Binary Response with Bounded Median Dependence

Xun TANG[1]

Department of Economics

University of Pennsylvania

November 29, 2008

Abstract

In this paper we study the identification and estimation of a class of binary response models where conditional medians of additive disturbances are bounded between known or exogenously identified functions of regressors. This class includes several important microeconometric models, such as simultaneous discrete games with incomplete information, binary regressions with endogenous regressors, and binary regressions with interval data or measurement errors on regressors. We characterize the identification region of linear coefficients in subutility functions in this class of models and show how point-identification can be achieved in various microeconometric models under fairly general restrictions on structural primitives. We define a two-step extreme estimator, and prove its consistency for the identification region of coefficients even when the model is not point-identified. We also provide encouraging Monte Carlo evidence of the estimator's performance in finite samples.

KEYWORDS: Binary response, median dependence, games with incomplete information, endogeneity, interval data, measurement error, partial identification, point identification

[1] I thank Stefan Hoderlein, Joel Horowitz, Charles Manski, Rosa Matzkin, Aureo De Paula, Robert Porter, Kyungchul Song, Elie Tamer, Petra Todd and the econometrics seminar participants at Northwestern and University of Pennsylvania for helpful discussions. Financial support from the Center of Studies in Industrial Organizations at Northwestern University is gratefully acknowledged.

1 Introduction

In this paper we study the identification and estimation of a general class of binary regression models that relaxes the assumption of median independence of unobserved disturbances. Specifically, the latent outcome in the binary decision is the sum of a linear index function of regressors and an additive structural disturbance whose conditional median is bounded between known or exogenously identified functions of regressors. The paper makes several contributions to the literature on semiparametric binary response models. First, we show how a variety of important micro-econometric models can be formulated as binary responses with bounded median dependence of the errors under quite general conditions. These models include:

(i) simultaneous discrete games with incomplete information, where players' private signals are independent of each other conditional on observable states and have zero conditional medians;

(ii) binary regressions with an endogenous regressor that is simultaneously determined with the latent outcome, where additive structural errors in the latent simultaneous system are symmetric around zero and independent conditional on exogenous regressors;

(iii) binary responses with interval data on one of the regressors, where the error is median independent from all other perfectly observed regressors, as well as endpoints of the interval;

(iv) binary regressions with a noisy measure of one of the regressors, where the sum of the noise in the measurement and the structural error are median independent from all regressors.

Our approach of estimating these models is novel in the sense that restrictions required to formulate and identify them as a binary response with bounded median dependence are different from (and in some cases weaker than) those used in the literature so far. Second, we characterize the convex identification region of coefficient parameters, and derive sufficient conditions for their point identification in the motivating models above. Remarkably exact identification is always possible under fairly general exclusion restrictions on the linear indices and some conditions on richness of the support of regressors. Our third contribution is to propose and prove the consistency of a novel two-step extreme estimator of the identified set of coefficients. In the first-step, we use kernel regressions to estimate conditional choice probabilities. In the second step, we use the first-step estimates to construct a sample analog of certain limiting function that penalizes coefficients outside the identification with positive numbers. The estimator is defined as minimizers of this sample analog. We do not derive the asymptotic distribution of the estimator, but give some Monte Carlo evidence on the estimator's performance in finite samples in two designs. The estimator is consistent for the identification region in general even when coefficients are not point-identified.

The rest of the paper is organized as follows. Section 2 reviews the related literature. Section 3 defines the class of binary response models with bounded conditional medians, and show how various microeconometric models can be included into this general framework under appropriate restrictions. Section 4 characterizes the identification region of the index coefficients, and proves its convexity. Section 5 defines the two-step extreme estimator of the identification region and proves its consistency regardless of set identification. Section 6 specifies sufficient conditions on structural primitives in motivating submodels that lead to the point-identification of the coefficients. Section 7 show Monte Carlo evidence of the estimator's performance in finite samples. Section 8 concludes.

2 Related Literature

Our paper is related to a vast semiparametric literature on binary response models with the latent variable additively separable in subutility functions of observed regressors and disturbances unobserved by econometricians. Various shape or stochastic restrictions have been introduced on the subutility functions and error distributions. A most popular identifying assumption is the statistical independence between errors and regressors. Matzkin (1992) showed under the independence assumption that subutility function $u(X)$ and the error distribution F_ϵ can be uniquely recovered up to a locational normalization from choice probabilities with additional shape restrictions on u such as monotonicity, concavity and homogeneity. Other authors studied the estimation of binary response models under statistical independence but with different restrictions on $u(.)$ (such as Cosslett (1983), Han (1987), Klein and Spady (1993), and Ichimura (1998)). Another strand of literature studies binary response models under a weaker assumption of median independence of the disturbances, which allows for heterogeneous disturbances. Manski (1985) showed the linear coefficients can be identified up to scale under median independence and restrictions on the richness of the support of regressors, and proposed a consistent maximum score estimator. Other authors have studied the asymptotic distribution and the refinement of maximum score estimators (see Sherman (1988) and Horowitz (1992)). We contribute to this branch of literature by replacing median independence with the restriction of bounded median dependence, and showing how to attain identification and consistent estimation in this case.

Our paper is also related to several previous works on the motivating models we list above. Aradillas-Lopez (2007) and Bajari, Hong, Krainer and Nekipelov (2007) studied discrete games with incomplete information under different sets of restrictions on the players' private signals. Our specification of players' payoffs is the same as in Aradillas-Lopez (2007), where payoffs are additively separable in the linear index of subutilities $X\beta_i$, the private signals ϵ_i, and a constant term that captures the strategic interaction. Our work differs from Aradillas-Lopez (2007) in that the latter requires private signals to be jointly independent from observable states X, while we require them to be independent of each other conditional on observable states. Thus our specification can accommodate private signals with heterogeneous distributions across games. Our identification and estimation strategies are also substantially different from that in Aradillas-Lopez, and are valid even in the presence of multiple equilibria. Bajari, Hong, Krainer and Nekipelov (2007) does not impose any restrictions on how players' payoffs depend on observable states and the interaction between their actions. However, this generality comes at the cost of stronger restrictions on unobservable disturbances. Their approach requires players' private signals to be independently and identically distributed conditional on X, and that the distribution is completely known to the econometrician. In contrast, we are less restrictive about the unobservable distributions of private signals, while the identifying power in our approach derives crucially from the additive form of the payoff functions. Some authors have studied the estimation of binary regressions with endogenous regressors. Hong and Tamer (2003) consistently estimated binary response models with an endogenous regressor when certain instruments are observable. Blundell and Powell (2004) studied a binary regression where the endogenous regressor is determined simultaneously with the unobservable latent variable. They develop a control function approach that accounts for the endogenous regressors, and identifies the average structural function as well as linear coefficients in the latent variable. The key identifying restriction is that the structural disturbance in the binary regression must be independent of all regressors conditional on reduced form errors in the equation defining the endogenous regressors. In comparison, our identification strategy is based on a more intuitive assumption that structural errors in the underlying simultaneous system are independent of each other conditional on exogenous regressors. Manski and Tamer (2002) are the first to study the inference of binary regressions with interval data on one of the regressors. Compared to their work, our approach is valid under a weaker restriction where the size and location of the interval can depend on both the true value of the imperfectly observed regressor and the structural disturbance jointly.

3 The Models

Throughout the paper, we use upper cases for random variables and lower cases for their realized values. Consider a binary choice model:

$$Y = 1(X\beta + \epsilon \geq 0),\ \beta \in \mathbb{R}^K,\ \beta \neq 0 \tag{1}$$

The conditional median of ϵ is defined as:

$$Med(\epsilon|X) = \{\eta \in \mathbb{R} : \Pr(\epsilon \geq \eta|X) \geq \frac{1}{2}, \Pr(\epsilon \leq \eta|X) \geq \frac{1}{2}\}$$

Let $S_X \subseteq \mathbb{R}^K$ denote the support of X and F_X denote its distribution. The distribution of the error term satisfies the following stochastic restriction.

BCQ (Bounded Conditional Medians): Conditional on any $x \in S_X$, ϵ is distributed as $F_{\epsilon|X=x}$ with well-defined continuous densities and $L(x) \leq \sup Med(\epsilon|x)$ and $\inf Med(\epsilon|x) \leq U(x)$ a.e.F_X, where $L(.), U(.)$ are known functions with $L \equiv \inf_{x \in S_X} L(x) > -\infty$, $U \equiv \sup_{x \in S_X} U(x) < +\infty$.

Under BCQ, the median of error terms may depend on the regressors, but the form of such dependence are known to be within certain boundaries. The restriction includes the classical median independence $Med(\epsilon|X) = 0$ as a special case when $\Pr(L(X) = 0 = U(X)) = 1$. It fails if and only if the conditional medians fall outside the interval $[L(x), U(x)]$ for some $x \in S_X$ with positive probability. Alternatively, this restriction can be represented as: $Med(\epsilon|x) \cap [L(x), U(x)] \neq \varnothing$ *a.e.*F_X. We do not require $F_{\epsilon|X=x}$ to be strictly increasing and therefore it may have interval-valued medians (rather than unique medians). A key identifying restriction is that $L(.)$ and $U(.)$ must be known or can be exogenously identified and consistently estimated outside the model. This requirement may appear to be quite restrictive at the first sight, but we argue this framework is general enough to include several interesting microeconometric models where researchers can attain knowledge of these bounds a priori.

3.1 Simultaneous discrete games with incomplete information

Consider a simultaneous 2-by-2 discrete game with the same space of pure strategies $\{1, 0\}$ for players $i = 1, 2$. The payoff structure is :

	0	1
0	$0, 0$	$0, X\beta_2 - \epsilon_2$
1	$X\beta_1 - \epsilon_1, 0$	$X\beta_1 + \delta_1 - \epsilon_1, X\beta_2 + \delta_2 - \epsilon_2$

where $X \in \mathbb{R}^K$ is a vector of payoff-related exogenous states observed by both players and econometricians, and $\epsilon \equiv (\epsilon_1, \epsilon_2)$ where $\epsilon_i \in \mathbb{R}^1$ is an idiosyncratic payoff-related component only observed by each player i but not by the rival or econometricians. The joint distribution of these disturbances conditional on X (denoted $F_{\epsilon|X}$), as well as the structural parameters $\theta \equiv \{\beta_i, \delta_i\}_{i=1,2}$, are common knowledge among both players. Econometricians do not know θ, but know $\delta_i < 0$ for $i = 1, 2$, ϵ_1 is independent of ϵ_2 conditional on X, and $Med(\epsilon_i|X) = 0$ for $i = 1, 2$.

Let S_X denote the support of X, and $S_{\epsilon_i}(x)$ denote the support of ϵ_i conditional on $X = x$. A *pure strategy* for a player is a mapping $g_i : S_X \otimes S_{\epsilon_i}(X) \to \{0, 1\}$. A *pure-strategy Bayesian Nash equilibrium* (BNE) is characterized by a pair of set-valued functions $A_i : S_X \to S_{\epsilon_i}(X)$ such that for all $x \in S_X$ and $\varepsilon_i \in S_{\epsilon_i}(x)$, $g_i(x, \varepsilon_i) = 1(\varepsilon_i \in A_i(x))$ (where $1(.)$ is the indicator function that equals 1 if the event "." happens) and

$$\begin{aligned} A_1^*(x) &= \{\varepsilon_1 : \varepsilon_1 \leq x\beta_1 + \delta_1 P(\epsilon_2 \in A_2^*(x)|\varepsilon_1, x)\} \\ A_2^*(x) &= \{\varepsilon_2 : \varepsilon_2 \leq x\beta_2 + \delta_2 P(\epsilon_1 \in A_1^*(x)|\varepsilon_2, x)\} \end{aligned}$$

In general $A_i^*(x)$ is a mapping from structural primitives $\{\delta_i, \beta_i\}_{i=1,2}$ and $F_{\epsilon|X=x}$ into subsets of $S_{\epsilon_i}(x)$, and is independent of realizations of (ϵ_1, ϵ_2).[2] We maintain that the data observed by econometricians are generated by players following pure strategies only, and that ϵ_1 and ϵ_2 are independent conditional on X with $Med(\epsilon_i|X) = 0$ for $i = 1, 2$. Hence choice probabilities $p(x) \equiv [p_1(x)\ p_2(x)]$ observed from data (where $p_i(x) \equiv \Pr(\textit{player } i \textit{ chooses } 1|X = x)$) must satisfy the following fixed-point equation in *any* pure-strategy BNE,

$$\begin{bmatrix} p_1(x) \\ p_2(x) \end{bmatrix} = \begin{bmatrix} F_{\epsilon_1|X=x}(x\beta_1 + p_2(x)\delta_1) \\ F_{\epsilon_2|X=x}(x\beta_2 + p_1(x)\delta_2) \end{bmatrix} \tag{2}$$

This characterization of BNE is identical with the definition of *Quantal Response Equilibrium* in McKinley and Palfrey (1995). The latter is a special case of BNE when error distributions are independent across the choices. The existence of BNE follows from the Brouwer's Fixed Point Theorem. A generic value of parameters θ can generate $p(x)$ if and only if it can generate $p_i(x)$ in the binary response $Y_i = 1(X\beta_i + p_{-i}(X)\delta_i - \epsilon_i \geq 0)$ for $i = 1, 2$. As in binary response models, δ_i needs to be normalized to -1 for $i = 1, 2$ to attain identification of the other parameters in β_i. Define $\tilde{\epsilon}_i = p_{-i}(x) + \epsilon_i$. Both decision process for $i = 1, 2$ fits in our general framework with $L_i(x) = U_i(x) = -p_{-i}(x)$, where reduced form choice probabilities $p_i(x)$ is known to econometricians completely from observable data. In Section 6 below, we formally define and prove identification of (β_1, β_2) under the conditional and

[2] More generally, pure-strategies should take the form $g_i(x, \varepsilon_i) = 1(\varepsilon_i \in A_i(\varepsilon_i, x))$, but this can be easily represented as $g_i(x, \varepsilon_i) = 1(\varepsilon_i \in A_i^*(x))$ with $A_i^*(x) \equiv \{\varepsilon_i : \varepsilon_i \in A_i(\varepsilon_i, x)\}$.

median independence restriction above, an exclusion restriction in the indices $X\beta_i$, as well as some other regularity conditions on model primitives.

