_{10}$ is zero.

[9] The aggregate probability of cooperation is $\pi = \frac{1}{3}$ in our simulations.

Table 3: Monte Carlo results for $\widehat{\pi}$

Quantiles of $\widehat{\pi}$ across our simulations						
True value of π	**$n = 500$**					
	0.15^{th} quantile	**0.25^{th} quantile**	**median**	**0.75^{th} quantile**	**0.85^{th} quantile**	**median $\lvert\widehat{\pi} - \pi\rvert$**
$\pi = \frac{1}{3}$	0.112	0.176	0.330	0.540	0.654	0.178

True value of π	**$n = 1,000$**					
	0.15^{th} quantile	**0.25^{th} quantile**	**median**	**0.75^{th} quantile**	**0.85^{th} quantile**	**median $\lvert\widehat{\pi} - \pi\rvert$**
$\pi = \frac{1}{3}$	0.124	0.178	0.325	0.503	0.633	0.161

True value of π	**$n = 2,000$**					
	0.05^{th} quantile	**0.15^{th} quantile**	**median**	**0.75^{th} quantile**	**0.85^{th} quantile**	**median $\lvert\widehat{\pi} - \pi\rvert$**
$\pi = \frac{1}{3}$	0.114	0.176	0.330	0.506	0.619	0.165

- 2,000 simulations, $c_h = 1.0$, Gaussian kernel

Table 4 includes our results for $\widehat{\pi}(z,\widehat{\theta})$ for $z \in \{-1,-0.675,-0.5,-0.25,0,0.25,0.5,0.675,1\}$, where we restrict attention to estimates inside $(0,1)$. While our estimators $\widehat{\pi}(z,\widehat{\theta})$ converge at a nonparametric rate, our results clearly illuminate the fact that *the probability of cooperation is decreasing in Z*. Furthermore, the true value of $\pi(z)$ was included within the 10^{th} and the 75^{th} simulation quantiles of $\widehat{\pi}(z,\widehat{\theta})$ for all values of z and all sample sizes, and $\pi(z)$ was included within the simulation interquartile range of $\widehat{\pi}(z,\widehat{\theta})$ for $z \in \{-1,-0.675,-0.5,0.0.25\}$. Overall, our results were able to reveal important properties of the underlying behavioral selection mechanism, such as the fact that the probability of cooperation is decreasing in Z, and the fact that the aggregate frequency of cooperation is around one-third. The Empirical Supplement includes results for $c_h = 0.80$ and $c_h = 1.40$, and they show that these findings were qualitatively robust across our bandwidth choices.

Table 4: Monte Carlo results for $\widehat{\pi}(z,\widehat{\theta})$ for various values of z.

$n = 500$							
	True value of $\pi(z)$	**Quantiles of $\widehat{\pi}(z,\widehat{\theta})$ across our simulations**					
		0.15^{th} quantile	**0.25^{th} quantile**	**median**	**0.75^{th} quantile**	**0.85^{th} quantile**	**median bias**
$z = -1$	0.652	0.167	0.295	0.526	0.740	0.842	0.225
$z = -0.675$	0.526	0.166	0.239	0.436	0.646	0.756	0.215
$z = -0.50$	0.457	0.150	0.233	0.422	0.625	0.739	0.199
$z = -0.25$	0.360	0.135	0.208	0.379	0.589	0.719	0.180
$z = 0$	0.271	0.094	0.164	0.339	0.528	0.653	0.176
$z = 0.25$	0.195	0.098	0.148	0.298	0.520	0.652	0.148
$z = 0.50$	0.134	0.085	0.142	0.286	0.502	0.658	0.152
$z = 0.675$	0.100	0.081	0.137	0.291	0.510	0.656	0.192
$z = 1$	0.054	0.081	0.123	0.290	0.504	0.633	0.236

$n = 1,000$							
	True value of $\pi(z)$	**Quantiles of $\widehat{\pi}(z,\widehat{\theta})$ across our simulations**					
		0.15^{th} quantile	**0.25^{th} quantile**	**median**	**0.75^{th} quantile**	**0.85^{th} quantile**	**median bias**
$z = -1$	0.652	0.243	0.342	0.547	0.744	0.825	0.199
$z = -0.675$	0.526	0.190	0.273	0.462	0.663	0.770	0.201
$z = -0.50$	0.457	0.161	0.243	0.408	0.609	0.730	0.188
$z = -0.25$	0.360	0.125	0.192	0.359	0.539	0.668	0.173
$z = 0$	0.271	0.104	0.162	0.308	0.509	0.633	0.157
$z = 0.25$	0.195	0.083	0.133	0.278	0.474	0.621	0.141
$z = 0.50$	0.134	0.081	0.128	0.264	0.462	0.598	0.131
$z = 0.675$	0.100	0.068	0.123	0.254	0.447	0.583	0.155
$z = 1$	0.054	0.067	0.108	0.248	0.445	0.569	0.195

$n = 2,000$							
	True value of $\pi(z)$	**Quantiles of $\widehat{\pi}(z,\widehat{\theta})$ across our simulations**					
		0.15^{th} quantile	**0.25^{th} quantile**	**median**	**0.75^{th} quantile**	**0.85^{th} quantile**	**median bias**
$z = -1$	0.652	0.215	0.338	0.565	0.744	0.833	0.193
$z = -0.675$	0.526	0.193	0.291	0.467	0.664	0.765	0.191
$z = -0.50$	0.457	0.169	0.250	0.420	0.620	0.731	0.188
$z = -0.25$	0.360	0.128	0.195	0.357	0.547	0.667	0.176
$z = 0$	0.271	0.098	0.155	0.313	0.498	0.608	0.161
$z = 0.25$	0.195	0.076	0.131	0.272	0.457	0.567	0.143
$z = 0.50$	0.134	0.064	0.109	0.237	0.421	0.519	0.123
$z = 0.675$	0.100	0.058	0.104	0.225	0.394	0.507	0.125
$z = 1$	0.054	0.056	0.086	0.210	0.405	0.515	0.156

2,000 simulations, $c_h = 1.0$, Gaussian kernel

6 An empirical illustration

A nonparametric test for cooperation in discrete games was proposed in Aradillas-López and Kosenkova (2023). As an empirical illustration, they analyze geographic-market entry decisions by Lowe's and Home Depot in the continental United States, and their results suggest evidence consistent with cooperation in small markets, but not in large markets. We revisit this question here by applying the methodology in Section 2, which assumes cooperation and PSNE as the two candidate behavioral models. We label player $p = 1$ as Lowe's and $p = 2$ as Home Depot, and we define a market i as a core-based statistical area (CBSA) in the contiguous United States. Our sample consists of $n = 954$ markets. We say that $Y_{pi} = 1$ if player p had presence in market i in 2022. Our payoff covariates include: X^1_{pi} ≡*Distance between market i and the nearest distribution center of player p*. X^2_i ≡*Small-market indicator*, equal to one if market i's population and number of businesses are both below their medians, and if i is categorized as a micropolitan statistical area. X^3_i ≡*Income per household in market i*. X^4_i ≡*Population density in market i*, and X^5_i ≡*Market size*, measured as $\log(Population_i)$. Our data comes from Aradillas-López and Kosenkova (2023), and its details are described in Section 6 of that paper. In our parameterization of payoffs, we assume that the slope coefficients of the non-strategic payoff shifters are the same for both players, and we include a player-specific fixed effect (i.e, a player-specific constant term) in players' payoffs. Motivated by the findings in Aradillas-López and Kosenkova (2023), who find evidence that noncooperative behavior is related to market size, we use as our instrument $Z_i = X^5_i$ =*Market size*. Unobserved payoff shocks $(\varepsilon_{1i}, \varepsilon_{2i})$ are assumed to be jointly Normal, with mean zero, unit variance, and correlation coefficient ρ_0. Our choice of regressors and our parameterization are compatible with the exclusion restrictions and the identification conditions assumed in Section 2.

6.1 Estimation results for $\widehat{\theta}$

Estimation results are included in Table 5. The non-strategic parameters are estimated using MLE, as described in Section 2.5.1, and the strategic parameters are estimated using the conditional GMM procedure described in Section 2.6, which maintains the assumption that these firms display noncooperative behavior with nonzero probability. Except for income-per-household, all the non-strategic payoff shifters were statistically significant at a 1% significance level, with the signs we would anticipate for distance-to-distribution-center (-), market size (+), and small-market (-). The estimated correlation coefficient ρ_0 was large (≈ 0.82) and statistically significant, suggesting the potential presence of market-specific "fixed effects" observed by players. As in our Monte Carlo experiments, we use a bandwidth of the form $h_n = c_h \cdot \widehat{\sigma}(Z) \cdot n^{-1/5}$ and a Gaussian kernel. The inference range $\mathcal{Z}$ was constructed exactly as in our Monte Carlo experiments. The results shown in Table 5 correspond to $c_h = 1$ (close to the "rule of thumb" bandwidth choice), which yields $h_n \approx 0.25 \cdot \widehat{\sigma}(Z)$. The strategic interaction effect estimates were statistically significant, and they suggest a stronger interaction effect for Lowe's than Home Depot (i.e, $\Delta_{20} < \Delta_{10}$), a conjecture that we formally test below. Our Empirical Supplement presents estimation results for alternative bandwidth choices, using $c_h = 0.80$ and $c_h = 1.40$. As we show there, even though the strategic interaction estimates naturally change[10], the main qualitative finding that both effects are statistically significant, and that $\Delta_{20} < \Delta_{10}$, was robust across our bandwidth choices.