Several recent literature have studied the estimation of such static discrete games with incomplete information, including Aradillas-Lopez (2007) and Bajari, Hong, Krainer and Nekipelov (2007). Aradillas-Lopez focuses on a case where (ϵ_1, ϵ_2) are jointly independent from observable states X. He extends the semiparametric likelihood estimator in Klein and Spady (1993) to this game-theoretic setup. The uniqueness of BNE is crucial for his approach, which requires a well-defined likelihood function. To address this issue of cardinality, Aradillas-Lopez gives sufficient and necessary conditions for the uniqueness of the equilibrium. Bajari et.al (2007) does not impose any form restrictions on the conditional median of the sub-utilities (i.e. $X\beta$ is replaced with a general function $u(X)$). They show $u(.)$ can be identified nonparametrically provided disturbances are independently and identically distributed across players conditional on X and that $F_{\varepsilon_1,\varepsilon_2|X}$ is completely known to the researcher. The main limitation of the approach is, of course, the distribution of the disturbances is rarely known to researchers in econometric implementations.

In comparison, our approach of formulating the BNE as a system of two binary regressions with identified form of median dependence has two advantages. First, the identification of structural parameters does not require any strong form restrictions on the distribution of disturbances. In particular, it allows for heterogenous games where the distribution of disturbances are determined by observable states. Second, multiplicity of the equilibria is not an issue for estimating the model, as the recoverability of parameters does not depend on the form of a well-defined likelihood function. Instead, identification is solely based on an universal characterization in (2) that is shared by possible equilibrium. In case of multiple pure-strategy equilibria, we only require that the equilibrium outcome follows a continuous path in observable states x. We remain otherwise agnostic about the equilibrium selection mechanism.

3.2 Binary regressions with endogeneity

Let $Y_1 = 1(Y_1^* \geq 0)$, where Y_1^* is a scalar latent variable simultaneously determined with Y_2 in the following linear system with $\gamma_1\gamma_2 \neq 1$:

$$
\begin{aligned}
Y_1^* &= X_0\beta_{01} + X_1\beta_1 + Y_2\gamma_1 + \epsilon_1 \\
Y_2 &= X_0\beta_{02} + X_2\beta_2 + Y_1^*\gamma_2 + \epsilon_2
\end{aligned}
$$

where X_0, X_1, X_2 are random vectors with mutually exclusive coordinates.[3] Blundell and Powell (2004) discusses an related empirical application where Y_1 is an individual's decision to participate in work. In their specifications, Y_1^* is the difference between the total number of hours an individual can supply and certain measure of "reservation hours", X_0, X_1, X_2 include social demographic characteristics of the individual as well as other determinants of the market wage, and Y_2 is the log of the individual's total income from some non-work activities.

It is easy to see that $\beta_{01}, \beta_1, \gamma_1, \gamma_2$ can only be identified up to scale even when β_{02}, β_2 are known ex ante. The data-generating process can be equivalently represented by a triangular system where Y_2 is first generated by exogenous regressors and reduced form errors

$$Y_2 = X_0\delta_0 + X_1\delta_1 + X_2\delta_2 + \eta \tag{3}$$

where $\delta_0 \equiv \frac{\beta_{02}+\gamma_2\beta_{01}}{1-\gamma_1\gamma_2}, \delta_1 \equiv \frac{\gamma_2\beta_1}{1-\gamma_1\gamma_2}, \delta_2 \equiv \frac{\beta_2}{1-\gamma_1\gamma_2}$ and $\eta \equiv \frac{\epsilon_2+\gamma_2\epsilon_1}{1-\gamma_1\gamma_2}$. Now suppose (i) ϵ_1 is independent of ϵ_2 conditional on all exogenous variables $\tilde{X} \equiv (X_0, X_1, X_2)$, and (ii) both ϵ_1 and ϵ_2 are symmetric around 0 conditional on $\tilde{X}$. Then any linear combinations of ϵ_1 and ϵ_2 are also symmetrically distributed around 0 conditional on $\tilde{X}$. If γ_1 is known to be positive and normalized to 1, the DGP can be represented as

$$Y_1 = 1(X_0\beta_{01} + X_1\beta_1 + E(Y_2|\tilde{X}) + u \geq 0) \tag{4}$$

where $E(Y_2|\tilde{X}) \equiv X_0\delta_0 + X_1\delta_1 + X_2\delta_2$ and $u \equiv \gamma_1\eta + \epsilon_1$. As Section 6 proves below, under these weak restrictions and normalization, both β_{01}, β_1 can be exactly identified under fairly general support conditions on $\tilde{X}$ and restrictions on the parameter space of structural coefficients. It follows that $\gamma_2, \beta_2, \beta_{02}$ are also uniquely recovered from reduced form coefficients $\delta_0, \delta_1, \delta_2$.

To our knowledge, Heckman (1978) was the first to model endogeneity in binary regressions through a underlying simultaneous system where the latent variable and the endogenous regressor are simultaneously determined. His estimation of the linear coefficients is based on the parametric specification of structural error distributions. Blundell and Powell (2004) develops a semiparametric approach that uses control functions to account for endogeneity. In our notation, their approach requires

$$u|(\tilde{X}, Y_2) \sim u|(X_0, X_1, Y_2, \eta) \sim u|\eta \tag{5}$$

[3] I am grateful to Stefan Hoderlein for pointing out that binary regressions with endogenous regressors (generated under simultaneity) can be included in the general framework of this paper.

where "$\sim$" denotes the equality of conditional distributions. We are not aware of any equivalent representation of (5) in terms of structural errors, though a sufficient condition for (5) is that (ϵ_1, ϵ_2) are jointly independent from $\tilde{X}$. In comparison, the identifying restrictions in our approach is the conditional symmetry and independence of the structural errors given exogenous states, and therefore has immediate structural interpretations. Also note that if we strengthen our assumptions by imposing the independence between structural errors and exogenous regressors, then the identification of structural average function follows immediately from knowledge of the identified β_{01}, β_1 and arguments in Blundell and Powell (2004). Hoderlein (2008) estimates binary regressions with endogenous regressors under a completely different assumption of median exclusion, where the median of the structural error is independent of all instruments conditional on reduced form errors. In a more recent work, Hoderlein and Tang (2008) extends arguments in this paper to estimate a model where X_0, X_2 enters the structural form of Y_2 in a nonparametric component that is additively separable from Y_1^* and ϵ_2.

3.3 Binary regressions with interval data or measurement error

(Binary regressions with interval data) Let $Y_i = 1(X\beta + V + \epsilon \geq 0)$, where $X \in \mathbb{R}^K$, $V \in \mathbb{R}$. Researchers observe a random sample of (Y, X, V_0, V_1) and (i) $\Pr(V_0 \leq V \leq V_1) = 1$ and both V_0 and V_1 have bounded support; (ii) $Med(\epsilon|x, v_0, v_1) = 0$ for all (x, v_0, v_1). Then $Y = 1(X\beta + \tilde{\varepsilon} \geq 0)$ where $\tilde{\epsilon} = V + \epsilon$. It follows from (i) and (ii) that $v_0 \leq \inf Med(\tilde{\epsilon}|x, v_0, v_1) \leq \sup Med(\tilde{\epsilon}|x, v_0, v_1) \leq v_1 \; \forall (x, v_0, v_1)$. Denote the $(k+2)$-vectors $[X \; V_0 \; V_1]$ by Z and $[\beta \; 0 \; 0]$ by α. Then the model is reformulated as $Y = 1(Z'\alpha + \tilde{\epsilon} \geq 0)$, where $L(Z) \leq \inf Med(\tilde{\epsilon}|Z) \leq \sup Med(\tilde{\epsilon}|Z) \leq U(Z) \; a.e.F_{\mathbf{Z}}$ with $L(Z) = V_0$ and $U(Z) = V_1$. The parameter space now considered is $\Theta = \{b \in \mathbb{R}^{k+2} : b_{k+1} = b_{k+2} = 0\}$. Thus the model fits in with our framework of binary regressions with bounded conditional medians. Note the identifying restrictions above is weaker than those in Manski and Tamer (2002). In addition to the classical median independence restriction (i.e. $Med(\epsilon|x, v) = 0$ for all x, v), they also require that conditional on the (unobservable) true regressor V, the disturbance ϵ is statistically independent from the random bounds (V_0, V_1). Among other things, this conditional independence restriction rules out an interesting case where the size or location of the interval depends on both V and ϵ jointly. In contrast, our model only requires the median independence of ϵ conditional on X and the bounds.

(Binary regressions with measurement error) Let $Y = 1(X_1\beta + X_2^* + \epsilon \geq 0)$ where one of the regressors $X_2^* \in \mathbb{R}^1$ can only be measured with an additive error, i.e. $X_2 = X_2^* - \eta$. The

key identifying restriction is that conditional on (X_1, X_2), ϵ and η are mutually independent and both symmetric around 0. Then $Y = 1(X_1\beta + X_2 + \tilde{\epsilon} \geq 0)$ with $\tilde{\epsilon} = \epsilon + \eta$ and $Median(\tilde{\epsilon}|X_1 = x_1, X_2 = x_2) = 0$ for all x_1, x_2. An empirical example of this model is an individual's decision for labor participation. Suppose each individual chooses to participate in the labor force if and only if he expects his net payoffs from working or active job searches to be non-negative. These net payoffs are determined by potential employer's perception of individual's abilities X_2^* and other demographic characteristics X_1 (including gender, education, previous job experience, etc). Let X_2 be a certain noisy measure of the individual's ability based on which employers form their perceptions X_2^* (e.g. X_2 may be individuals' scores in standard tests such as SAT). Then the key identifying assumption requires that noises in the employers' perception (i.e. η) and other unobserved factors affecting net payoffs from labour participation are mutually independent and both symmetric around 0 given the demographic features and the test scores.

4 Partial Identification of β_0

In this section, we characterize the identification region of linear coefficients β_0 in the general framework of binary response with median dependence. Specializations into various motivating models is straightforward. Let Γ denote the set of conditional distributions $F_{\epsilon|X}$ that satisfy BCQ, let $\beta_0, F^0_{\epsilon|X}$ denote the true structural parameters in the model, and $p^*(x; \beta_0, F^0_{\epsilon|X})$ denote observed conditional choice probabilities $\Pr(d = 1|x; \beta_0, F^0_{\epsilon|X})$. In this section we characterize the set of coefficients $b \in \mathbb{R}^K$ which, for some choice of $F_{\epsilon|\mathbf{X}} \in \Gamma$, can generate the observed choice probabilities $p^*(x)$ almost everywhere on the support of X (denoted S_X). This reveals the limit of what can be learned about the true parameter β_0 from observables under BCQ, and leads to the definition of our two-step extreme estimator. For any generic pair of coefficients b and conditional error distribution $G_{\varepsilon|X}$, let $p(x; b, G_{\varepsilon|X})$ denote the probability that the person chooses $d = 1$ given x, b and $G_{\varepsilon|X}$ (i.e. $p(x; b, G_{\varepsilon|X}) \equiv \int 1(xb + \epsilon \geq 0) dG_{\epsilon|X=x}$), and let $X(b, G_{\varepsilon|X})$ denote the set $\{x : p(x; b, G_{\epsilon|X}) \neq p^*(x; \beta_0, F^0_{\epsilon|X})\}$.

Definition 1 *The true coefficient β_0 is identified relative to b under BCQ if for all $F_{\epsilon|X} \in \Gamma$, $\Pr(X \in X(b, F_{\epsilon|\mathbf{X}})) > 0$. Furthermore, β_0 is observationally equivalent to b under BCQ if it is not identified relative to b under BCQ. The identification region of β_0 is the set of b that is observationally equivalent to β_0 under BCQ.*

Obviously, identification is always defined under specific restrictions on the nuisance parameter $F_{\epsilon|X}$. By construction, the size of the identification region becomes smaller as stronger restrictions are imposed on the distribution of unobserved disturbances $F_{\epsilon|X}$. As the *BCQ* incorporates median independence as a special case, the identification region under *BCQ* must be bigger than under the latter. *Lemma 1* below fully characterizes the identification region. For the rest of the paper, we use $p^*(x)$ as a shorthand for $p^*(x;\beta_0,F^0_{\epsilon|X})$.

Lemma 1 *In the binary response model (1), b is observationally equivalent to β_0 under BCQ if and only if* $\Pr(X\in\xi'_b)=0$, *where*

$$\xi'_b\equiv\{x:(-xb\le L(x),\ p^*(x)<\frac{1}{2})\ or\ (-xb\ge U(x),\ p^*(x)>\frac{1}{2})\}$$

That $F_{\epsilon|X}$ has continuous conditional densities is only a regularity condition in *BCQ* that enhances the asymptotic properties of our estimators proposed below, but is not necessary for identification lemma above. Instead, the lemma is still valid under a weaker restriction $\sup Med(\epsilon|x)\in Med(\epsilon|x)\ a.e.F_X$. An immediate implication of *Lemma 1* is that the identification region under *BCQ* is $\Theta'_I\equiv\{b:\Pr(X\in\xi'_b)=0\}$. To understand how Lemma 1 helps with estimation and inference, note the characterization of Θ'_I is independent of the unknown structural elements $(\beta_0,F^0_{\epsilon|X})$ given the joint distribution $F_{Y,X}$ observed. Thus, it leads to the definition of a non-stochastic function $Q(b)$ that can be constructed from $F_{Y,X}$ only, and more importantly, is minimized if and only if b is in the identification region. Then an extreme estimator can be constructed by optimizing a properly defined sample analog $\hat{Q}_n(b)$. In general this set of observationally equivalent coefficients Θ'_I will not be a singleton. As additional restrictions are imposed on the nuisance parameter $F_{\epsilon|X}$, the size of the identification region of β will be reduced.

BCQ-2: ϵ has continuous densities conditional on all $x\in S_X$ and $L(x)\le\inf Med(\varepsilon|x)\le\sup Med(\varepsilon|x)\le U(x)\ a.e.F_X$, where $L(.),U(.)$ are known functions with $L\equiv\inf_{x\in S_X}L(x)>-\infty$, and $U\equiv\sup_{x\in S_X}U(x)<+\infty$.

Corollary 1 *In the binary response model (1), the identification region of β under BCQ-2 is $\Theta_I\equiv\{b:\Pr(X\in\xi_b)=0\}$, where*

$$\xi_b\equiv\xi'_b\cup\{x:-xb\notin[L(x),U(x)],\ p^*(x)=\frac{1}{2}\}$$

Median independence is a special case of *BCQ-2* when $L(x) = U(x) = 0$ $a.e.F_{\mathbf{X}}$. Under median independence, the identification region is $\Theta_I^0 \equiv \{b : \Pr(x \in \xi_b^0) = 0\}$, where

$$\xi_b^0 \equiv \{x : (-x'b \leq 0, p^*(x) < \frac{1}{2}) \text{ or } (-x'b \geq 0, p^*(x) > \frac{1}{2}) \text{ or } (-x'b \neq 0, p^*(x) = \frac{1}{2})\}$$

Note $\xi_b' \subseteq \xi_b \subseteq \xi_b^0$ when $L(x) \leq 0 \leq U(x)$ $a.e.F_X$. Thus $\Theta_I^0 \subseteq \Theta_I \subseteq \Theta_I'$. The exact difference between the sizes of these sets is determined by F_X and the linear coefficient b considered. Despite the lack of analytical form of the identification regions of β_0, we know they are convex.

Corollary 2 *Under BCQ, the identification region* Θ_I' *is convex. Under BCQ-2, the identification region* Θ_I *is convex.*

Convexity of Θ_I is a desirable property that brings computational advantages in estimation and inference using our extreme estimator defined below. In particular, convexity facilitates the estimation of the identification region through grid searches. Convexity also helps with constructing confidence regions using the criterion function approach in Chernozhukov, Hong and Tamer (2008), which relies on recovering the distribution of the supreme of the objective function over the identification region.

5 A Smooth Extreme Estimator

We define an extreme estimator for β_0 under *BCQ-2* by minimizing a non-negative, random function $\hat{Q}_n(b)$ constructed from empirical distribution of (X, Y). The idea is that the limiting function of $\hat{Q}_n(.)$ as $n \to \infty$ (denoted $Q(.)$) equals zero if and only if $b \in \Theta_I$, where Θ_I is the identification region of β under *BCQ-2*. Thus the set of minimizers of $\hat{Q}_n(.)$ converges to Θ_I in probability (denoted $\xrightarrow{p}$) under certain set metrics, provided $\hat{Q}_n \xrightarrow{p} Q$ uniformly over the parameter space. Let Θ denote the parameter space of interests. *Lemma 2* below defines the appropriate limiting function $Q(.)$.

Lemma 2 *Define the non-stochastic function*

$$Q(b) \equiv E[1(p^*(X) \geq 1/2)(-U(X) - Xb)_+^2 + 1(p^*(X) \leq 1/2)(-L(X) - Xb)_-^2]$$

where $1(.)$ *is the indicator function,* $a_+ \equiv \max(0, a)$ *and* $a_- \equiv \max(0, -a)$. *Suppose* $\Pr\{-X'b = U(X) \text{ or } -X'b = L(X)\} = 0$ $\forall b \in \Theta$. *Then* $Q(b) \geq 0$ $\forall b \in \Theta$ *and* $Q(b) = 0$ *if and only if* $b \in \Theta_I$, *where* $\Theta_I \subset \Theta$ *is the identification region of* β_0 *under BCQ-2.*

For better asymptotic performance of the estimator, we replace the indicator function in $Q(b)$ with a certain smoothing function $\Lambda : [-1/2, 1/2] \to [0, 1]$. Then *Corollary 3* below proves the identification region under *BCQ-2* is completely characterized as the minimizer of the smoothed version of $Q(b)$. The additional regularity condition necessary for identification under the smooth Λ is $\Pr\{p^*(X) = \frac{1}{2}\} = 0$. This is quite general, and, for instance, can be satisfied if $Med(\epsilon|X) = X\gamma$ for some $\gamma \neq \beta_0$ in $\mathbb{R}^K$, and S_X is not contained in any linear subspace in $\mathbb{R}^K$.

Corollary 3 *Define the non-stochastic function*

$$Q(b) \equiv E[\Lambda(p^*(X) - 1/2)(-U(X) - Xb)_+^2 + \Lambda(1/2 - p^*(X))(-L(X) - Xb)_-^2]$$

where $\Lambda : [-\frac{1}{2}, \frac{1}{2}] \to [0, +\infty)$ *is a smoothing function such that* $\Lambda(c) = 0\ \forall c \leq 0$ *and* $\Lambda(c) > 0$ $\forall c > 0$. *Suppose* $\Pr\{-X'b = U(X)$ *or* $-X'b = L(X)\} = 0\ \forall b \in \Theta$, *and* $\Pr\{p^*(X) = \frac{1}{2}\} = 0$. *Then under BCQ-2,* $Q(b) \geq 0\ \forall b \in \Theta$ *and* $Q(b) = 0$ *if and only if* $b \in \Theta_I$.

Our extreme estimator is defined by replacing $Q(b)$ with its sample analog $\hat{Q}_n(b)$. In the first step, we use kernel regressions to estimate the choice probabilities nonparametrically. In the second step, a sample analog $\hat{Q}_n(b)$ is constructed using the empirical distribution of X and the first-step kernel estimate. The two-step extreme estimator is then defined as the minimizer of $\hat{Q}_n$. For simplicity in exposition, we construct the estimator below for the case where all regressors are continuous. Extensions to cases with discrete regressors entail no conceptual or technical challenge for estimation, and is omitted.[4]

Define the kernel density estimates for $f_0(x_i)$ and $h_0(x_i) \equiv E(Y_i|X_i = x_i)f_0(x_i)$ as

$$\hat{f}(x_i) \equiv \frac{1}{n\sigma_n^K} \sum_{j=1, j\neq i}^n K(\frac{x_j - x_i}{\sigma_n}), \quad \hat{h}(x_i) \equiv \frac{1}{n\sigma_n^K} \sum_{j=1, j\neq i}^n y_j K(\frac{x_j - x_i}{\sigma_n})$$

where $K(.)$ is the kernel function and σ_n is the bandwidth chosen. The nonparametric estimates for $p(x_i)$ is $\hat{p}(x_i) \equiv \hat{h}(x_i)/\hat{f}(x_i)$. Now construct the sample analog of $Q(b)$:

$$\hat{Q}_n(b) = \frac{1}{n} \sum_{i=1}^n \Lambda(\hat{p}(x_i) - \frac{1}{2})[-x_i b - U(x_i)]_+^2 + \Lambda(\frac{1}{2} - \hat{p}(x_i))[-x_i b - L(x_i)]_-^2$$

The two-step extreme estimator is defined as:

$$\hat{\Theta}_n = \arg\min_{\mathbf{b}\in\Theta} \hat{Q}_n(b)$$

[4] The inclusion of discrete regressors affects sufficient conditions for the point identification of β. However, in this section we focus on general partial identification only.

In general, the true parameter $\boldsymbol{\beta}$ may not be point identified (i.e. $Q(b) = 0$ on a non-degenerate set). Therefore we need to choose a metric of differences in sets before defining and proving consistency of our extreme estimator. We choose the same metric as in Manski and Tamer (2002). The Hausdorff metric for the distance between two sets A and B in $\mathbb{R}^K$ is defined as

$$d(A,B) \equiv \max\{\rho(A,B),\ \rho(B,A)\}, \text{ where } \rho(A,B) = \sup_{a\in A}\inf_{b\in B}||a-b||$$

where $||.||$ is the Euclidean norm. *Proposition 1* below proves the two-step extreme estimator is a consistent estimator of the identification region Θ_I under the Hausdorff metric. Regularity conditions for set consistency are collected below.

PAR (Parameter space) The identification region Θ_I is in the interior of a compact, convex parameter space Θ.

RD (Regressors and disturbance) (i) the $(K+1)$-dimensional random vector (X_i', ε_i) is independently and identically distributed; (ii) the support of X (denoted S_X) is bounded, and its continuous coordinates have bounded joint density $f_0(x_1,.,x_K)$, which is m times continuously differentiable on the interior of S_X with $m > k$; (iii) $\Pr\{-L(X) = Xb\} = 0$, $\Pr\{-U(X) = Xb\} = 0$ *and* $\Pr\{p^*(X) = \frac{1}{2}\} = 0$.

KF (Kernel function) (i) $K(.)$ is continuous and zero outside a bounded set; (ii) $\int K(u)du = 1$ and for all $l_1 + .. + l_k < m$, $\int u_1^{l_1}...u_k^{l_k}K(u)du = 0$; (iii) $(\ln n)n^{-1/2}\sigma_n^{-K} \to 0$ *and* $\sqrt{n}(\ln n)\sigma_n^{2m} \to 0$.

SF (Smoothing functions) (i) $\Lambda : [-\frac{1}{2}, \frac{1}{2}] \to [0,1]$ is such that $\Lambda(c) = 0\ \forall c \leq 0$ and $\Lambda(c) > 0\ \forall c > 0$; (ii) Λ is bounded with continuous, bounded first and second derivatives on the interior of the support.

BF (Bounding functions) $\sup_{x\in S_X}|\hat{L}(x) - L(x)| = o_p(1)$ *and* $\sup_{x\in S_X}|\hat{U}(x) - U(x)| = o_p(1)$.

The conditions in $RD(iii)$ are regularity conditions for the identification lemma. Restrictions in $SF(i)$ are also essential for the formulation of the identification region as the set of minimizers of Q. Conditions in $RD(i),(ii)$ and KF imply $\hat{p} \xrightarrow{p} p$ uniformly over S_X at a rate faster than $n^{-1/4}$, which, combined with smoothness property of Λ in $SF(ii)$, facilitates our proof of point-wise convergence of $\hat{Q}_n$ to Q in probability. Given that $\hat{Q}_n$ is convex and continuous over the convex parameter space Θ, this point-wise convergence can be strengthened to uniform convergence over any compact subsets of Θ. That Θ is compact and that the support of X are bounded are regularity conditions that help simplify the proof of consistency by making the integrand of the limiting function uniformly bounded over Θ. However, this may be stronger than necessary for our consistency result, as we only need $\hat{Q}_n \xrightarrow{p} Q$ point-wise in Θ. Finally, not on most occasions, such as in our examples of discrete games with incomplete information and binary regressions with an endogenous regressor, the bounding functions $L(x)$ and $U(x)$ are estimated along with $p^*(x)$ from the sample data in the first step. The conditions in BF require such estimates to converge to the truth uniformly over the support of X at an appropriate rate.