Table 5: Estimation results (standard errors in parenthesis)

Non-strategic parameters	
Player 1 (Lowe's) fixed effect	Player 2 (Home Depot) fixed effect
3.226* (0.552)	2.767* (0.804)
Distance to nearest distribution center	Small-market indicator
-0.323* (0.054)	-0.695* (0.198)
Income per household	Payroll per business establishment
-0.100 (0.088)	-0.231* (0.043)
Population density	Market size
-0.443* (0.174)	0.467* (0.146)
ρ_0	
0.824* (0.263)	

Strategic-interaction parameters	
Player 1 (Lowe's): Δ_{10}	Player 2 (Home Depot): Δ_{20}
0.946* (0.143)	0.257* (0.022)

(∗) denotes statistically significant at a 1% significance level.
Non-strategic parameters were estimated using MLE.
(Δ_1, Δ_2) were estimated using $c_h = 1$ for our bandwidth and a Gaussian kernel.

6.1.1 Testing for asymmetric interaction effects

Our results suggest that the strategic effect for Lowe's is larger than that of Home Depot, i.e, $\Delta_{20} < \Delta_{10}$. Following the steps described in our Monte Carlo experiments, we construct the test-statistic $\widehat{t}_{\widehat{\Delta}_2-\widehat{\Delta}_1} \equiv \frac{\sqrt{n}\cdot(\widehat{\Delta}_2-\widehat{\Delta}_1)}{\widehat{\sigma}_{\widehat{\Delta}_2-\widehat{\Delta}_1}}$, which is asymptotically $\mathcal{N}(0,1)$ by Proposition 1, if $\Delta_{20} - \Delta_{10} = 0$. The value of our test-statistic using $c_h = 1$ for our bandwidth is $\widehat{t}_{\widehat{\Delta}_2-\widehat{\Delta}_1} = -5.329$, leading us to reject the null hypothesis $H_0 : \Delta_{20} \geq \Delta_{10}$ in favor of $H_1 : \Delta_{20} < \Delta_{10}$ with a p-value of 4.9×10^{-8}. We repeat this exercise in the Empirical Supplement for alternative bandwidth choices and, in all cases, we reject the null hypothesis $H_0 : \Delta_{20} \geq \Delta_{10}$ in favor of $H_1 : \Delta_{20} < \Delta_{10}$ with p-values much smaller than 1%.

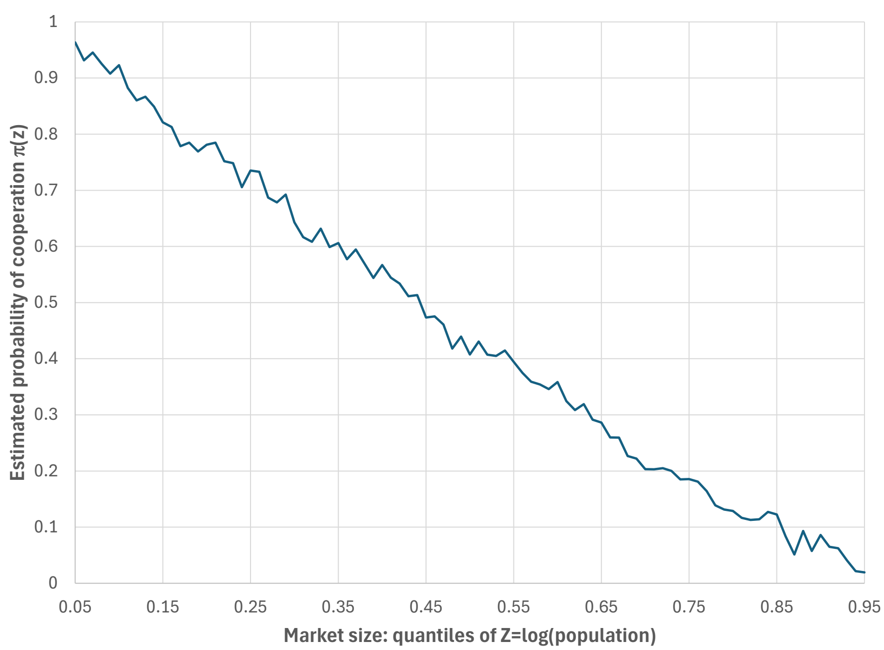
[10]The non-strategic parameters are estimated by MLE and they are unaffected by our bandwidth choice.

Our results suggest, at a statistically significant level, that the strategic effect is greater for Lowe's than for Home Depot.

6.2 Probability of cooperation and market size

The nonparametric test in Aradillas-López and Kosenkova (2023) suggested evidence consistent with cooperative behavior in smaller markets. That approach is based on testable implications of cooperation, but it was not designed to estimate the probability of cooperation. The method proposed here allows us to estimate $\pi(Z)$, the probability of cooperation conditional on Z (market size in our case). We estimate $\pi(z)$ with $\widehat{\pi}(z,\widehat{\theta})$, as described in (22). Focusing on markets where $\widehat{\pi}(Z_i,\widehat{\theta})$ is in the $(0,1)$ interval, figure 3 plots $\widehat{\pi}(Z_i,\widehat{\theta})$ against Z_i. Our results suggest that the probability of cooperation is *decreasing* in market size, lending support to the assertion that entry decisions by these firms are more likely to be consistent with noncooperative behavior in larger markets, and more likely to be consistent with cooperative behavior in smaller ones. The average estimated probability of cooperation across the markets in our sample was 0.452, and the median was 0.418. These results as well as those shown in Figure 3 correspond to $c_h = 1$ for our bandwidth choice. In Appendix A (Figure E1), we replicate this plot for $c_h = 0.80$ and $c_h = 1.40$ and we show that the finding that $\pi(z)$ is decreasing in z (market size) is robust across our bandwidth choices.

Figure 3: Estimated probability of cooperation, $\widehat{\pi}(z,\widehat{\theta})$ and market size



- Values of market size shown range from the 5th to the 95th quantiles.
- Results shown correspond to $c_h = 1$ for our bandwidth, and a Gaussian kernel.

7 Concluding remarks

Econometric analysis of games in non-experimental settings typically relies on the assumption that every observation in the data was produced by the same behavioral model. This paper aims to relax this restriction by assuming instead a collection of candidate behavioral models, each one with potentially multiple solutions, that could have produced each observation in the data, with different observations possibly having been produced by different behavioral models with this collection. In static discrete games with parametric normal-form representations, we show that if there is an observable instrument Z that controls for the dependence between the underlying behavioral and solution selection mechanisms and the game's payoff covariates, the predictions of our global model can be written as a convolution of the predictions of the candidate solutions, with the convolution weights being nonparametric functionals of Z. If there exist outcomes for which our candidate solutions are not observationally equivalent, these weights can be identified and estimated in a first step as functionals of the parameters of interest, and the resulting convolution can be used for estimation and inference. Point identification of the normal-form parameters can result if the behavioral selection mechanism assigns positive probability to behavioral models that have sufficient identification power. However, in the absence of such restrictions, a CS for the parameters can be constructed from our semiparametric convolution. Our setup allows us to recover not only the parameters of the normal-form, but also the convolution weights, which contain information about the propensity of the selection mechanisms to select each behavioral model and each solution within each behavioral model. While our framework relies on the availability of an instrument Z, the restrictions that any candidate Z has to satisfy are testable, and they can be formally evaluated through consistent specification tests. Conditional GMM procedures for estimation and inference were proposed, and Monte Carlo experiments showed that the resulting normal-form estimators, and the behavioral weights recovered had good properties. As an empirical illustration, we applied our conditional-GMM estimation procedure to model geographic entry decisions by Lowe's and Home Depot. Assuming cooperation and pure-strategy Nash equilibrium as the two candidate behavioral models, we found evidence of a statistically significant (and asymmetric) presence of strategic-interaction effects, and we also found evidence of the presence of both behavioral models across markets. Our estimates for the propensity of behavioral selection were consistent with the conjecture that the probability of cooperation decreases with market size.

Appendix A

Every section in this appendix has the format **AX.X** and every equation has the format **(AX.X.X)**. Any section or equation that we reference here which does not have this format refers to the main paper. Similarly, all assumptions referenced here refer to the main paper.