Proposition 1 *Suppose BCQ-2, PAR, RD, TF, KF and BF are satisfied. Then (i)* $\hat{\Theta}_n$ *exists with probability approaching* 1 *and* $\Pr(\rho(\hat{\Theta}_n, \Theta_I) > \varepsilon) \to 0$ *as* $n \to \infty$ *for all* $\varepsilon > 0$*; (ii) Let* $\tilde{\Theta}_n = \{b \in \Theta : \hat{Q}_n(b) \leq \min_{\beta \in \Theta} \hat{Q}_n(\beta) + \delta_n\}$*, where* $\delta_n \to 0$ *almost surely. If* $\sup_{b \in \Theta} |\hat{Q}_n(b) - Q(b)|/\delta_n \xrightarrow{p} 0$*, then* $\Pr(d(\Theta_I, \tilde{\Theta}_n) > \varepsilon) \to 0$ *as* $n \to \infty$.

The proof proceeds by first using the uniform convergence of $\hat{p}$ to p^* to show point-wise convergence of the convex objective function $\hat{Q}_n$ in probability to the (continuous) limiting function Q in Θ. Then the convexity of $\hat{Q}_n$ and Q implies the point-wise convergence can be strengthened into uniform convergence in probability over Θ. This is sufficient for showing (i), and therefore sufficient for the consistency of $\hat{\Theta}_n = \arg\min_{\mathbf{b} \in \Theta} \hat{Q}_n(b)$ for β_0 when β_0 is point-identified (i.e. Θ_I is a singleton). The introduction of the sequence δ_n in the perturbed estimator $\tilde{\Theta}_n$ is due to the subtle issue that in general the identification region Θ_I is not a singleton. $\tilde{\Theta}_n$ is consistent for β_0 (just as $\hat{\Theta}_n$) if β_0 is point-identified. More importantly, it is also consistent for Θ_I in the Hausdorff metric under partial identification.

6 Point identification of β_0

Point identification is the special case where the identification region is reduced to a singleton. Despite the generality in the characterization of Θ_I, point identification of β_0 is possible under fairly weak conditions on the parameter space, the support of regressors, and the form of bounding functions. In this section we first specify conditions for point-identification under the general framework of binary response with bounded median dependence. Later in the subsections, we will discuss in greater detail how these conditions are satisfied by restrictions on structural primitives in the motivating micro-econometric models.

EX (Exclusion restriction) $\exists J \subset \{1, 2, ..., K\}$ *such that for all* $b \in \Theta$, $b_j = 0\ \forall j \in J$.

SX (Support of X) (a) There exists no nonzero vector $\lambda \in \mathbb{R}^{K-\#\{J\}}$ *such that* $\Pr(X'_{-J}\lambda = 0) = 1$ *where* $X_{-J} \equiv (X_j)_{j\in\{1,..,K\}\backslash J}$*; (b) For all* $b, \tilde{b} \in \Theta$ *and* $b_{-J} \neq \tilde{b}_{-J}$, $\Pr\{X_{-J} \in T(b_{-J}, \tilde{b}_{-J})\} > 0$ *where* $T(b_{-J}, \tilde{b}_{-J}) \equiv \{x_{-J} : (L,U) \cap R(x_{-J}; b_{-J}, \tilde{b}_{-J}) \neq \varnothing, x'_{-J}(b_{-J} - \tilde{b}_{-J}) \neq 0\}$ *and* $R(X_{-J}; b_{-J}, \tilde{b}_{-J})$ *is the random interval with endpoints* $X'_{-J}b_{-J}$ *and* $X'_{-J}\tilde{b}_{-J}$*; (c)* $\Pr(a_0 < L(X), U(X) < a_1 | X_{-J} = x_{-J}) > 0$ *for all open interval* $(a_0, a_1) \subset [L, U]$ *and almost everywhere* x_{-J}.

These conditions deliver a point-identification of β_0 through an extended approach of exclusion restrictions. Essentially, it suffices to show for all $b \neq \beta_0$,

$$\Pr\begin{pmatrix} \text{"} -X\beta_0 \leq L(X) \leq U(X) < -Xb\text{"} \\ \textit{or } \text{"} -Xb \leq L(X) \leq U(X) < -X\beta_0\text{"} \end{pmatrix} > 0$$

First, SX(a) requires coordinates $\{X_j\}_{j\in J}$ to be excluded from all index functions considered in the parameter space. Then SX(b) requires the support of the other regressors X_{-J} not to be contained in any linear subspaces. Together they guarantee there is a positive probability that there is a non-degenerate random interval between the true index and any other index with $b \neq \beta_0$ in Θ (denoted $R(x_{-J}; b_{-J}, \tilde{b}_{-J})$). Next, SX(c) requires the excluded regressors X_J to generate enough variation in the bounding functions even conditional on X_{-J}. This means: *(i)* X_J enters both $L(.)$ and $U(.)$; and *(ii)* the joint distribution of these two functions of X_J is so rich that for any given X_{-J}, the probability for both of them to fall within any open interval on $[L, U]$ (in particular, the intersection of $[L, U]$ with $R(x_{-J}; b_{-J}, \tilde{b}_{-J})$) is positive. Thus for all $b \neq \beta_0$ in Θ, at least one of the two events above in Lemma 1 (and Corollary 1) happen with a positive probability. The proof of the proposition below formalizes this idea.

Proposition 2 *Under BCQ-2, PAR, EX, and SX,* β_0 *is identified relative to all other* $b \in \Theta$.

Remark 1 The support conditions in *SX* are quite general, and in particular, allow for both discrete coordinates and bounded support of the regressors. This is an important feature, for the compact support for regressors may come in handy in the proof of root-n consistency and asymptotic normality of the estimator when β_0 is point-identified.

Remark 2 Perhaps a more intuitive explanation of the identifying restrictions is by establishing a link with those conditions in Manski (1985) under median independence. Suppose $L(x) = U(x) = M(x)$ for all $x \in S_X$. Then $Y = 1(X_{-J}\beta_{-J} + M(X) + \tilde{\epsilon} \geq 0)$, where $\tilde{\epsilon} \equiv \epsilon - M(X)$ and $Med(\tilde{\epsilon}|X) = 0$. Hence $M(X)$ can be interpreted as an augmented regressor, whose coefficient is known to be positive, and normalized to be 1. The restrictions in Proposition 2 can be interpreted as a generalization of Manski's identifying conditions in our current framework.

For the subsections below, we revisit several motivating models in Section 3, and show how primitive conditions on model structures can deliver point-identification of β_0.

6.1 Discrete games with incomplete information revisited

Consider the 2-by-2 discrete game with incomplete information in Section 3. The definition of identification of parameters needs to be slightly modified to accommodate the game-theoretic framework and, more importantly, the possibility of multiple equilibria. Multiplicity in Bayesian Nash Equilibria is a concern, as it implies the mapping from structural primitives to observable distributions is a correspondence (rather than a function). Let $\Theta \otimes \mathcal{F}_{CMI}$ denote the parameter space for $\theta \equiv (\beta_i, \delta_i)_{i=1,2}$ and $F_{\epsilon_1,\epsilon_2|\mathbf{X}}$, where $\mathcal{F}_{CMI}$ is the set of distributions of (ϵ_1, ϵ_2) conditional on X that satisfy the conditional independence and median independence restrictions specified in Section 3.1. Let $p^*(X;\theta^0, F^0_{\epsilon_1,\epsilon_2|X}) \in [0,1]^2$ denote the profile of choice probabilities observed for some $\theta^0 \equiv (\theta_1^0, \theta_2^0) \in \Theta$ and $F^0_{\epsilon_1,\epsilon_2|X} \in \mathcal{F}_{CMI}$. For any $\theta, F_{\epsilon_1,\epsilon_2|X} \in \Theta \otimes \mathcal{F}_{CMI}$ and $x \in S_X$, let $\Upsilon(x;\theta, F_{\epsilon_1,\epsilon_2|X})$ denote the set of all choice profiles $p(x) \equiv [p_1(x), p_2(x)]$ that solves the fixed point equations in (2). Note $\Upsilon(x;\theta, F_{\epsilon_1,\epsilon_2|X})$ must be a non-empty correspondence for all $\theta \in \Theta$, $x \in S_X$ and $F_{\epsilon_1,\epsilon_2|X} \in \mathcal{F}_{CMI}$ by Brower's Fixed Point Theorem. Define $\varkappa(\theta, F_{\epsilon_1,\epsilon_2|X}, p^*) \equiv \{x \in S_X : p^*(x) \in \Upsilon(x;\theta, F_{\epsilon_1,\epsilon_2|X})\}$.

Definition 2 *Given an equilibrium outcome p^* observed, θ is observationally equivalent (denoted $\overset{o.e.}{\sim}$) to θ^0 under $\mathcal{F}_{CMI}$ if $\exists F_{\epsilon_1,\epsilon_2|X} \in \mathcal{F}_{CMI}$ such that* $\Pr\{X \in \varkappa(\theta, F^0_{\epsilon_1,\epsilon_2|X}, p^*)\} = 1$. *The identification region of θ^0 in Θ under $\mathcal{F}_{CMI}$ is the subset of $\theta \in \Theta$ such that $\theta \overset{o.e.}{\sim} \theta^0$ under $\mathcal{F}_{CMI}$. We say θ^0 is point-identified under $\mathcal{F}_{CMI}$ if the identification region of θ^0 under $\mathcal{F}_{CMI}$ is θ^0.*

In words, if there exists a $F_{\epsilon_1,\epsilon_2|X} \in \mathcal{F}_{CMI}$ such that $(\theta, F_{\epsilon_1,\epsilon_2|X})$ can always rationalize the observed choice probabilities $p^*(x;\theta^0, F^0_{\epsilon_1,\epsilon_2|X})$ as one of the solutions of the fixed point equations in (2), then θ is considered observationally equivalent to the true parameter θ^0 under the CMI restriction. Two remarks about the definitions are in order. First, identification is always relative to the Bayesian Nash equilibria p^* observed. Second, the definition of " $\overset{o.e.}{\sim}$ " only requires marginal distributions of both players' actions to be rationalizable by observed equilibria, even though econometricians get to observe the joint distribution of both players' actions. This is justified by our choice of focus in this paper. Obviously, the conditional independence restriction "$\epsilon_1 \perp \epsilon_2$ given X" implies

$$\Pr(i \text{ } chooses \text{ } 1, j \text{ } chooses \text{ } 1|X) = \Pr(i \text{ } chooses \text{ } 1|X)\Pr(j \text{ } chooses \text{ } 1|X)$$

and this is a testable implication of the CMI restriction. However, our focus in this paper is on what can be learned about θ *given* that the model is known to be correctly specified, or that the choice profiles p^* observed are known to be rationalizable by structural primitives. Below we establish a link between identifying simultaneous Bayesian games and single-agent binary choice models. Given an equilibrium outcome p^* observed, a player i faces a binary choice with an augmented vector of regressors: $Y_i = 1\{\epsilon_i \leq X\beta_i + p^*_{-i}(X)\delta_i\}$. Let $\theta_i \equiv (\beta_i, \delta_i)$ and Θ_i denote the corresponding parameter space. Suppose the model is correctly specified for some $\theta^0, F^0_{\epsilon_1,\epsilon_2|X} \in \Theta \otimes \mathcal{F}_{CMI}$. Let $\mathcal{F}^i_{MI}$ denote the set of marginal distributions of ϵ_i corresponding to a certain joint distribution $F_{\epsilon_1,\epsilon_2|X}$ in $\mathcal{F}_{CMI}$, and define $\varkappa_i(\theta_i, F_{\epsilon_i|X}, p^*) \equiv \{x : p^*_i(x) = \int 1(\epsilon_i \leq x\beta_i + p^*_{-i}(x)\delta_i)dF_{\epsilon_i|X=x}\}$ for any $\theta_i \in \Theta_i$.

Definition 3 *Given an equilibrium outcome p^* observed, θ_i is unilaterally observationally equivalent to θ^0_i (denoted $\overset{u.o.e.}{\sim}$) under $\mathcal{F}^i_{MI}$ if $\exists F_{\epsilon_i|\mathbf{X}} \in \mathcal{F}^i_{MI}$ such that* $\Pr(X \in \varkappa_i(\theta_i, F_{\epsilon_i|X}, p^*)) = 1$*. The truth θ^0_i is unilaterally point-identified in Θ_i under $\mathcal{F}^i_{MI}$ (given p^*) if $\forall F_{\epsilon_i|X} \in \mathcal{F}^i_{MI}$,* $\Pr(X \in \varkappa_i(\theta_i, F_{\epsilon_i|\mathbf{X}}, p^*)) < 1$ *for all $\theta_i \neq \theta^0_i$ in Θ_i.*

Lemma 3 *Given an equilibrium outcome p^* observed, $\theta \overset{o.e.}{\sim} \theta^0$ in Θ under $\mathcal{F}_{CMI}$ if and only if $\theta_i \overset{u.o.e.}{\sim} \theta^0_i$ under $\mathcal{F}^i_{MI}$ for both $i = 1, 2$.*

The lemma provides a convenient framework for identifying simultaneous discrete games with incomplete information by decomposing it into two binary choice models, each with the rival's choice probabilities entering the player's payoffs as an additional regressor. Equivalently, player i's decision takes the form of a binary regression model with bounded median dependence where $L(x) = p_{-i}(x) = U(x)$. Thus $(\beta^0_{-i}, \delta^0_{-i})$ affects the identification of (β^0_i, δ^0_i) only through the choice probabilities $p^*_{-i}(;\theta^0)$ observed from data. An immediate consequence of this lemma is that θ_i can only be identified up to scale. Hence we normalize $\delta_i = -1$ for $i = 1, 2$ for the rest of this subsection. (With a slight abuse of notation, we continue to use $\Theta \equiv \Theta_1 \otimes \Theta_2$ to denote the parameter space for β^0_1, β^0_2 after normalizing δ^0_i to -1.) Below we specify conditions on model primitives to attain point-identification of $(\beta_i)_{i=1,2}$.

(CMI) For $i = 1, 2$, and for all $x \in S(X)$, the conditional disturbance distributions $F_{\varepsilon_i|X=x}$ are continuously differentiable for all ε_i with conditional median 0.

(PS) For $i = 1, 2$, (i) the true parameter β_i^0 is in the interior of Θ_i, where Θ_i is a convex, compact subset of $\mathbb{R}^K$; (ii) $\exists h_i \in \{1, 2, ..., K\}$ such that $b_{i,h_i} = 0$, $b_{-i,h_i} \neq 0$ for all $b_i \in \Theta_i$; (iii) $\delta_i < 0$ for all δ_i in the parameter space.

Under the conditional independence restriction in CMI, the outcome of a Bayesian Nash Equilibrium is characterized a profile of conditional choice probabilities $p = (p_1, p_2)$ such that (2) is satisfied. The PS assumption requires each player's payoff depends on at least one state variable that does not affect the rival's payoff. This exclusion restriction is instrumental to our identification strategy, as it implies the rival's choice probabilities $p_{-i}(x)$ can vary even when the player's own subutility $X\beta_i$ is held constant. This, combined with the zero conditional median restriction in CMI, enables us to extend the identifying arguments in Proposition 2 to recover $(\beta_i)_{i=1,2}$. Yet a major departure from the general framework in Proposition 2 is that the variability of $p(x)$ now depends on conditions of the structure of the model, and therefore cannot be simply assumed. Below we specify several primitive conditions on support of X and the conditional disturbance distribution $F^0_{\epsilon_1,\epsilon_2|X}$ that solves this issue. Let $S(W)$ denote the support of any generic random variable W, and let $S(W|z)$ denote the conditional support of W given a realized value of another generic random variable $Z = z$.

(DDF) (i) for all $x \in S(X)$, $F^0_{\varepsilon_i|x}$ are Lipschitz continuous on S_{ϵ_i} with an unknown constant C_{F_i}; (ii) there exists an unknown constant $K^i_{F_j} > 0$ such that $\sup_{t\in[\varepsilon^j_L,\varepsilon^j_U]} |F^0_{\varepsilon_j|\bar{x}_{-h_i},\tilde{x}_{h_i}}(t) - F^0_{\varepsilon_j|\bar{x}_{-h_i},x_{h_i}}(t)| \leq K^i_{F_j}|\tilde{x}_{h_i} - x_{h_i}|$ for all $\bar{x}_{-h_i} \in S(X_{-h_i})$ and $x_{h_i}, \tilde{x}_{h_i} \in S(X_{h_i}|\bar{x}_{-h_i})$.

Loosely speaking, *DDF-(i)* requires conditional distributions of ϵ_1, ϵ_2 given any x not to increase too fast, while *DDF-(ii)* requires the marginal distributions of ϵ_1 and ϵ_2 given any $\bar{x}_{-h_i}$ "not to perturb too much" as x_{h_i} changes. These two restrictions enables an application of a version of the Fixed Point Theorem to show choice probabilities p_i, as solutions to (2), are continuous in the excluded regressors X_{h_i} conditional on all the other regressors.

Lemma 4 *Suppose PS and DDF (i)-(ii) are satisfied with $|C_{F_1}C_{F_2}| < 1$. Then for all (b_1, b_2) in the parameter space, there exist solutions $\{p_i(.)\}_{i=1,2}$ to the fixed point equation in (2) such that for $i = 1, 2$, $p_i(x_{h_i}, \bar{x}_{-h_i})$ is continuous in x_{h_i} for any $\bar{x}_{-h_i} \in S(X_{-h_i})$, $x_{h_i} \in S(X_{h_i}|\bar{x}_{-h_i})$.*

In the next lemma, we show for any b_1, b_2 in the parameter space, the support of the equilibrium outcome p_i to be rich enough on $[0, 1]$ given any x_{-h_i}, so that there is a positive probability that $p_i(X)$ falls within any open interval on $[0,1]$. In particular, this implies there is positive probability that the true equilibrium outcome observed p_i^* falls within the intersection of $[0,1]$ and the random interval between $X\beta_i^0$ and Xb_i with $b_i \neq \beta_i^0$ in Θ_i). We prove this by using the continuity of p_{-i} and the following regularity conditions on the distribution and support of X_{h_i} given any $\bar{x}_{-h_i}$.

(REG) For $i = 1, 2$, for all $\bar{x}_{-h_i} \in S(X_{-h_i})$, X_{h_i} is continuously distributed on $\mathbb{R}^1$ and $\Pr(X_{h_i} \in I|\bar{x}_{-h_i}) > 0$ *for any open interval I in $\mathbb{R}^1$.*

Lemma 5 *Under PS, REG and DDF, for all (b_1, b_2) in the parameter space, there exist solutions $\{p_i(.)\}_{i=1,2}$ to the fixed point equation in (2) such that for $i = 1, 2$ and any $(a_1, a_2) \subset [0,1]$,* $\Pr\{p_{-i}(X) \in (a_1, a_2)|\ X_{-h_i} = \bar{x}_{-h_i}\} > 0$ *for all $\bar{x}_{-h_i} \in S(X_{-h_i})$.*

Finally, we need restrictions on the support of X_{-h_i} for $i = 1, 2$ such that for all $b_i \neq \beta_i^0$ in Θ_i, there is a positive probability that $X\beta_i \neq Xb_i$ and the random interval between $X\beta_i$ and Xb_i intersects with the open interval $(0, 1)$.

(RSX) For $i = 1, 2$, (i) for all nonzero vector $\lambda \in \mathbb{R}^{K-1}$, $\Pr(X'_{-h_i}\lambda \neq 0) > 0$*; (ii) there exists an unknown constant $C < \infty$ such that* $\Pr(\min\{|X'b_i|, |X'b_i'|\} \leq C) > 0$ $\forall b_i, b_i' \in \Theta_i$ *where Θ_i is the parameter space for b_i; (iii) Let $X^c_{-h_i}$ and $X^d_{-h_i}$ denote respectively subvectors of continuous and discrete coordinates of X_{-h_i}. For all S such that $P(X^c_{-h_i} \in S) > 0$, $P(X^d_{-h_i} = 0, X^c_{-h_i} \in \alpha S) > 0$ $\forall \alpha \in (-1, 1)$ where $\alpha S \equiv \{\tilde{x}_{-h_i} : \tilde{x}_{-h_i} = \alpha x_{-h_i}$ $for \quad some \quad x_{-h_i} \in S\}$.*

Condition *RSX-(i)* is the standard full-rank restriction on the support of X_{-h_i} so that $\Pr(X\beta_i \neq Xb_i) > 0$. *RSX-(ii)* requires payoff indices to be bounded by unknown constants for all $b_i \in \Theta_i$ with probability 1, while *RSX-(iii)* requires the support of X_{-h_i} to be closed under scalar multiplications with $|\alpha| < 1$. Restrictions in *RSX-(i)* maps into part *(a)* in *SX*, while *RSX-(ii), (iii)* satisfy part *(b)* in *SX*.

Proposition 3 *Under CMI, PS, RSX, REG and DDF, (β_1^0, β_2^0) is point-identified under $\mathcal{F}_{CMI}$.*

As mentioned , *REG, RSX,* and *DDF* admit point-identification of (β_1^0, β_2^0) with bounded support of the regressors, as well as the presence of discrete regressors among those not excluded from either players' payoffs. X_{-h_1,h_2}.

6.2 Binary regressions with endogeneity or interval data revisited

(Binary regressions with endogeneity) Let $\mathcal{F}_{CSI}$ denote the set of conditional distributions of structural errors $F_{\epsilon_1,\epsilon_2|\tilde{X}}$ that are symmetric around 0 and satisfies mutual independence between ϵ_1 and ϵ_2 conditional on $\tilde{X}$. Let $S_{\tilde{X}}$ denote the support of $\tilde{X}$ and $F_{Y_1,Y_2|\tilde{X}=\tilde{x}}$ denote the observable distribution of Y_1, Y_2 conditional on $\tilde{x} \in S_{\tilde{X}}$. Below we extend the general definition of identification in Section 4 to account for the presence of simultaneity. Let Θ denote the parameter space of the vector of structural coefficients, with its generic element denoted by $\theta \equiv (\beta_{01}, \beta_{02}, \beta_1, \beta_2, \gamma_1, \gamma_2)$. The set Θ could reflect any exogenous restrictions on the coefficients, such as scale normalizations. A feature of θ (denoted $\Gamma(\theta)$) is a function that maps from the parameter space into some space of features. For example, $\Gamma(\theta)$ could be a subset of θ (such as structural parameters in the binary regression β_{01}, β_1) or a real- or vector-valued function of θ (such as reduced form coefficients $\delta_0, \delta_1, \delta_2$).

Definition 4 *(Identification under simultaneity of Y_1, Y_2) A vector of coefficients θ is observationally equivalent to θ' under $\mathcal{F}_{CSI}$ if $\exists F_{\epsilon_1,\epsilon_2|\tilde{X}}, F'_{\epsilon_1,\epsilon_2|\tilde{X}} \in \mathcal{F}_{CSI}$ such that*

$$F_{Y_1,Y_2|\tilde{X}=\tilde{x}}(.,.;\theta, F_{\epsilon_1,\epsilon_2|\tilde{X}}) = F_{Y_1,Y_2|\tilde{X}=\tilde{x}}(.,.;\theta', F'_{\epsilon_1,\epsilon_2|\tilde{X}})$$

almost everywhere on $S_{\tilde{X}}$. A feature of the true parameter $\theta^0 \in \Theta$, denoted $\Gamma(\theta^0)$, is identified under $\mathcal{F}_{CSI}$ in Θ if $\Gamma(\theta) = \Gamma(\theta^0)$ for all θ in Θ that are observationally equivalent to θ^0 under $\mathcal{F}_{CSI}$.

It is easy to see that any vector $\theta \equiv (\beta_{01}, \beta_1, \gamma_1, \beta_{02}, \beta_2, \gamma_2)$ must be observationally equivalent to $\theta^c \equiv (c\beta_{01}, c\beta_1, c\gamma_1, \beta_{02}, \beta_2, \gamma_2/c)$ for some constant $c > 0$ under $\mathcal{F}_{CSI}$.[5] Hence a scale normalization of γ_1 is necessary for further discussions in identification. We collect the identifying restrictions below.

(SIM) (i) For all $\theta \in \Theta$, none of the structural coefficients is zero and $\gamma_1 = 1$, $\beta_{02} \neq -\gamma_2\beta_{01}$, $\gamma_1\gamma_2 \neq 1$; (ii) The support of $\tilde{X}$ is not contained in any linear subspace of $\mathbb{R}^{K_0+K_1+K_2}$, where K_i is the number of coordinates in X_i for $i = 0, 1, 2$; (iii) For all $(x_0, x_1) \in S_{X_0,X_1}$

[5] Consider any $F_{\epsilon_1,\epsilon_2|\tilde{X}} \in \mathcal{F}_{CSI}$ and let $F^c_{\epsilon_1,\epsilon_2|\tilde{X}} \equiv F_{c\epsilon_1,\epsilon_2|\tilde{X}}$. By construction the reduced form coefficients corresponding to θ and θ^c are the same. Furthermore, reduced form error distributions corresponding to $(\theta, F_{\epsilon_1,\epsilon_2|\tilde{X}})$ and $(\theta^c, F^c_{\epsilon_1,\epsilon_2|\tilde{X}})$ are also identical. Hence $F_{Y_2|\tilde{X}}(.;\theta, F_{\epsilon_1,\epsilon_2|\tilde{X}}) = F_{Y_2|\tilde{X}}(.;\theta^c, F^c_{\epsilon_1,\epsilon_2|\tilde{X}})$ almost everywhere on $S_{\tilde{X}}$. That $F_{Y_1|Y_2,\tilde{X}}(.;\theta, F_{\epsilon_1,\epsilon_2|\tilde{X}}) = F_{Y_1|Y_2,\tilde{X}}(.;\theta^c, F^c_{\epsilon_1,\epsilon_2|\tilde{X}})$ follows from the definition of Y_1 as a binary outcome, with coefficients β_{01}, β_1 and γ_1 included in the linear index component of the latent variable Y_1^*.

and all δ_2 *in its relevant parameter space,*[6] $\Pr(X_2\delta_2 \in (c_0, c_1)|x_0, x_1) > 0$ *where* (c_0, c_1) *is any non-degenerate open interval in* $\mathbb{R}^1$.