A1 Deriving the expression for $\frac{\partial^2 Q_S(\theta)}{\partial\Delta\partial\Delta'}$ in Section 2.5.2

We have,

$$\begin{aligned}\frac{\partial^2 Q_S(\theta)}{\partial\Delta\partial\Delta'} &= E\left[\frac{\partial^2\tau(U,\theta)}{\partial\Delta\partial\Delta'}\cdot\tau(U,\theta)\right]+E\left[\frac{\partial\tau(U,\theta)}{\partial\Delta}\cdot\frac{\partial\tau(U,\theta)'}{\partial\Delta}\right]\\ &= -\int E\left[\frac{\partial^2\mu_S(U,\theta)}{\partial\Delta\partial\Delta'}\cdot\mathbb{1}\{U\le u\}\right]\cdot E\left[\varphi_S(V,\theta)\cdot\mathbb{1}\{U\le u\}\right]dF_U(u)\\ &\quad+\int E\left[\frac{\partial\mu_S(U,\theta)}{\partial\Delta}\cdot\mathbb{1}\{U\le u\}\right]\cdot E\left[\frac{\partial\mu_S(U,\theta)'}{\partial\Delta}\cdot\mathbb{1}\{U\le u\}\right]dF_U(u).\end{aligned}$$

Evaluated at θ_0, this simplifies to $\frac{\partial^2 Q_S(\theta_0)}{\partial\Delta\partial\Delta'} = \int E\left[\frac{\partial\mu_S(U,\theta_0)}{\partial\Delta}\cdot\mathbb{1}\{U\le u\}\right]\cdot E\left[\frac{\partial\mu_S(U,\theta_0)'}{\partial\Delta}\cdot\mathbb{1}\{U\le u\}\right]dF_U(u)$. Invertibility of $\frac{\partial^2 Q_S(\theta_0)}{\partial\Delta\partial\Delta'}$ will require that $\int E\left[\frac{\partial\mu_S(U,\theta)}{\partial\Delta}\cdot\mathbb{1}\{U\le u\}\right]\cdot E\left[\frac{\partial\mu_S(U,\theta)'}{\partial\Delta}\cdot\mathbb{1}\{U\le u\}\right]dF_U(u)$ have full rank in an open neighborhood $\mathcal{N}$ of θ_0. We have $\frac{\partial\mu_S(U,\theta)}{\partial\Delta} = \left(\frac{\partial\mu_S(U,\theta)'}{\partial\Delta_1}, \frac{\partial\mu_S(U,\theta)'}{\partial\Delta_2}\right)'$, where $\frac{\partial\mu_S(U,\theta)}{\partial\Delta_p} = \frac{\partial m_S^{NC}(X,\theta)}{\partial\Delta_p} + \pi(Z,\theta)\cdot\frac{\partial\Xi_S(X,\theta)}{\partial\Delta_p} + \frac{\partial\pi(Z,\theta)}{\partial\Delta_p}\cdot\Xi_S(X,\theta)$. Let us analyze the terms on the right hand side of this expression. For $p=1,2$, denote $\nabla_p F_{1,2}(\epsilon_1,\epsilon_2|\rho)\equiv\frac{\partial F_{1,2}(\epsilon_1,\epsilon_2|\rho)}{\partial\epsilon_p}$ and let $H_p^{NC}(X,\theta)\equiv \nabla_p F_{1,2}(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1\ ,\ X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2|\rho)$, and $H^C(X,\theta)\equiv\sum_{q=1}^2\nabla_q F_{1,2}(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1 - X_2^{s\prime}\Delta_2\ ,\ X_2^{ns\prime}\beta_2 - X_1^{s\prime}\Delta_1 - X_2^{s\prime}\Delta_2|\rho)$. From (10),

$$\begin{gathered}\frac{\partial m_S^{NC}(X,\theta)}{\partial\Delta_p} = -H_p^{NC}(X,\theta)\cdot X_p^s,\quad \frac{\partial m_S^{C}(X,\theta)}{\partial\Delta_p} = -H^{C}(X,\theta)\cdot X_p^s,\\ \frac{\partial\Xi_S(X,\theta)}{\partial\Delta_p} = \left(H_p^{NC}(X,\theta)-H^C(X,\theta)\right)\cdot X_p^s\end{gathered}\tag{A1}$$

Next, from (11) and (13),

$$\begin{aligned}\frac{\partial\pi(Z,\theta)}{\partial\Delta_p} &= \frac{E\left[-\frac{\partial m_S^{NC}(X,\theta)}{\partial\Delta_p}\cdot\Xi_S(X,\theta)+\left(m_S^{NC}(X,\theta_0)-m_S^{NC}(X,\theta)+\pi(Z)\cdot\Xi_S(X,\theta_0)\right)\cdot\frac{\partial\Xi_S(X,\theta)}{\partial\Delta_p}\,\middle|Z\right]}{E\left[\Xi_S(X,\theta)^2\,\middle|Z\right]}\\ &\quad-2\cdot\frac{E\left[\left(m_S^{NC}(X,\theta_0)-m_S^{NC}(X,\theta)+\pi(Z)\cdot\Xi_S(X,\theta_0)\right)\cdot\Xi_S(X,\theta)\,\middle|Z\right]\cdot E\left[\Xi_S(X,\theta)\frac{\partial\Xi_S(X,\theta)}{\partial\Delta_p}\,\middle|Z\right]}{\left(E\left[\Xi_S(X,\theta)^2\,\middle|Z\right]\right)^2}\end{aligned}$$

Denote $J_p(U,\theta) \equiv \pi(Z,\theta)\cdot H^C(X,\theta) + (1-\pi(Z,\theta))\cdot H_p^{NC}(X,\theta)$. Evaluated at θ_0, we have,

$$\begin{aligned}\frac{\partial \pi(Z,\theta_0)}{\partial \Delta_p} &= -\frac{E\left[\left(\pi(Z)\cdot \frac{\partial m_S^C(X,\theta_0)}{\partial \Delta_p} + (1-\pi(Z))\cdot \frac{\partial m_S^{NC}(X,\theta_0)}{\partial \Delta_p}\right)\cdot \Xi_S(X,\theta_0) \,\middle|\, Z\right]}{E\left[\Xi_S(X,\theta_0)^2 \,\middle|\, Z\right]} \\ &= \frac{E\left[\left(\pi(Z)\cdot H^C(X,\theta_0) + (1-\pi(Z))\cdot H_p^{NC}(X,\theta_0)\right)\cdot \Xi_S(X,\theta_0)\cdot X_p^s \,\middle|\, Z\right]}{E\left[\Xi_S(X,\theta_0)^2 \,\middle|\, Z\right]} \\ &= \frac{E\left[J_p(U,\theta_0)\cdot \Xi_S(X,\theta_0)\cdot X_p^S \,\middle|\, Z\right]}{E\left[\Xi_S(X,\theta_0)^2 \,\middle|\, Z\right]}\end{aligned}$$

From here, $\frac{\partial \mu_S(U,\theta_0)}{\partial \Delta_p}$ simplifies to $\frac{\partial \mu_S(U,\theta_0)}{\partial \Delta_p} = -J_p(U,\theta_0)\cdot X_p^S + \frac{E\left[J_p(U,\theta_0)\cdot \Xi_S(X,\theta_0)\cdot X_p^S \,\middle|\, Z\right]}{E\left[\Xi_S(X,\theta_0)^2 \,\middle|\, Z\right]}\cdot \Xi_S(X,\theta_0)$ and,

$$\frac{\partial \mu_S(U,\theta_0)}{\partial \Delta} = \begin{pmatrix} -J_1(U,\theta_0)\cdot X_1^S + \frac{E\left[J_1(U,\theta_0)\cdot \Xi_S(X,\theta_0)\cdot X_1^S \,\middle|\, Z\right]}{E\left[\Xi_S(X,\theta_0)^2 \,\middle|\, Z\right]}\cdot \Xi_S(X,\theta_0) \\ -J_2(U,\theta_0)\cdot X_2^S + \frac{E\left[J_2(U,\theta_0)\cdot \Xi_S(X,\theta_0)\cdot X_2^S \,\middle|\, Z\right]}{E\left[\Xi_S(X,\theta_0)^2 \,\middle|\, Z\right]}\cdot \Xi_S(X,\theta_0) \end{pmatrix}$$

From here, let

$$\dot{\mu}_{S,\Delta}(U,\theta) \equiv \begin{pmatrix} -J_1(U,\theta)\cdot X_1^S + \frac{E\left[J_1(U,\theta)\cdot \Xi_S(X,\theta)\cdot X_1^S \,\middle|\, Z\right]}{E\left[\Xi_S(X,\theta)^2 \,\middle|\, Z\right]}\cdot \Xi_S(X,\theta) \\ -J_2(U,\theta)\cdot X_2^S + \frac{E\left[J_2(U,\theta)\cdot \Xi_S(X,\theta)\cdot X_2^S \,\middle|\, Z\right]}{E\left[\Xi_S(X,\theta)^2 \,\middle|\, Z\right]}\cdot \Xi_S(X,\theta) \end{pmatrix}$$

Then, $\frac{\partial^2 Q_S(\theta_0)}{\partial \Delta \partial \Delta'} = \int E\left[\dot{\mu}_{S,\Delta}(U,\theta_0)\cdot \mathbb{1}\{U \le u\}\right]\cdot E\left[\dot{\mu}_{S,\Delta}(U,\theta_0)'\cdot \mathbb{1}\{U \le u\}\right] dF_U(u)$. This is the expression in equation (20). ■

A2 Gale Nikaido conditions for the BNE model in Section 3.3

The Jacobian $\nabla_\pi \mathcal{H}(\pi;X,\theta)$ of the BNE system is given by,

$$\nabla_\pi \mathcal{H}(\pi;X,\theta) \equiv \begin{pmatrix} 1 - \frac{\partial H_1(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1\cdot\pi_2\,;\,X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2\cdot\pi_1|\rho)}{\partial \pi_1} & -\frac{\partial H_1(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1\cdot\pi_2\,;\,X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2\cdot\pi_1|\rho)}{\partial \pi_2} \\ -\frac{\partial H_2(X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2\cdot\pi_1\,;\,X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1\cdot\pi_2|\rho)}{\partial \pi_1} & 1 - \frac{\partial H_2(X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2\cdot\pi_1\,;\,X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1\cdot\pi_2|\rho)}{\partial \pi_2} \end{pmatrix}$$

The Gale Nikaido conditions will be satisfied if,

$$
\begin{aligned}
&1-\frac{\partial H_1(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1\cdot\pi_2\ ;\ X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2\cdot\pi_1|\rho)}{\partial\pi_1}\neq 0\\
&1-\frac{\partial H_2(X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2\cdot\pi_1\ ;\ X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1\cdot\pi_2|\rho)}{\partial\pi_2}\neq 0\\
&\left(1-\frac{\partial H_1(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1\cdot\pi_2\ ;\ X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2\cdot\pi_1|\rho)}{\partial\pi_1}\right)\cdot\left(1-\frac{\partial H_2(X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2\cdot\pi_1\ ;\ X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1\cdot\pi_2|\rho)}{\partial\pi_2}\right)\\
&-\frac{\partial H_1(X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1\cdot\pi_2\ ;\ X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2\cdot\pi_1|\rho)}{\partial\pi_2}\cdot\frac{\partial H_2(X_2^{ns\prime}\beta_2 - X_2^{s\prime}\Delta_2\cdot\pi_1\ ;\ X_1^{ns\prime}\beta_1 - X_1^{s\prime}\Delta_1\cdot\pi_2|\rho)}{\partial\pi_1}\neq 0\\
&\forall\ (\pi_1,\pi_2)\in[0,1]^2.
\end{aligned}
\tag{A2}
$$

If (A2) holds, BNE will be unique for (X,θ), so $J(X,\theta)=1$. (A2) can be verified directly $\forall\ (X,\theta)$.

A3 Showing that the restrictions in Assumption I6 are satisfied by $y\in\{(1,0)\ ,(0,1)\}$ under the conditions of Proposition 1

First, we show that the full-rank condition in Assumption I6 can *only* be satisfied for $y\in\{(1,0)\ ,(0,1)\}$. To see why, note that $g^1_{\mathcal{M}_1,NE}(y|X,\theta)=g^2_{\mathcal{M}_1,NE}(y|X,\theta)=0$ for $y\in\{(0,0)\ ,(1,1)\}$. Thus, from (39), for $y\in\{(0,0)\ ,(1,1)\}$,

$$
\begin{aligned}
&\Xi(y|X,\theta)=\\
&\left(g^C_{\mathcal{R}}(y|X,\theta)-g^1_{BNE}(y|X,\theta)\ ,\ g^{NE}_{\mathcal{R}}(y|X,\theta)-g^1_{BNE}(y|X,\theta)\ ,\ 0\ ,\ 0\ ,\ \left(g^j_{BNE}(y|X,\theta)-g^1_{BNE}(y|X,\theta)\right)_{j=2}^{J(X,\theta)}\right)'
\end{aligned}
$$

And the full-rank condition in Assumption I6 cannot be satisfied. Now, for $y\in\{(1,0)\ ,(0,1)\}$,

$$
\begin{aligned}
m_2((1,0)|X,\theta)&=\left(g^C_{\mathcal{R}}((1,0)|X,\theta)-g^1_{BNE}((1,0)|X,\theta)\ ,\ g^{NE}_{\mathcal{R}}((1,0)|X,\theta)+G_{\mathcal{M}_1}(X,\theta)-g^1_{BNE}((1,0)|X,\theta)\right)'\\
m_3((1,0)|X,\theta)&=\left(-G_{\mathcal{M}_1}(X,\theta)\ ,\ g^3_{\mathcal{M}_1,NE}((1,0)|X,\theta)-G_{\mathcal{M}_1}(X,\theta)\ ,\ \left(g^j_{BNE}((1,0)|X,\theta)-g^1_{BNE}((1,0)|X,\theta)\right)_{j=2}^{J(X,\theta)}\right)'\\
m_2((0,1)|X,\theta)&=\left(g^C_{\mathcal{R}}((0,1)|X,\theta)-g^1_{BNE}((0,1)|X,\theta)\ ,\ g^{NE}_{\mathcal{R}}((0,1)|X,\theta)-g^1_{BNE}((0,1)|X,\theta)\right)'\\
m_3((0,1)|X,\theta)&=\left(G_{\mathcal{M}_1}(X,\theta)\ ,\ g^3_{\mathcal{M}_1,NE}((0,1)|X,\theta)\ ,\ \left(g^j_{BNE}((0,1)|X,\theta)-g^1_{BNE}((0,1)|X,\theta)\right)_{j=2}^{J(X,\theta)}\right)'
\end{aligned}
$$

Since $\Xi(y|X,\theta)=\left(m_2(y|X,\theta)'\ ,\ m_3(y|X,\theta)'\right)'$, the restriction in Assumption I6 would be satisfied for $y\in\{(1,0)\ ,(0,1)\}$ if, for each $\theta\in\Theta$ and a.e Z, the support of

$$
\left(g^C_{\mathcal{R}}(y|X,\theta)\ ,\ G_{\mathcal{M}_1}(X,\theta)\ ,\ g^{NE}_{\mathcal{R}}(y|X,\theta)\ ,\ g^3_{\mathcal{M}_1,NE}(y|X,\theta)\ ,\ \left(g^j_{BNE}(y|X,\theta)\right)_{j=1}^{J(X,\theta)}\right)'
$$

conditional on Z, is not contained in any proper linear subspace of $\mathbb{R}^{J(X,\theta)+4}$. This can follow from the conditions in Assumption I6 and the support/exclusion restrictions leading to Proposition 1.

A4 Deriving an expression for $Pr(Y = y|U)$ for the general model in Section 4

Again, group $U \equiv X \cup Z$. For any $y \in \mathcal{Y}$, our model predicts,

$$\mathbb{1}\{Y = y\} = \sum_{b=1}^{\mathcal{B}} \mathbb{1}\{\xi = b\} \sum_{r_b=1}^{\overline{R}_b(X,\theta_0)} \mathbb{1}\{\varepsilon \in \mathcal{R}_{b,r_b}(X,\theta_0)\} \sum_{s_{r_b}=1}^{\overline{S}_{b,r_b}(X,\theta_0)} \mathbb{1}\{\lambda_{b,r_b} = s_{r_b}\} \cdot \mathbb{1}\{Y_{(s_{r_b})} = y\} \tag{A3}$$

Note that $Pr(Y = y \mid \xi = b, \lambda_{b,r_b} = s_{r_b}, U, \varepsilon) = Pr(Y_{(s_{r_b})} = y|X,\varepsilon) = \sigma^{s_{r_b}}_{b,r_b}(y|X,\varepsilon,\theta_0)$. Thus, (A3) yields,

$$Pr(Y = y|\xi,\lambda,U,\varepsilon) = \sum_{b=1}^{\mathcal{B}} \mathbb{1}\{\xi = b\} \sum_{r_b=1}^{\overline{R}_b(X,\theta_0)} \mathbb{1}\{\varepsilon \in \mathcal{R}_{b,r_b}(X,\theta_0)\} \sum_{s_{r_b}=1}^{\overline{S}_{b,r_b}(X,\theta_0)} \mathbb{1}\{\lambda = s_{r_b}\} \cdot \sigma^{s_{r_b}}_{b,r_b}(y|X,\varepsilon,\theta_0).$$

From here, the exclusion restriction $Pr(\lambda = s_{r_b}|\xi = b,\varepsilon,X,Z) = \omega_{b,r_b}(s_{r_b}|Z)$ yields,

$$Pr(Y = y|\xi,U,\varepsilon) = \sum_{b=1}^{\mathcal{B}} \mathbb{1}\{\xi = b\} \sum_{r_b=1}^{\overline{R}_b(X,\theta_0)} \mathbb{1}\{\varepsilon \in \mathcal{R}_{b,r_b}(X,\theta_0)\} \sum_{s_{r_b}=1}^{\overline{S}_{b,r_b}(X,\theta_0)} \omega_{b,r_b}(s_{r_b}|Z) \cdot \sigma^{s_{r_b}}_{b,r_b}(y|X,\varepsilon,\theta_0).$$

Define $g^{s_{r_b}}_{b,r_b}(y|X,\theta_0) \equiv E\Big[\mathbb{1}\{\varepsilon \in \mathcal{R}_{b,r_b}(X,\theta_0)\} \cdot \sigma^{s_{r_b}}_{b,r_b}(y|X,\varepsilon,\theta_0) \; |U\Big] = \int \mathbb{1}\{\varepsilon \in \mathcal{R}_{b,r_b}(X,\theta_0)\} \cdot \sigma^{s_{r_b}}_{b,r_b}(y|X,\varepsilon,\theta_0) \cdot f_\varepsilon(\varepsilon|\rho_0)d\varepsilon$. The expression for $g^{s_{r_b}}_{b,r_b}(y|X,\theta_0)$ is characterized parametrically from our normal-form parameterization. The independence restriction[11] $\varepsilon \perp \xi$ yields,

$$Pr(Y = y|\xi,U) = \sum_{b=1}^{\mathcal{B}} \mathbb{1}\{\xi = b\} \sum_{r_b=1}^{\overline{R}_b(X,\theta_0)} \sum_{s_{r_b}=1}^{\overline{S}_{b,r_b}(X,\theta_0)} \omega_{b,r_b}(s_{r_b}|Z) \cdot g^{s_{r_b}}_{b,r_b}(y|X,\theta_0).$$