That $\gamma_1 = 1$ is more than a scale normalization, for it also restricts the sign of γ_1 to be positive. Such a priori knowledge is often possible in empirical applications. That $\beta_{02} \neq -\gamma_2\beta_{01}$, $\gamma_1\gamma_2 \neq 1$ and *SIM-(ii)* are necessary for identification of reduced form coefficients. This allows us to hold $E(Y_2|\tilde{X})$ as known while discussing identification of β_{01}, β_1 in the semi-reduced form (4) defining Y_1. The rich support condition in *SIM-(iii)* ensures the identifying argument similar to Proposition 2 can be applied.

Proposition 4 *Under* SIM*,* θ^0 *is identified under* $\mathcal{F}_{CSI}$ *in* Θ*.*

Hoderlein and Tang (2008) show that structural linear coefficients in Y_1^* can be point-identified even when Y_2 is nonlinear in X_0, X_2, provided these two regressors are additively separable from Y_1^* and ϵ_2 in the structural equation defining Y_2. That is, $Y_2 = g(X_0, X_2) + Y_1^*\gamma_2 + \epsilon_2$. They extended the argument in Proposition 4 to show β_{01}, β_1 can be uniquely recovered from observable distributions as long as the structural errors satisfy restrictions on $\mathcal{F}_{CSI}$.

(Binary regressions with measurement errors revisited) Consider the binary choice model with interval data on one of the regressors. The augmented vector of regressors is $Z \equiv [X\ V_0\ V_1] \in \mathbb{R}^{K+2}$. Note by construction, $Z_J = [V_0\ V_1]$, and $L(Z) = V_0$, $U(Z) = V_1$, and $\beta_J = [0\ 0]$. Let V_0 and V_1 have unbounded support and the support of X not to be contained in a linear subspace of $\mathbb{R}^K$. Then conditions *SX1-(a)* and *(b)* are satisfied. And β is point identified if $\Pr(a_0 < L(Z) \leq U(Z) < a_1|X = x) > 0$ for all open interval $(a_0, a_1) \subset \mathbb{R}^1$ and all $x \in S(X)$. This shows that conditions in Manski and Tamer (2002) are in fact sufficient for point-identification of the true coefficient β even under the weaker restrictions of $Med(\epsilon|x, v_0, v_1) = 0$ only.

7 Monte Carlo Experiments

In this section, we study the finite sample performance of the two-step extreme estimator through two Monte Carlo experiments where linear coefficients are set-identified and point-identified respectively. The first experiment consists of two designs of binary response models

[6] That is, $\Theta_{\delta_2} \equiv \{\lambda \in \mathbb{R}^{K_2} : \lambda = \frac{\beta_2}{1-\gamma_1\gamma_2}$ *for some* $\theta \in \Theta\}$

with interval data on one of the regressors, and the second consists an example of a binary regression with an endogenous regressor.

7.1 Binary response with interval data on a regressor

Specifically, $Y = 1\{\beta_0+\beta_1 X+V+\varepsilon \geq 0\}$. In the first design, $V\tilde{}N(0,2)$ and $X\tilde{}N(0,4)$, and in the second design $V\tilde{}Uniform(-2,3)$ and $X\tilde{}Uniform(0,5)$. In both designs, $\varepsilon\tilde{}N(0,1)$, (V,X,ε) are statistically independent, and V is not observed by the researcher. Instead, only $V_0 = int(V)$ and $V_1 = int(V)+1$ are observed. These are exactly the same designs as considered in Manski and Tamer (2002). The sufficient conditions for point identification in Section 6 are not satisfied and there is no reason to believe the coefficients β_0 and β_1 are point identified.

We do not derive the closed form of the identification region. Instead we simulate a large data set with 10^5 observations, and treat it as the population for our Monte Carlo studies. We apply the two-step extreme estimators to this data set and use the estimates to approximate the real identification region. For both designs, we reported the performance of the estimator in samples with $N = 500$, 1000 and 3000 respectively. For each sample size N considered, we simulate 100 different samples by making random draws from the population with replacement. We calculate the two-step extreme estimates for each of the 100 samples. We use Naradaya-Watson kernel regressions to estimate conditional choice probabilities in the first step. The bandwidths are chosen through cross-validations and Gaussian kernels are used. The maximization procedure in the second step is done by a two dimensional grid-search.

In this design, we do not report any confidence regions. Rather, for each of the 100 estimated sets, we record the percentage of the identification region it covers (denoted $P1$), as well as the proportion of the estimated set contained in the identification region (denoted $P2$). We use these two proportions as measures of discrepancies between the two-step estimates and the real identification region. *Table 1* below reports different percentiles of these two measures among the 100 simulations.

Table 1 (a): Normal Design

	$P1$			$P2$		
percentile	$n = 500$	$n = 1000$	$n = 3000$	$n = 500$	$n = 1000$	$n = 3000$
10%	0	0	0.054	0	0	0.344
25%	0.017	0.097	0.345	0.228	0.296	0.487
50%	0.370	0.444	0.571	0.409	0.520	0.594
75%	0.653	0.724	0.787	0.597	0.686	0.701
90%	0.841	0.860	0.934	0.808	0.839	0.853

In the normal design, *Table 1(a)* suggests the discrepancies between the worst estimates and the identification region is quite noticeable for small samples. In particular, the first quartile of $P1$ (the percentage of identification region covered by an estimated set) is smaller than 10% for $n = 500$ and $n = 1000$. And the medians for $P1$ are both lower than 50%. The performance is remarkable enhanced when the sample size is increased. In particular, the first quartile for $P1$ with $n = 3000$ reports a much higher proportion. In comparison, the estimators have higher first quartile for $P2$ (the percentage of an estimated set covered by the identification region). The difference between $P1$ and $P2$ for higher quartiles are less pronounced.

Table 1 (b): Uniform Design

	$P1$			$P2$		
percentile	$n = 500$	$n = 1000$	$n = 3000$	$n = 500$	$n = 1000$	$n = 3000$
10%	0.581	0.627	0.814	0.509	0.635	0.786
25%	0.664	0.755	0.859	0.600	0.719	0.848
50%	0.782	0.854	0.911	0.685	0.807	0.908
75%	0.925	0.949	0.968	0.843	0.895	0.954
90%	0.989	0.984	0.994	0.923	0.965	0.982

Table 1(b) suggests the performance of the estimator under the uniform design is much better than under the normal design. This is best illustrated by lower percentiles for smaller sample sizes. The median for $P1$ and $P2$ are remarkably high for all sample sizes.

Table 1 (c): Min{P1,P2}

	Normal			*Uniform*		
percentile	$n = 500$	$n = 1000$	$n = 3000$	$n = 500$	$n = 1000$	$n = 3000$
10%	0	0	0.054	0.503	0.583	0.758
25%	0.013	0.086	0.341	0.582	0.673	0.817
50%	0.294	0.367	0.477	0.659	0.748	0.858
75%	0.453	0.541	0.618	0.756	0.818	0.902
90%	0.546	0.633	0.688	0.854	0.863	0.940

Another measure of the discrepancies between the estimates and the identification region is $\min\{P1, P2\}$ reported in *Table 1(c)*. By this criterion, the estimator also performs obviously better under the uniform design than under the normal design. Recall the consistency of our estimator was proved for the case where the support of X is bounded. Though we have argued that the boundedness of the support S_X may be stronger than necessary for consistency, these monte carlo simulation results seem to suggest our estimator works better in the uniform design where S_X is bounded than in the normal design where it is not.

7.2 Binary regressions with an endogenous regressor

Next consider a binary regression with one of the regressors simultaneously determined with the latent outcome. That is $Y_1 = 1(Y_1^* \geq 0)$, where

$$\begin{aligned} Y_1^* &= X_0\beta_{01} + X_1\beta_1 + Y_2\gamma_1 + \epsilon_1 \\ Y_2 &= X_0\beta_{02} + X_2\beta_2 + Y_1^*\gamma_2 + \epsilon_2 \end{aligned}$$

and Y_1, Y_2 are observed along with exogenous regressors X_0, X_1, X_2. We adopt the following values for the parameters: $\beta_{01} = 1$, $\beta_1 = 2$, $\gamma_1 = 1$, $\beta_{02} = 1/2$, $\beta_2 = -3$, $\gamma_2 = 2$. The exogenous regressors follow a multivariate normal distribution with a mean and covariance matrix in the following form:

$$\mu_X = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \; \Sigma_X = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & 0 \\ \rho\sigma^2 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$$

The private signals (ϵ_1, ϵ_2) follow a bivariate normal distribution with

$$\mu_\epsilon = [0, 0], \; \Sigma_\epsilon = \sigma_\epsilon \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; \sigma_\epsilon = |\bar{X}|$$

where $\bar{X}$ is the mean of X_0, X_1, X_2. We follow the criterion function approach proposed in Chernozhukov, Hong and Tamer (2003) to construct confidence regions for the linear coefficients using our extreme estimator. For any confidence level α, we use a subsampling procedure to approximate the distribution of $\sup_{b\in\Theta_I} n\hat{Q}_n(b)$. Once we have an estimate of the α-th quantile of this distribution (denoted $\hat{t}_\alpha$), the α-th confidence region for the true coefficient is defined as $\{b : n\hat{Q}_n(b) \leq \hat{t}_\alpha\}$.[7]

In the first step of estimation, the bandwidth for estimating $\hat{p}$ was chosen following standard cross-validation procedures using the $AMISE$ criterion. The resulting bandwidth is 0.82. In the second step, the identification region is computed using simple grid-search. The length of each grid in the two dimensions are both 1/50, and therefore the number of grid points can be used as a measure of the size of different estimated identification regions. For each set of parameter values in Table 2, we run $S = 200$ simulation, each generating a sample of $N = 3000$ observations. For each sample, we first estimate $\hat{\Theta}_I$ using our two-step extreme estimator, and then approximate the distribution of $\sup_{b\in\Theta_I} n\hat{Q}_n(b)$ by drawing $N_b = 200$ subsamples from that sample (each contains $b = 500$ observations), and calculating the supreme of our estimand (objective function) over $\hat{\Theta}_I$ in each of the subsamples. Then α-th quantile of the distribution of $\sup_{b\in\Theta_I} n\hat{Q}_n(b)$ is estimated by the α-th quantile out of the $N_b = 200$ extreme values (denoted $\hat{t}_\alpha$). Finally, the confidence region is constructed as $\{b : n\hat{Q}_n(b) \leq \hat{t}_\alpha\}$. Table 2 reports the actual coverage probabilities of these confidence regions, and the average size of the $S = 200$ estimated confidence regions in terms of the grid points contained.

Table 2: size and coverage probabilities of C.R.

Nominal CP	ρ	σ	*Actual C.P.*	*Avg. size (10^3 grid points)*
70%	0.5	1	59%	1.55
70%	0.5	2	84%	2.23
90%	0.5	1	89%	4.40
90%	0.5	2	94%	5.83
90%	-0.5	1	59%	2.45
90%	-0.5	2	91%	2.88

Estimated confidence regions seem to have smaller errors in coverage probabilities as the variance of the regressors increase relative to that of the unobserved disturbances.

[7] We follow an algorithm in CHT (2007) for constructing confidence regions that cover the *identification region* with the predetermined confidence level, rather than covering the *true parameter*. However, in our design here, the truce parameter is point-identified, and this distinction is superfluous.

8 Conclusion

We have studied the estimation of a class of binary response models where the conditional median of the disturbances is bounded between known (or exogenously identified) functions of the regressors. We focus on the case where the latent outcome is additively separable in a linear index subutility function and the disturbance. Though the index coefficients are not exactly identified in general, we can characterize their convex identification regions, and propose a two-step extreme estimator that estimates the identification region consistently regardless of point identification. More interestingly, we show how this approach provides a novel approach of inference of several important micro-econometric submodels under alternative (and sometimes weaker) assumptions that are not studied in the literature. We prove point-identification of the coefficients in these motivating submodels under fairly general restrictions on the structural primitives. Monte Carlo experiments in various designs also provide encouraging evidence of our estimator's performance in finite samples. Directions for future research include the search for conditions for point-identification when the latent outcome takes a more general form than linear indices. Besides, a priority for future study of this class of models is to derive the asymptotic distribution of the estimator when there is point-identification. Yet another interesting issue is the estimation of the model when the bounding functions L and U are only known up to certain parametric or shape restrictions.