[11] We can relax the independence restriction $\varepsilon \perp \xi$ to $\varepsilon \perp \xi|U$ if we parameterize the conditional distribution of $\varepsilon|U$. If we denote it as $F_{\varepsilon|U}(\cdot|U,\rho_0)$, we would have

$$\begin{aligned} g^{s_{r_b}}_{b,r_b}(y|U,\theta_0) &\equiv E\Big[\mathbb{1}\{\varepsilon \in \mathcal{R}_{b,r_b}(X,\theta_0)\} \cdot \sigma^{s_{r_b}}_{b,r_b}(y|X,\varepsilon,\theta_0) \; |U\Big] \\ &= \int \mathbb{1}\{\varepsilon \in \mathcal{R}_{b,r_b}(X,\theta_0)\} \cdot \sigma^{s_{r_b}}_{b,r_b}(y|X,\varepsilon,\theta_0) \cdot f_{\varepsilon|U}(\varepsilon|U,\rho_0)d\varepsilon. \end{aligned}$$

And (A4) becomes,

$$Pr(Y = y|U) = \sum_{b=1}^{\mathcal{B}} \pi_b(Z) \sum_{r_b=1}^{\overline{R}_b(X,\theta_0)} \sum_{s_{r_b}=1}^{\overline{S}_{b,r_b}(X,\theta_0)} \omega_{b,r_b}(s_{r_b}|Z) \cdot g^{s_{r_b}}_{b,r_b}(y|U,\theta_0).$$

Finally, the exclusion restriction $Pr(\xi = b|X,Z) = \pi_b(Z)$ yields,

$$Pr(Y = y|U) = \sum_{b=1}^{\mathcal{B}} \pi_b(Z) \sum_{r_b=1}^{\overline{R}_b(X,\theta_0)} \sum_{s_{r_b}=1}^{\overline{S}_{b,r_b}(X,\theta_0)} \omega_{b,r_b}(s_{r_b}|Z) \cdot g_{b,r_b}^{s_{r_b}}(y|X,\theta_0). \tag{A4}$$

Writing $\pi_1(Z) = 1 - \sum_{b=2}^{\mathcal{B}} \pi_b(Z)$ and $\omega_{b,r_b}(1|Z) = 1 - \sum_{s_{r_b}=2}^{\overline{S}_{b,r_b}(X,\theta_0)} \omega_{b,r_b}(s_{r_b}|Z)$, (A4) becomes,

$$\begin{aligned} Pr(Y = y|U) = & \sum_{r_1=1}^{\overline{R}_1(X,\theta_0)} g_{1,r_1}^1(y|X,\theta_0) + \sum_{b=2}^{\mathcal{B}} \pi_b(Z) \cdot \left(\sum_{r_b=1}^{\overline{R}_b(X,\theta_0)} g_{b,r_b}^1(y|X,\theta_0) - \sum_{r_1=1}^{\overline{R}_1(X,\theta_0)} g_{1,r_1}^1(y|X,\theta_0) \right) \\ & + \sum_{b=1}^{\mathcal{B}} \sum_{r_b=1}^{\overline{R}_b(X,\theta_0)} \sum_{s_{r_b}=2}^{\overline{S}_{b,r_b}(X,\theta_0)} \vartheta_{b,r_b}(s_{r_b}|Z) \cdot \left(g_{b,r_b}^{s_{r_b}}(y|X,\theta_0) - g_{b,r_b}^1(y|X,\theta_0) \right) \end{aligned} \tag{A5}$$

where $\vartheta_{b,r_b}(s_{r_b}|Z) \equiv \pi_b(Z) \cdot \omega_{b,r_b}(s_{r_b}|Z)$. The previous expression can be simplified as follows. Let

$$\begin{aligned} m_1(y|X,\theta_0) &\equiv \sum_{r_1=1}^{\overline{R}_1(X,\theta_0)} g_{1,r_1}^1(y|X,\theta_0) \\ m_2(y|X,\theta_0) &\equiv \left(\sum_{r_b=1}^{\overline{R}_b(X,\theta_0)} g_{b,r_b}^1(y|X,\theta_0) - \sum_{r_1=1}^{\overline{R}_1(X,\theta_0)} g_{1,r_1}^1(y|X,\theta_0) \right)_{b=2}^{\mathcal{B}} \\ m_3(y|X,\theta_0) &\equiv \left(\left(\left(g_{b,r_b}^{s_{r_b}}(y|X,\theta_0) - g_{b,r_b}^1(y|X,\theta_0) \right)_{s_{r_b}=2}^{\overline{S}_{b,r_b}(X,\theta_0)} \right)_{r_b=1}^{\overline{R}_b(X,\theta_0)} \right)_{b=1}^{\mathcal{B}} \\ \Xi(y|X,\theta_0) &\equiv \left(m_2(y|X,\theta_0)', m_3(y|X,\theta_0)' \right)' \end{aligned}$$

and $\delta_1(Z) \equiv \left(\pi_b(Z)\right)_{b=2}^{\mathcal{B}}$, $\delta_2(Z) \equiv \left(\left(\left(\vartheta_{b,r_b}(s_{r_b}|Z) \right)_{s_{r_b}=2}^{\overline{S}_{b,r_b}(X,\theta_0)} \right)_{r_b=1}^{\overline{R}_b(X,\theta_0)} \right)_{b=1}^{\mathcal{B}}$ and $\delta(Z) \equiv \left(\delta_1(Z)', \delta_2(Z)'\right)'$. Then, (A5) can be expressed as, $Pr(Y = y|U) \equiv p_Y(y|U) = m_1(y|X,\theta_0) + \delta(Z)'\Xi(y|X,\theta_0)$. This is the behavioral convolution expression described in Section 4.4. Estimation and inference can then be approached along the lines outlined in Sections 3.5 and 3.7. ■

A5 Details of Proposition 1

We will describe here the structure of the influence function $\psi_{\Delta,n}(V_i;\theta_0)$ in the statement of Proposition 1. Our result follows from standard arguments in semiparametric models, and the step-by-step details can be found in Section S1 of the online Econometric Supplement, which is available at `https://aaradill.github.io/econometric_supplement_uncertain_behavior.pdf`. Let $m_S^C(X,\theta)$, $m_S^{NC}(X,\theta)$ and $\Xi_S(X,\theta)$ be as defined in equation (5) of the paper. From (3), $E[S|U] = m_S^{NC}(X,\theta_0) + \pi(Z) \cdot \Xi_S(X,\theta_0)$. Recall that the population objective function of our conditional GMM estimator is $Q_{S,\mathcal{Z}}(\theta) \equiv \frac{1}{2} \cdot E[\tau_{\mathcal{Z}}(U,\theta)^2]$, with $\tau_{\mathcal{Z}}(u,\theta) \equiv E[\varphi_S(V,\theta) \cdot \mathbb{1}_{\mathcal{Z}}\{U \leq u\}]$, and $\varphi_S(V_i,\theta) \equiv$

$S_i - m_S^{NC}(X_i,\theta) - \pi(Z_i,\theta)\cdot\Xi_S(X_i,\theta)$. From these definitions, it follows therefore that, $\frac{\partial Q_{S,\mathcal{Z}}(\theta)}{\partial\theta} = E\Big[\frac{\partial\tau_{\mathcal{Z}}(U,\theta)}{\partial\theta}\cdot\tau_{\mathcal{Z}}(U,\theta)\Big]$, and[12] $\frac{\partial^2 Q_{S,\mathcal{Z}}(\theta)}{\partial\theta\partial\theta'} = E\Big[\frac{\partial^2\tau_{\mathcal{Z}}(U,\theta)}{\partial\theta\partial\theta'}\cdot\tau_{\mathcal{Z}}(U,\theta)\Big] + E\Big[\frac{\partial\tau_{\mathcal{Z}}(U,\theta)}{\partial\theta}\cdot\frac{\partial\tau_{\mathcal{Z}}(U,\theta)}{\partial\theta}'\Big]$, where $\frac{\partial\tau_{\mathcal{Z}}(u,\theta)}{\partial\theta} = E\Big[\frac{\partial\varphi_S(V,\theta)}{\partial\theta}\cdot\mathbb{1}_{\mathcal{Z}}\{U\le u\}\Big]$, and $\frac{\partial^2\tau_{\mathcal{Z}}(u,\theta)}{\partial\theta\partial\theta'} = E\Big[\frac{\partial^2\varphi_S(V,\theta)}{\partial\theta\partial\theta'}\cdot\mathbb{1}_{\mathcal{Z}}\{U\le u\}\Big]$. For a given (z,θ), denote $R_b^\pi(z,\theta)\equiv E[\Xi_S(X,\theta)^2|Z=z]\cdot f_Z(z)$, and let

$$\begin{aligned}
\psi_{a,n}^\pi(V_i;z,\theta) &\equiv \left(S_i - m_S^{NC}(X_i,\theta)\right)\cdot\Xi_S(X_i,\theta)\cdot K\left(\frac{Z_i-z}{h_n}\right) - E\left[\left(S - m_S^{NC}(X,\theta)\right)\cdot\Xi_S(X,\theta)\cdot K\left(\frac{Z-z}{h_n}\right)\right],\\
\psi_{b,n}^\pi(U_i;z,\theta) &\equiv \Xi_S(X_i,\theta)^2\cdot K\left(\frac{Z_i-z}{h_n}\right) - E\left[\Xi_S(X,\theta)^2\cdot K\left(\frac{Z-z}{h_n}\right)\right],\\
\psi_n^\pi(V_i;z,\theta) &\equiv \frac{1}{R_b^\pi(z,\theta)}\cdot\psi_{a,n}^\pi(V_i;z,\theta) - \frac{\pi(z,\theta)}{R_b^\pi(z,\theta)}\cdot\psi_{b,n}^\pi(U_i;z,\theta)
\end{aligned} \tag{A6}$$