9 Appendix: Proofs

Proof of Lemma 1. Suppose b is such that $\Pr(X \in \xi_b') > 0$, and let Γ denote the set of $F_{\epsilon|X}$ that satisfy BCQ. Then by definition, $\forall x \in \xi_b'$ $s.t.$ $-xb \leq L(x)$ and $p^*(x) < \frac{1}{2}$, $p(x;b,F_{\varepsilon|X}) = \int 1(\varepsilon \geq -xb)dF_{\varepsilon|X=x} \geq \frac{1}{2}$ $\forall F_{\varepsilon|X} \in \Gamma$. Likewise $\forall x \in \xi_b'$ $s.t.$ $-x^T b \geq U(x)$ and $p^*(x) > \frac{1}{2}$, $p(x;b,F_{\varepsilon|X}) = \int 1(\varepsilon \geq -xb)dF_{\varepsilon|X=x} \leq \frac{1}{2}$ $\forall F_{\varepsilon|X} \in \Gamma$. As a result $\forall x \in \xi_b'$, $p^*(x) \neq p(x;b,F_{\varepsilon|X})$ $\forall F_{\varepsilon|X} \in \Gamma$. Therefore $\Pr(X \in X(b,F_{\varepsilon|X})) > 0$ $\forall F_{\varepsilon|X} \in \Gamma$, and β is identified relative to b under BCQ. Now suppose b is such that $\Pr(X \in \xi_b') = 0$. Then $\Pr(X \in S_X \backslash \xi_b') = 1$ where $S_X\backslash\xi_b' \equiv \{x \in S_X : (-xb \leq L(x), p^*(x) \geq \frac{1}{2})$ *or* $(-xb \geq U(x), p^*(x) \leq \frac{1}{2})$ *or* $(-xb \in (L(x),U(x)))\}$. Then $\forall x$ $s.t. -xb \leq L(x), p^*(x) \geq \frac{1}{2}$, pick $F_{\epsilon|X=x}(.;b,p^*(x))$ $s.t.$ (i) $F_{\epsilon|\mathbf{X}=\mathbf{x}}$ is continuous in ε, $L(x) \leq \sup Med(\epsilon|x)$ and $\inf Med(\epsilon|x) \leq U(x)$; (ii) $\int 1(xb+\epsilon \geq 0)dF_{\epsilon|X=x} = p^*(x)$. This can be done because $-xb \leq L(x) \leq \sup Med(\epsilon|x)$ of $F_{\epsilon|X=x}$ requires $\int 1(xb+\epsilon \geq 0)dF_{\epsilon|X=x} \geq \frac{1}{2}$. Likewise $\forall x$ $s.t.$ $-xb \geq U(x)$ and $p^*(x) \leq \frac{1}{2}$, we can pick $F_{\epsilon|X=x}(.;b,p^*(x))$ $s.t.$ (i) holds and $\int 1(xb+\epsilon \geq 0)dF_{\epsilon|X=x} = p^*(x)$. And $\forall x$ $s.t.$ $L(x) < -xb < U(x)$, we can always pick $F_{\epsilon|X=x}(.;b,p^*(x))$ $s.t.$ $\int 1(xb+\epsilon \geq 0)dF_{\epsilon|X=x} = p^*(x)$ (regardless of the value of $p^*(x)$) while (i) still holds. Finally for a given $p^*(x)$ and any b such that $\Pr(X \in \xi_b') = 0$, let $F_{\epsilon|X}(.;b,p^*(x))$ be such that $F_{\epsilon|X=x}(.;b,p^*(x))$ is picked as above $\forall x \in S(X)\backslash\xi_b'$. We have shown $p(x;b,F_{\varepsilon|X}) = p^*(x)$ $\forall x \in S_X\backslash\xi_b$ (equivalent to $a.e.F_X$ as $\Pr(X \in \xi_b') = 0$). Hence $\exists F_{\varepsilon|X} \in \Gamma$ $s.t.$ $\Pr(X \in X(b,F_{\varepsilon|X})) = 0$, and b is observationally equivalent to β under BCQ.

Proof of Corollary 1. The proof is similar to *Lemma 1* and is omitted for brevity.

Proof of Corollary 2. Suppose $b_1 \in \Theta_I', b_2 \in \Theta_I'$. Then $\Pr(X \in \xi_{b_1}') = \Pr(X \in \xi_{b_2}') = 0$. Let $b_\alpha \equiv \alpha b_1 + (1-\alpha)b_2 \in \Theta_I'$ for some $\alpha \in (0,1)$ and ξ_{b_α}' be defined as before for b_α. Note $\forall x \in \xi_{b_\alpha}'$, either $(-xb_\alpha \geq U(x), p^*(x) > 1/2)$ or $(-xb_\alpha \leq L(x), p^*(x) < 1/2)$. Consider the former case. Then it must be $p^*(x) > 1/2$ and "either $-xb_1 \geq U(x)$ or $-xb_2 \geq U(x)$". This implies either $x \in \xi_{b_1}'$ or $x \in \xi_{b_2}'$. Symmetric argument applies to the case $(-xb_\alpha \leq L(x), p^*(x) < 1/2)$. It follows that $\xi_{b_\alpha}' \subseteq (\xi_{b_1}' \cup \xi_{b_2}')$. Then $\Pr(X \in \xi_{b_\alpha}') \leq \Pr(X \in \xi_{b_1}') + \Pr(X \in \xi_{b_2}') = 0$, and $b_\alpha \in \Theta_I'$. The proof of the convexity of Θ_I follows from similar arguments and is omitted for brevity.

Proof of Lemma 2. By construction, $Q(b)$ is non-negative $\forall b \in \Theta$. By the law of iterated expectations,

$$\begin{aligned} Q(b) &= E[(-U(X)-Xb)_+^2 \mid p^*(X)>1/2]\Pr(p^*(X)>1/2) \\ + & \quad E[(-L(X)-Xb)_-^2 \mid p^*(X)<1/2]\Pr(p^*(X)<1/2) \\ + & \quad E[(-L(X)-Xb)_-^2 + (-U(X)-Xb)_+^2 \mid p^*(X)=1/2]\Pr(p^*(X)=1/2) \end{aligned}$$

By definition $\forall b \in \Theta_I$, all of the four following events must have zero probability

$$\begin{aligned} \text{"} -Xb &\geq U(X),\ p^*(X)>1/2\text{"} \qquad \text{"} -Xb \leq L(X),\ p^*(X)<1/2\text{"} \\ \text{"} -Xb &< L(X),\ p^*(X)=1/2\text{"} \qquad \text{"} -Xb > U(X),\ p^*(X)=1/2\text{"} \end{aligned}$$

Therefore $Q(b)=0$ for all $b \in \Theta_I$. On the other hand, for any $b \notin \Theta_I$, at least one of the four events above must have positive probability. Without loss of generality, let the first event occur with positive probability. Then $\Pr\{-X'b=U(X)\}=0$ ensures $\Pr\{-Xb > U(X),\ p^*(X)>1/2\}>0$, which implies the first term in $Q(b)$ will be strictly positive. Similar arguments can be applied to prove $Q(b)>0$ for $b \notin \Theta_I$ if any of the other events also has positive probability.

Proof of Corollary 3. By construction, $Q(b)$ is non-negative $\forall b \in \Theta$. By the law of iterated expectations and the condition that $\Pr(p^*(X)=\frac{1}{2})=0$,

$$\begin{aligned} Q(b) &= E[\Lambda(p^*(X)-\frac{1}{2})(-U(X)-Xb)_+^2 \mid p^*(X)>1/2]\Pr(p^*(X)>1/2) \\ + & \quad E[\Lambda(\frac{1}{2}-p^*(X))(-L(X)-Xb)_-^2 \mid p^*(X)<1/2]\Pr(p^*(X)<1/2) \end{aligned}$$

By definition $\forall b \in \Theta_I$, both of the following events must have zero probability

$$\text{"} -Xb \geq U(X),\ p^*(X)>1/2\text{"} \qquad \text{"} -Xb \leq L(X),\ p^*(X)<1/2\text{"}$$

Note the two events in the proof of *Lemma 2* with $p^*(X)=\frac{1}{2}$ need not be addressed under current regularity condition. Therefore $Q(b)=0$ for all $b \in \Theta_I$. On the other hand, for any $b \notin \Theta_I$, at least one of the two events above must have positive probability. Without loss of generality, let the first event occur with positive probability. Then $\Pr\{-X'b=U(X)\}=0$ implies $\Pr\{-Xb>U(X),\ p^*(X)>1/2\}>0$, which implies the first term in $Q(b)$ is strictly positive for any $b \notin \Theta_I$. Similar arguments can be applied to prove the second term in $Q(b)$ is strictly positive for all $b \notin \Theta_I$ if the other event has positive probability.

Proof of Proposition 1. First, we show $\sup_{b\in\Theta}|\hat{Q}_n(b)-Q(b)| \xrightarrow{p} 0$. By Lemma 8.10 in Newey and McFadden (1994), under *RD*, *TF* and *K*,

$$\sup_{x\in S_X}|\hat{p}(x)-p^*(x)| = o_p(n^{-1/4}) \tag{6}$$

Apply a mean value expansion of $\Lambda(\hat{p}(x)-\frac{1}{2})$, $\Lambda(\frac{1}{2}-\hat{p}(x))$, $(-\hat{L}(x)-xb)_-^2$ and $(-\hat{U}(x)-xb)_+^2$ around $\Lambda(p^*(x)-\frac{1}{2})$, $\Lambda(\frac{1}{2}-p^*(x))$, $(-L(x)-xb)_-^2$ and $(-U(x)-xb)_+^2$ respectively. Then by the Law of Large Numbers, the uniform convergence of $\hat{p}$ in (6) and $\hat{L},\hat{U}$ in BF, as well as the uniform boundedness of Λ ,$(-L(x)-xb)_-^2$ and $(-U(x)-xb)_+^2$ over S_X and Θ, we have $\hat{Q}_n(b)\xrightarrow{p}Q(b)\ \forall b\in\tilde{\Theta}$. Now note $\hat{Q}_n(b)$ is continuous and convex in b over Θ for all n. Convexity is preserved by pointwise limits, and hence Q_0 is also convex and therefore continuous on the interior of Θ. Furthermore, by Andersen and Gill (1982) (and Theorem 2.7 in Newey and McFadden (1994)), the convergence in probability of $\hat{Q}_n(b)$ to $Q(b)$ must be uniform over Θ. The rest of the proof follows from arguments in Proposition 3 in Manski and Tamer (2002) and that $\hat{Q}_n$ converges to Q uniformly over Θ at a rate faster than $n^{-1/4}$, and and is omitted for brevity.

Proof of Proposition 2. By *BCQ-2* and *Lemma 1*, it suffices to show $\Pr(X\in\xi_b)>0$ for all $b\neq\beta_0$, where $\xi_b\equiv\{x:(-xb\leq L(x),-x\beta_0>U(x))\ or\ (-xb\geq U(x),-x\beta_0<L(x))\}$. By *SX-(a)*, $\Pr(X_{-J}(\beta_{0,-J}-b_{-J})\neq 0)>0$. Without loss of generality, let $\Pr(X_{-J}\beta_{0,-J}<X_{-J}b_{-J})>0$. Then by *EX* and *SX-(b),(c)*, $\Pr(-X\beta_0<L(X)\leq U(X)<-Xb)>0$.

Proof of Lemma 3. *(Sufficiency)* Suppose $\theta_i\overset{u.o.e.}{\sim}\theta_i^0$ under $\mathcal{F}_{MI}^i$ for $i=1,2$. By definition $\exists\bar{F}_{\epsilon_i|X}\in\mathcal{F}_{MI}^i$ such that $\Pr\{p_i^*(X)=\bar{F}_{\epsilon_i|X}(X\beta_i+p_{-i}^*(X)\delta_i)\}=1$ for $i=1,2$. Hence $\Pr(p^*(X)\in\Upsilon(X;\theta,\bar{F}_{\epsilon|X}))=1$ where $\bar{F}_{\epsilon|X}\equiv\prod_{i=1,2}\bar{F}_{\epsilon_i|X}\in\mathcal{F}$, and $\theta\overset{o.e.}{\sim}\theta^0$ under $\mathcal{F}_{CMI}$. *(Necessity)* That $\theta\overset{o.e.}{\sim}\theta^0$ under $\mathcal{F}_{CMI}$ implies $\exists\bar{F}_{\epsilon|X}\in\mathcal{F}_{CMI}$ such that $\Pr\{p^*(X)\in\Upsilon(x;\theta,\bar{F}_{\epsilon|X})\}=1$. It follows that $\Pr\{p_i^*(X)=\bar{F}_{\epsilon_i|X}(X\beta_i+p_{-i}^*(X)\delta_i)\}=1$ for $i=1,2$, where $\bar{F}_{\epsilon_i|X}$ are marginal distributions corresponding to $\bar{F}_{\epsilon|X}$. By definition, this means $\theta_i\overset{u.o.e.}{\sim}\theta_i^0$ under $\mathcal{F}_{MI}^i$ for both $i=1,2$.