In Section S1.1 of the online Econometric Supplement we show that $\widehat{\pi}(z,\theta)$ satisfies a linear representation result under the restrictions of Proposition 1,

$$\widehat{\pi}(z,\theta) = \pi(z,\theta) + \frac{1}{n\cdot h_n^{d_z}}\sum_{i=1}^n \psi_n^\pi(V_i;z,\theta) + \vartheta_n^\pi(z,\theta), \quad \text{where} \sup_{\substack{\theta\in\Theta\\ z\in\mathcal{Z}}}\left|\vartheta_n^\pi(z,\theta)\right| = o_p\left(n^{-1/2}\right) \tag{A7}$$

The linear representation in (A7) follows from standard semiparametric arguments given our assumptions. In what follows, let $(V_1,V_2,V_3,V_4)\sim F_V\otimes F_V\otimes F_V\otimes F_V$ (four randomly drawn observations from our i.i.d sample $(V_i)_{i=1}^n$). Let

$$\begin{aligned}
\Gamma_{\Delta,n}(V_1;\theta_0) &\equiv \varphi_S(V_1,\theta_0)\cdot E\left[E\left[\frac{\partial\varphi_S(V_3,\theta_0)}{\partial\Delta}\cdot\mathbb{1}_{\mathcal{Z}}\{U_3\le U_2\}\,\middle|\,U_2\right]\cdot\mathbb{1}_{\mathcal{Z}}\{U_1\le U_2\}\,\middle|\,U_1\right]\\
&- E\left[\frac{\partial\varphi_S(V_4,\theta_0)}{\partial\Delta}\cdot E\left[\Xi_S(X_2,\theta_0)\cdot\frac{1}{h_n^{d_Z}}\psi_n^\pi(V_1;Z_2,\theta_0)\cdot\mathbb{1}_{\mathcal{Z}}\{U_2\le U_3\}\middle|V_1,U_3\right]\cdot\mathbb{1}_{\mathcal{Z}}\{U_4\le U_3\}\,\middle|\,V_1\right].
\end{aligned} \tag{A8}$$

Note that $E\left[\Gamma_{\Delta,n}(V_1;\theta_0)\right] = 0$. Next, let $\psi_\gamma(V_i;\gamma_0)$ denote the MLE influence function of $\widehat{\gamma}$, as described in Section 2.5.1. Let

$$\psi_{\Delta,n}(V_i;\theta_0) \equiv -\frac{\partial^2 Q_{S,\mathcal{Z}}(\theta_0)}{\partial\Delta\partial\Delta'}^{-1}\times\left(\Gamma_{\Delta,n}(V_i;\theta_0) + \frac{\partial^2 Q_{S,\mathcal{Z}}(\theta_0)}{\partial\Delta\partial\gamma'}\cdot\psi_\gamma(V_i;\gamma_0)\right) \tag{A9}$$

Note that $E\left[\psi_{\Delta,n}(V;\theta_0)\right] = 0$. In Section S1 of the online Econometric Supplement we show that,

$$\widehat{\Delta} - \Delta_0 = \frac{1}{n}\sum_{i=1}^n \psi_{\Delta,n}(V_i;\theta_0) + o_p\left(n^{-1/2}\right), \tag{A10}$$

This result follows from the asymptotic properties of our semiparametric weights described in (A7) and standard steps in semiparametric extremum-estimation models. It is the result stated in

[12]Note that, evaluated at θ_0, we have $\frac{\partial^2 Q_{S,\mathcal{Z}}(\theta)}{\partial\theta\partial\theta'} = E\Big[\frac{\partial\tau_{\mathcal{Z}}(U,\theta_0)}{\partial\theta}\cdot\frac{\partial\tau_{\mathcal{Z}}(U,\theta_0)}{\partial\theta}'\Big]$. This follows since $\tau_{\mathcal{Z}}(U,\theta_0) = 0$ a.s, and therefore $E\Big[\frac{\partial^2\tau_{\mathcal{Z}}(U,\theta_0)}{\partial\theta\partial\theta'}\cdot\tau_{\mathcal{Z}}(U,\theta_0)\Big] = 0$.

Proposition 1. ■

A5.1 Estimating Ω_Δ, the asymptotic variance of $\sqrt{n}\cdot(\widehat{\Delta}-\Delta_0)$

Using Proposition 1, our proposal is to estimate Ω_Δ, the asymptotic variance of $\sqrt{n}\cdot(\widehat{\Delta}-\Delta_0)$, as $\widehat{\Omega}_\Delta \equiv \frac{1}{n}\sum_{i=1}^n \widehat{\psi}_{\Delta,n}(V_i;\widehat{\theta})\cdot\widehat{\psi}_{\Delta,n}(V_i;\widehat{\theta})'$, where $\widehat{\psi}_{\Delta,n}(V_i;\widehat{\theta})$ is an estimator of the influence function $\psi_{\Delta,n}(V_i;\widehat{\theta})$, described in (A9). Let $\widehat{R}^\pi_a(z,\theta) \equiv \frac{1}{n\cdot h_n^{d_Z}}\sum_{i=1}^n \left(S_i - m_S^{NC}(X_i,\theta)\right)\cdot \Xi_S(X_i,\theta)\cdot K\left(\frac{Z_i-z}{h_n}\right)$ and $\widehat{R}^\pi_b(z,\theta) \equiv \frac{1}{n\cdot h_n^{d_Z}}\sum_{i=1}^n \Xi_S(X_i,\theta)^2\cdot K\left(\frac{Z_i-z}{h_n}\right)$ and note that $\widehat{\pi}(z,\theta)=\frac{\widehat{R}^\pi_a(z,\theta)}{\widehat{R}^\pi_b(z,\theta)}$. We estimate the influence function $\psi^\pi_n(V_i;z,\theta)$ in (A6) as,

$$\widehat{\psi}^\pi_n(V_i;z,\theta) \equiv \frac{1}{\widehat{R}^\pi_b(z,\theta)}\cdot\widehat{\psi}^\pi_{a,n}(V_i;z,\theta) - \frac{\widehat{\pi}(z,\theta)}{\widehat{R}^\pi_b(z,\theta)}\cdot\widehat{\psi}^\pi_{b,n}(U_i;z,\theta), \quad \text{where}$$

$$\widehat{\psi}^\pi_{a,n}(V_i;z,\theta) \equiv \left(S_i - m_S^{NC}(X_i,\theta)\right)\cdot\Xi_S(X_i,\theta)\cdot K\left(\frac{Z_i-z}{h_n}\right) - h_n^{d_Z}\cdot\widehat{R}^\pi_a(z,\theta),$$

$$\widehat{\psi}^\pi_{b,n}(U_i;z,\theta) \equiv \Xi_S(X_i,\theta)^2\cdot K\left(\frac{Z_i-z}{h_n}\right) - h_n^{d_Z}\cdot\widehat{R}^\pi_b(z,\theta)$$

Next, we estimate[13] $\frac{\partial^2\widehat{Q}_{S,\mathcal{Z}}(\theta)}{\partial\theta\partial\theta'} = \frac{1}{n}\sum_{j=1}^n\left(\frac{\partial^2\widehat{\tau}_{\mathcal{Z}}(U_j,\theta)}{\partial\theta\partial\theta'}\cdot\widehat{\tau}_{\mathcal{Z}}(U_j,\theta) + \frac{\partial\widehat{\tau}_{\mathcal{Z}}(U_j,\theta)}{\partial\theta}\cdot\frac{\partial\widehat{\tau}_{\mathcal{Z}}(U_j,\theta)}{\partial\theta}'\right)$, where, $\frac{\partial\widehat{\tau}_{\mathcal{Z}}(u,\theta)}{\partial\theta} = \frac{1}{n}\sum_{i=1}^n\frac{\partial\widehat{\varphi}_S(V_i,\theta)}{\partial\theta}\cdot\mathbb{1}_{\mathcal{Z}}\{U_i\le u\}$, $\frac{\partial^2\widehat{\tau}_{\mathcal{Z}}(u,\theta)}{\partial\theta\partial\theta'} = \frac{1}{n}\sum_{i=1}^n\frac{\partial^2\widehat{\varphi}_S(V_i,\theta)}{\partial\theta\partial\theta'}\cdot\mathbb{1}_{\mathcal{Z}}\{U_i\le u\}$ and,

$$\widehat{\varphi}_S(V_i,\theta) \equiv S_i - m_S^{NC}(X_i,\theta) - \widehat{\pi}(Z_i,\theta)\cdot\Xi_S(X_i,\theta),$$

$$\frac{\partial\widehat{\varphi}_S(V_i,\theta)}{\partial\theta} = -\left(\frac{\partial m_S^{NC}(X_i,\theta)}{\partial\theta} + \frac{\partial\widehat{\pi}(Z_i,\theta)}{\partial\theta}\cdot\Xi_S(X_i,\theta) + \frac{\partial\Xi_S(X_i,\theta)}{\partial\theta}\cdot\widehat{\pi}(Z_i,\theta)\right],$$

$$\frac{\partial^2\widehat{\varphi}_S(V_i,\theta)}{\partial\theta\partial\theta'} = -\left(\frac{\partial^2 m_S^{NC}(X_i,\theta)}{\partial\theta\partial\theta'} + \frac{\partial\widehat{\pi}(Z_i,\theta)}{\partial\theta}\cdot\frac{\partial\Xi_S(X_i,\theta)}{\partial\theta}' + \frac{\partial^2\widehat{\pi}(Z_i,\theta)}{\partial\theta\partial\theta'}\cdot\Xi_S(X_i,\theta)\right.$$
$$\left. + \frac{\partial\Xi_S(X_i,\theta)}{\partial\theta}\cdot\frac{\partial\widehat{\pi}(Z_i,\theta)}{\partial\theta}' + \frac{\partial^2\Xi_S(X_i,\theta)}{\partial\theta\partial\theta'}\cdot\widehat{\pi}(Z_i,\theta)\right).$$

[13]Note that $\frac{\partial^2 Q_{S,\mathcal{Z}}(\theta_0)}{\partial\theta\partial\theta'} = E\left[\frac{\partial\tau_{\mathcal{Z}}(U,\theta_0)}{\partial\theta}\cdot\frac{\partial\tau_{\mathcal{Z}}(U,\theta_0)}{\partial\theta}'\right]$ (since $\tau_{\mathcal{Z}}(U,\theta_0)=0$ a.s), and we can estimate $\frac{\partial^2 Q_{S,\mathcal{Z}}(\theta_0)}{\partial\theta\partial\theta'}$ as

$$\frac{\partial^2\widehat{Q}_{S,\mathcal{Z}}(\widehat{\theta})}{\partial\theta\partial\theta'} = \frac{1}{n}\sum_{j=1}^n\frac{\partial\widehat{\tau}_{\mathcal{Z}}(U_j,\widehat{\theta})}{\partial\theta}\cdot\frac{\partial\widehat{\tau}_{\mathcal{Z}}(U_j,\widehat{\theta})}{\partial\theta}'.$$

The Jacobian $\frac{\partial\widehat{\pi}(Z_i,\theta)}{\partial\theta}$ and Hessian $\frac{\partial^2\widehat{\pi}(Z_i,\theta)}{\partial\theta\partial\theta'}$ of our estimated weights $\widehat{\pi}(Z_i,\theta)$ are estimated as,

$$
\begin{aligned}
\frac{\partial\widehat{\pi}(z,\theta)}{\partial\theta_\ell} &= \frac{1}{\widehat{R}_b^\pi(z,\theta)}\cdot\left(\frac{\partial\widehat{R}_a^\pi(z,\theta)}{\partial\theta_\ell}-\widehat{\pi}(z,\theta)\cdot\frac{\partial\widehat{R}_b^\pi(z,\theta)}{\partial\theta_\ell}\right),\\
\frac{\partial^2\widehat{\pi}(z,\theta)}{\partial\theta_j\partial\theta_\ell} &= -\frac{1}{\widehat{R}_b^\pi(z,\theta)^2}\cdot\frac{\partial\widehat{R}_b^\pi(z,\theta)}{\partial\theta_j}\cdot\left(\frac{\partial\widehat{R}_a^\pi(z,\theta)}{\partial\theta_\ell}-\widehat{\pi}(z,\theta)\cdot\frac{\partial\widehat{R}_b^\pi(z,\theta)}{\partial\theta_\ell}\right)\\
&\quad+\frac{1}{\widehat{R}_b^\pi(z,\theta)}\cdot\left(\frac{\partial^2\widehat{R}_a^\pi(z,\theta)}{\partial\theta_j\partial\theta_\ell}-\frac{\partial\widehat{\pi}(z,\theta)}{\partial\theta_j}\cdot\frac{\widehat{R}_b^\pi(z,\theta)}{\partial\theta_\ell}-\widehat{\pi}(z,\theta)\cdot\frac{\partial^2\widehat{R}_b^\pi(z,\theta)}{\partial\theta_j\partial\theta_\ell}\right),
\end{aligned}
$$

where,

$$
\begin{aligned}
\frac{\partial\widehat{R}_a^\pi(z,\theta)}{\partial\theta_\ell} &= \frac{1}{n\cdot h_n^{d_Z}}\sum_{i=1}^n\left(\frac{\partial\Xi_S(X_i,\theta)}{\partial\theta_\ell}\cdot\left(S_i-m_S^{NC}(X_i,\theta)\right)-\frac{\partial m_S^{NC}(X_i,\theta)}{\partial\theta_\ell}\cdot\Xi_S(X_i,\theta)\right)\cdot K\left(\frac{Z_i-z}{h_n}\right)\\
\frac{\partial\widehat{R}_b^\pi(z,\theta)}{\partial\theta_\ell} &= \frac{2}{n\cdot h_n^{d_Z}}\sum_{i=1}^n\frac{\partial\Xi_S(X_i,\theta)}{\partial\theta_\ell}\cdot\Xi_S(X_i,\theta)\cdot K\left(\frac{Z_i-z}{h_n}\right),\\
\frac{\partial^2\widehat{R}_a^\pi(z,\theta)}{\partial\theta_j\partial\theta_\ell} &= \frac{1}{n\cdot h_n^{d_Z}}\sum_{i=1}^n\left(\frac{\partial^2\Xi_S(X_i,\theta)}{\partial\theta_j\partial\theta_\ell}\cdot\left(S_i-m_S^{NC}(X_i,\theta)\right)-\frac{\partial\Xi_S(X_i,\theta)}{\partial\theta_\ell}\cdot\frac{\partial m_S^{NC}(X_i,\theta)}{\partial\theta_j}\right.\\
&\quad\left.-\frac{\partial^2 m_S^{NC}(X_i,\theta)}{\partial\theta_j\partial\theta_\ell}\cdot\Xi_S(X_i,\theta)-\frac{\partial m_S^{NC}(X_i,\theta)}{\partial\theta_\ell}\cdot\frac{\partial\Xi_S(X_i,\theta)}{\partial\theta_j}\right)\cdot K\left(\frac{Z_i-z}{h_n}\right),\\
\frac{\partial^2\widehat{R}_b^\pi(z,\theta)}{\partial\theta_j\partial\theta_\ell} &= \frac{2}{n\cdot h_n^{d_Z}}\sum_{i=1}^n\left(\frac{\partial^2\Xi_S(X_i,\theta)}{\partial\theta_j\partial\theta_\ell}\cdot\Xi_S(X_i,\theta)+\frac{\partial\Xi_S(X_i,\theta)}{\partial\theta_\ell}\cdot\frac{\partial\Xi_S(X_i,\theta)}{\partial\theta_j}\right)\cdot K\left(\frac{Z_i-z}{h_n}\right).
\end{aligned}
$$

For $m\in\mathbb{N}$ let $(m)_k\equiv m\cdot(m-1)\cdots(m-k)$. From the definition in (A8), we estimate

$$
\begin{aligned}
\widehat{\Gamma}_{\Delta,n}(V_i;\theta) &= \widehat{\varphi}_S(V_i,\theta)\cdot\frac{1}{(n-1)_1}\sum_{j\neq i}\sum_{k\neq i,j}\frac{\partial\widehat{\varphi}_S(V_k,\theta)}{\partial\Delta}\cdot\mathbb{1}_{\mathcal{Z}}\{U_k\leq U_j\}\cdot\mathbb{1}_{\mathcal{Z}}\{U_i\leq U_j\}\\
&-\frac{1}{(n-1)_2}\sum_{j\neq i}\sum_{k\neq i,j}\sum_{\ell\neq i,j,k}\frac{\partial\widehat{\varphi}_S(V_\ell,\theta)}{\partial\Delta}\cdot\Xi_S(X_j,\theta)\cdot\frac{1}{h_n^{d_Z}}\widehat{\psi}_n^\pi(V_i;Z_j,\theta)\cdot\mathbb{1}_{\mathcal{Z}}\{U_j\leq U_k\}\cdot\mathbb{1}_{\mathcal{Z}}\{U_\ell\leq U_k\}
\end{aligned}
$$

Let $\widehat{\psi}_\gamma(V_i;\widehat{\gamma})$ be the estimated MLE influence function for $\widehat{\gamma}$. Using (A9), we estimate[14],

$$
\widehat{\psi}_{\Delta,n}(V_i;\widehat{\theta})\equiv-\frac{\partial^2\widehat{Q}_{S,\mathcal{Z}}(\widehat{\theta})}{\partial\Delta\partial\Delta'}^{-1}\times\left(\widehat{\Gamma}_{\Delta,n}(V_i;\widehat{\theta})+\frac{\partial^2\widehat{Q}_{S,\mathcal{Z}}(\widehat{\theta})}{\partial\Delta\partial\gamma'}\cdot\widehat{\psi}_\gamma(V_i;\widehat{\gamma})\right)\tag{A11}
$$

From here, we estimate Ω_Δ, the asymptotic variance of $\sqrt{n}\cdot(\widehat{\Delta}-\Delta_0)$, as $\widehat{\Omega}_\Delta\equiv\frac{1}{n}\sum_{i=1}^n\widehat{\psi}_{\Delta,n}(V_i;\widehat{\theta})\cdot\widehat{\psi}_{\Delta,n}(V_i;\widehat{\theta})'$. In Section S1 of the online Econometric Supplement we show that, under the restrictions of Proposition 1, $\left\|\frac{1}{n}\sum_{i=1}^n\widehat{\psi}_{\Delta,n}(V_i;\widehat{\theta})\cdot\widehat{\psi}_{\Delta,n}(V_i;\widehat{\theta})'-\frac{1}{n}\sum_{i=1}^n\psi_{\Delta,n}(V_i;\theta_0)\cdot\psi_{\Delta,n}(V_i;\theta_0)'\right\|\xrightarrow{p}0$, so $\left\|\widehat{\Omega}_\Delta-\Omega_\Delta\right\|\xrightarrow{p}0$. Once again, these results are obtained using standard arguments. ■

[14]See footnote 13.

A6 Details of Proposition 2

The CS for Δ_0 proposed in Section 2.7 is based on the population statistic described as follows. Let $g:\mathbb{R}^{d_U}\to\mathbb{R}$ be a real-valued, pre-specified function. As before, let $\varphi_S(V,\theta)\equiv S-m_S^{NC}(X,\theta)-\pi(Z,\theta)\cdot\Xi_S(X,\theta)$. For a given $u\in\mathbb{R}^{d_U}$, we defined $\tau_g(u,\theta)\equiv E[\varphi_S(V,\theta)\cdot g(U)\cdot\mathbb{1}_{\mathcal{Z}}\{U\leq u\}]$, and $M_g(\theta)\equiv E[\tau_g(U,\theta)]$. By iterated expectations, $M_g(\theta_0)=0$. Since γ_0 is identified and estimable, our proposal is to use the population statistic $M_g(\gamma_0,\Delta)$ to construct a CS for Δ. Let $\Theta_g^I=\{\Delta\in\Theta: M_g(\gamma_0,\Delta)=0\}$ be our target identified set for Δ based on the moment restriction $M_g(\theta_0)=0$. Our sample statistic is $\widehat{M}_g(\widehat{\gamma},\Delta)=\frac{1}{n}\sum_{j=1}^n\widehat{\tau}_g(U_j,\widehat{\gamma},\Delta)=\frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n\widehat{\varphi}_S(V_i,\widehat{\gamma},\Delta)\cdot g(U_i)\cdot\mathbb{1}_{\mathcal{Z}}\{U_i\leq U_j\}$. The linear representation result in Proposition 2 is obtained using standard arguments. We will describe the structure of the influence function $\psi_{M_g,n}(V_i;\gamma_0,\Delta)$ here. The interested reader can find the step-by-step details in Section S2 of the online Econometric Supplement. In what follows, let $(V_1,V_2,V_3)\sim F_V\otimes F_V\otimes F_V$ (three randomly drawn observations from our i.i.d sample $(V_i)_{i=1}^n$). First, let

$$H_\gamma^{M_g}(\theta)\equiv E\left[\frac{\partial\varphi_S(V_1,\theta)}{\partial\gamma}\cdot g(U_1)\cdot\mathbb{1}_{\mathcal{Z}}\{U_1\leq U_2\}\right]\tag{A12}$$

Next, define $q_{M_g}^a(V_1,V_2;\theta)\equiv\varphi_S(V_1,\theta)\cdot g(U_1)\cdot\mathbb{1}_{\mathcal{Z}}\{U_1\leq U_2\}$ and note that $M_g(\theta)=E\Big[q_{M_g}^a(V_1,V_2;\theta)\Big]$. Next, denote $\widetilde{q}_{M_g}^a(V_1,V_2;\theta)=\frac{1}{2}\cdot\Big(q_{M_g}^a(V_1,V_2;\theta)+q_{M_g}^a(V_2,V_1;\theta)\Big)$ and note that, by construction, $E\Big[\widetilde{q}_{M_g}^a(V_1,V_2;\theta)\Big]=M_g(\theta)$. Let

$$\overline{q}_{M_g}^a(V_1;\theta)\equiv 2\cdot\Big(E\Big[\widetilde{q}_{M_g}^a(V_1,V_2;\theta)\,\Big|V_1\Big]-M_g(\theta)\Big)\tag{A13}$$

Note that, by construction, $E[\overline{q}_{M_g}^a(V_1;\theta)]=0\;\forall\;\theta$. By inspection, we can also see that, $\overline{q}_{M_g}^a(V_1;\gamma_0,\Delta)=\varphi_S(V_1,\gamma_0,\Delta)\cdot g(U_1)\cdot E\Big[\mathbb{1}_{\mathcal{Z}}\{U_1\leq U_2\}\Big|U_1\Big]\;\forall\;\Delta\in\Theta_g^I$. Next, let $\psi_n^\pi(V_i;z,\theta)$ be as defined in (A6) and define,

$$\overline{q}_{M_g,n}^b(V_1;\theta)=E\left[\frac{1}{h_n^{d_Z}}\cdot\psi_n^\pi(V_1;Z_2,\theta)\cdot\Xi_S(X_2,\theta)\cdot g(U_2)\cdot\mathbb{1}_{\mathcal{Z}}\{U_2\leq U_3\}\,\middle|V_1\right].\tag{A14}$$

Finally, let

$$\psi_{M_g,n}(V_i;\gamma_0,\Delta)\equiv H_\gamma^{M_g}(\gamma_0,\Delta)'\psi_\gamma(V_i;\gamma_0)+\overline{q}_{M_g}^a(V_i;\gamma_0,\Delta)-\overline{q}_{M_g,n}^b(V_i;\gamma_0,\Delta).\tag{A15}$$

where $\psi_\gamma(V_i;\gamma_0)$ is the influence function of our MLE estimator $\widehat{\gamma}$. Note that $E[\psi_\gamma(V_i;\gamma_0)]=0$, and that $E[\overline{q}_{M_g}^a(V_i;\gamma_0,\Delta)]=0$ and $E[\overline{q}_{M_g,n}^b(V_i;\gamma_0,\Delta)]=0\;\forall\;\Delta\in\Theta$. Therefore, $E[\psi_{M_g,n}(V_i;\gamma_0,\Delta)]=0$ $\forall\;\Delta\in\Theta$. In Section S2 of the online Econometric Supplement, we show that under the restrictions of Proposition 2, our CS statistic satisfies,

$$\widehat{M}_g(\widehat{\gamma},\Delta)=M_g(\gamma_0,\Delta)+\frac{1}{n}\sum_{i=1}^n\psi_{M_g,n}(V_i;\gamma_0,\Delta)+\varsigma_n^{M_g}(\Delta),\;\forall\;\Delta\in\Theta,$$
$$\text{where}\quad\sup_{\Delta\in\Theta}\Big|\varsigma_n^{M_g}(\Delta)\Big|=o_p\Big(n^{-1/2}\Big).$$

And over our target identified set Θ_g^I, this result simplifies to,

$$\widehat{M}_g(\widehat{\gamma},\Delta) = \frac{1}{n}\sum_{i=1}^{n}\psi_{M_g,n}(V_i;\gamma_0,\Delta) + \varsigma_n^{M_g}(\Delta),\ \forall\ \Delta \in \Theta_g^I, \quad \text{where} \quad \sup_{\Delta\in\Theta}\left|\varsigma_n^{M_g}(\Delta)\right| = o_p\left(n^{-1/2}\right).$$

This is the statement in Proposition 2. ■

References

Andrews, D. (1987). Asymptotic results for generalized wald tests. *Econometric Theory 3*, 348–358.

Andrews, D. W. K. and X. Shi (2013). Inference for parameters defined by conditional moment inequalities. *Econometrica 81*(2), 609–666.

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

Aradillas-López, A. and L. Kosenkova (2023). A nonparametric test for cooperation in discrete games. *The Econometrics Journal 221*(1), 257–278.

Berry, S. and E. Tamer (2007). Identification in models of oligopoly entry. In R. Blundell, W. Newey, and T. Persson (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Volume II*, pp. 46–85. Cambridge University Press.

Bierens, H. (1982). Consistent model specification tests. *Journal of Econometrics 20*, 105–134.

Bierens, H. and W. Ploberger (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica 65*, 1129–1151.

Billingsley, P. (1995). *Probability and Measure*. New York: Wiley and Sons.

Bjorn, P. and Q. Vuong (1984). Simultaneous equations models for dummy endogenous variables: A game theoretic formulation with an application to labor force participation. CIT working paper, SSWP 537.

Bresnahan, T. F. and P. J. Reiss (1990). Entry in monopoly markets. *Review of Economic Studies 57*, 531–553.

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

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

Chen, X. and Y. Fan (1999). Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics 91*, 373–401.

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

Dominguez, M. and I. Lobato (2004, July). Consistent estimation of models defined by conditional moment restrictions. *Econometrica 72*(5), 1601–1615.

Fan, Y. and Q. Li (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* (4), 865–890.

Gale, D. and H. Nikaido (1965). The jacobian matrix and the global univalence of mappings. *Mathematische Annalen 159*, 81–93.

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

Lehmann, E. and J. Romano (2005). *Testing Statistical Hypotheses*. Springer.

Mas-Colell, A., M. Whinston, and J. Green (1995). *Microeconomic Theory*. Oxford University Press.

Myerson, R. (1990). Fictitious-transfer solutions in cooperative game theory. The center for Mathematical Studies in Economics and Management Science. Northwestern University. Discussion Paper No. 907.

Newey, W. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. Engle and D. McFadden (Eds.), *The Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111–2245. North-Holland.

Nolan, D. and D. Pollard (1987). U-processes: Rates of convergence. *Annals of Statistics 15*, 780–799.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica 39*(3), 577–591.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley. New York, NY.

Sherman, R. (1994). Maximal inequalities for degenerate u-processes with applications to optimization estimators. *Annals of Statistics 22*, 439–459.

Tamer, E. (2003, January). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies 70*(1), 147–167.

Thomson, W. (1994). Cooperative models of bargaining. In R. Aumann and S. Hart (Eds.), *The Handbook of Game Theory with Economic Applications*, Volume 2, pp. 1237–1284. North-Holland.

Watson, J. (2013). *Strategy*. W.W. Norton & Company.

Zheng, J. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics 75*, 263–289.