Proof of Lemma 4. We prove the lemma from the perspective of player 1. The proof for the case of player 2 follows from the same argument. Fix $\bar{x}_{-h_1}\in S(X_{-h_1})$. By definition of a BNE,

$$\begin{bmatrix}p_1(\bar{x}_{-h_1},x_{h_1})\\p_2(\bar{x}_{-h_1},x_{h_1})\end{bmatrix}=\begin{bmatrix}F_{\epsilon_1|\bar{x}_{-h_1},x_{h_1}}(x_{-h_1}b_{1,-h_1}+x_{h_1}b_{1,h_1}-p_2(\bar{x}_{-h_1},x_{h_1}))\\F_{\epsilon_2|\bar{x}_{-h_1},x_{h_1}}(x_{-h_1}b_{2,-h_1}+x_{h_1}b_{2,h_1}-p_1(\bar{x}_{-h_1},x_{h_1}))\end{bmatrix}\tag{7}$$

Let $C[S(X_{h_1}|\bar{x}_{-h_1})]$ denote the space of bounded, continuous functions on the compact support $S(X_{h_1}|\bar{x}_{-h_1})$ under the sup-norm. By standard arguments, $C[S(X_{h_1}|\bar{x}_{-h_1})]$ is a Banach Space. Define $C^{K_1}(\bar{x}_{-h_1})$ as a subset of functions in $C[S(X_{h_1}|\bar{x}_{-h_1})]$ that map from $S(X_{h_1}|\bar{x}_{-h_1})$ to $[0,1]$, and are Lipschitz continuous with some constant $k\leq K_1$. Then

$C^{K_1}(\bar{x}_{-h_1})$ is bounded in the sup-norm and equicontinuous by the Lipschitz continuity. Note $C^{K_1}(\bar{x}_{-h_1})$ is also closed in $C[S(X_{h_1}|\bar{x}_{-h_1})]$. To see this, consider a sequence f_n in $C^{K_1}(\bar{x}_{-h_1})$ that converges in the sup-norm to f_0. By the completeness of $C[S(X_{h_1}|\bar{x}_{-h_1})]$, $f_0 \in C[S(X_{h_1}|\bar{x}_{-h_1})]$. Now suppose $f_0 \notin C^{K_1}(\bar{x}_{-h_1})$. Then $\exists x^a_{h_1}, x^b_{h_1} \in S(X_{h_1}|\bar{x}_{-h_1})$ such that $|f_0(x^a_{h_1}) - f_0(x^b_{h_1})| > K'_1|x^a_{h_1} - x^b_{h_1}|$ for some $K'_1 > K_1$. By convergence of f_n, for all $\varepsilon > 0$, $|f_n(x^j_{h_1}) - f_0(x^j_{h_1})| \leq \frac{\varepsilon}{2}|x^a_{h_1} - x^b_{h_1}|$ for $j = a, b$ for n big enough. Hence $\frac{|f_n(x^a_{h_1}) - f_n(x^b_{h_1})|}{|x^a_{h_1} - x^b_{h_1}|} > K'_1 - \varepsilon$ for n big enough. For any $\varepsilon < K'_1 - K_1$, this implies for n big enough, f_n is not Lipschitz continuous with $k \leq K_1$. Contradiction. Hence $C^{K_1}(\bar{x}_{-h_1})$ is bounded, equicontinuous, and closed in $C[S(X_{h_1}|\bar{x}_{-h_1})]$. By the Arzela-Ascoli Theorem, $C^{K_1}(\bar{x}_{-h_1})$ is a convex, compact subset of the normed linear space $C[S(X_{h_1}|\bar{x}_{-h_1})]$. Now substitute the second equation in (7) into the first one, and we have

$$\bar{p}_1(x_{h_1}) = \bar{F}_{\epsilon_1|x_{h_1}}\{x'b_1 - \bar{F}_{\epsilon_2|x_{h_1}}[x'b_2 - \bar{p}_1(x_{h_1})]\} \tag{8}$$

where $\bar{p}_1$ and $\bar{F}_{\epsilon_i|x_{h_1}}$ are shorthand notations for conditioning on $\bar{x}_{-h_1}$. Let $\tau(x_{h_1})$ denote the right-hand side of (8). Suppose $\bar{p}_1(x_{h_1})$ is Lipschitz continuous with constant $k \leq K_1$ for some $K_1 > 0$. Then by the definition of the Lipschitz constants in *DDF (i)-(ii)*, for all x_{h_1}, $\tilde{x}_{h1} \in S(X_{h_1}|\bar{x}_{-h_1})$, $|\tau(x_{h_1}) - \tau(\tilde{x}_{h_1})| \leq D(K_1)|x_{h_1} - \tilde{x}_{h_1}|$, where

$$D(K_1) \equiv K^1_{F_1} + (|b_{2,h_1}|C_{F_2} + K_1 C_{F_2} + K^1_{F_2})C_{F_1}$$

Since $b_{2,h_1} \neq 0$ and $|C_{F_1}C_{F_2}| < 1$, K_1 can be chosen such that $D(K_1) \leq K_1$. Therefore the right hand side of (8) is a continuous self-mapping from $C^{K_1}(\bar{x}_{-h_1})$ to $C^{K_1}(\bar{x}_{-h_1})$ for the K_1 chosen. It follows from Schauder's Fixed Point Theorem that the solution $p_1(X_{h_1,}\bar{x}_{-h_1})$ is continuous in X_{h_1} for all $\bar{x}_{-h_1} \in S(X_{-h_1})$.

Proof of Lemma 5. Fix $\bar{x}_{-h_1} \in S(X_{-h_1})$ from the perspective of player 1, then for any $b \equiv (b_1, b_2)$ in the parameter space $\Theta \equiv \Theta_1 \otimes \Theta_2$, all $x_{h_1} \in S(X_{h_1}|x_{-h_1})$,

$$\bar{p}_2(x_{h_1}) = \bar{F}_{\epsilon_2|x_{h_1}}[x'b_2 - \bar{p}_1(x_{h_1})] \tag{9}$$

where $\bar{p}_i$ and $\bar{F}_{\epsilon_2|x_{h_1}}$ are both conditioned on $\bar{x}_{-h_1}$. Lemma 4 has shown $\bar{p}_1(x_{h_1})$ is Lipschitz continuous with a certain constant on the compact, connected support $S(X_{h_i}|\bar{x}_{-h_i})$ for all $b \in \Theta$. By similar arguments, it follows from *REG* and *DDF* that $\bar{p}_2(x_{h_1})$ as defined in (9) is also Lipschitz continuous on $S(X_{h_1}|\bar{x}_{-h_1})$ for all $b \in \Theta$. Then *REG* implies the image of $\bar{p}_2(x_{h_1})$ is the connected interval $(0,1)$ for all $b \in \Theta$. Hence $\Pr\{\bar{p}^*_2(X_{h_1}) \in I_{b,\beta^0}(X)|\bar{x}_{-h_i}\} > 0$ for all $\bar{x}_{-h_i} \in S(X_{-h_i})$ and $b \neq \beta^0$ in Θ, where $\bar{p}^*_2$ is the equilibrium outcome under the truth β^0, and $I_{b,\beta^0}(X)$ is the intersection of $(0,1)$ with the random interval between Xb and $X\beta^0$. Proof from the perspective of player 2 follows from similar arguments.

Proof of Proposition 3. We prove for the case of player 1. The case for player 2 follows from symmetric arguments. Note all any $b \in \Theta$, X_{-h_1} impacts $p_2(X;b)$ and $p_1(X;b)$ but not $X'b_1$ (as $b_{1,h_1} = 0$). By *RSX-(i)*, $\Pr\{X'_{-h_1}(b_{1,-h_1} - \beta^0_{1,-h_1}) \neq 0\} > 0$ for all $b_1 \neq \beta^0_1$ in Θ_1. Hence *SX-1 (a)* of Proposition 2 is satisfied. Suppose

$$\Pr\{X'_{-h_1}(b_{-h_1} - \beta^0_{-h_1}) \neq 0, sgn(X'_{-h_1}b_{-h_1}) \neq sgn(X'_{-h_1}\beta^0_{-h_1})\} > 0$$

then *SX-(b)* is satisfied. Otherwise, without loss of generality, consider the case $\Pr(X'_{-h_1}b_{1,-h_1} > X'_{-h_1}{\beta^0}_{1,-h_1} > 0) > 0$. Then *RSX-(ii)* and the closedness under scalar multiplications in *RSX-(iii)* guarantee that *SX-(b)* is satisfied. Let p^* denote the true equilibrium outcome induced by β^0, and let $\bar{p}^*_1(x_{h_1})$ be a shorthand for conditioning on $\bar{x}_{-h_1}$. By Lemma 4, $\bar{p}^*_1(x_{h_1})$ is continuous in x_{h_1} on $S(X_{h_1}|\bar{x}_{-h_1})$ for all $\bar{x}_{-h_1} \in S(X_{-h_1})$. By Lemma 5, $\Pr\{\bar{p}^*_2(X_{h_1}) \in (a_1, a_2)|\bar{x}_{-h_1}\} > 0$ for all $(a_1, a_2) \subset [0,1]$ and all $\bar{x}_{-h_i} \in S(X_{-h_i})$ under *REG*. It follows immediately that *SX-(c)* is satisfied.

Proof of Proposition 4. We first prove the reduced form coefficients are identified under $\mathcal{F}_{CSI}$. Let $F^0_{\epsilon_1,\epsilon_2|\tilde{X}}$ denote the true distribution of structural errors. For any θ in Θ, observational equivalence to θ_0 under $\mathcal{F}_{CSI}$ implies $E(Y_2|\tilde{X};\theta_0, F^0_{\epsilon_1,\epsilon_2|\tilde{X}}) = E(Y_2|\tilde{X};\theta, F_{\epsilon_1,\epsilon_2|\tilde{X}})$ for some $F_{\epsilon_1,\epsilon_2|\tilde{X}} \in \mathcal{F}_{CSI}$. This in turn implies "$\tilde{X}'\delta(\theta_0) = \tilde{X}'\delta(\theta)$ almost everywhere on $S_{\tilde{X}}$", where $\delta(\theta_0)$, $\delta(\theta)$ are reduced form coefficients in $\mathbb{R}^{K_0+K_1+K_2}$ corresponding to θ_0, θ respectively. By SIM-1, 2, this event is only possible when $\delta(\theta_0) = \delta(\theta)$. Hence $\delta(\theta_0)$ is identified under $\mathcal{F}_{CSI}$. Next we argue β_1, β_{01} must also be among the identified features under $\mathcal{F}_{CSI}$. Suppose not. Then $\exists\theta \in \Theta$ such that θ is observational equivalent to the truth θ^0 under $\mathcal{F}_{CSI}$, but its components β_1, β_{01} are not equal to the truth's component β^0_1, β^0_{01}. Note SIM-2 also implies "support of (X_0, X_1) (denoted $S_{X_0,X_1} \subseteq \mathbb{R}^{K_0+K_1}$) is not contained in any linear subspace". Then it must be the case $\Pr(X_0\beta_{01} + X_1\beta_1 \neq X_0\beta^0_{01} + X_1\beta^0_1) > 0$. Without loss of generality, let the event $\Omega_{NE} \equiv$ "$X_0\beta_{01} + X_1\beta_1 > X_0\beta^0_{01} + X_1\beta^0_1$" happen with positive probability. Recall δ_0, δ_1 are among the identified features under $\mathcal{F}_{CSI}$. Thus for all x_0, x_1 such that Ω_{NE} occurs, SIM-3 implies

$$\Pr([x_0\ x_1]\begin{bmatrix} \beta_{01} - \delta_0 \\ \beta_1 - \delta_1 \end{bmatrix} > X_2\delta_2 > [x_0\ x_1]\begin{bmatrix} \beta^0_{01} - \delta_0 \\ \beta^0_1 - \delta_1 \end{bmatrix} |x_0, x_1) > 0$$

Let $F_{u|\tilde{X}}(.;\theta^0, F^0_{\epsilon_1,\epsilon_2|\tilde{X}})$ denote the true semi-reduced form error distribution in (4), and $F_{u|\tilde{X}}(.;\theta, F_{\epsilon_1,\epsilon_2|\tilde{X}})$ denote its analog derived from $\theta, F_{\epsilon_1,\epsilon_2|\tilde{X}}$ that is observationally equivalent to the truth $\theta^0, F^0_{\epsilon_1,\epsilon_2|\tilde{X}}$ under $\mathcal{F}_{CSI}$. By construction, both $F_{u|\tilde{X}}(.;\theta^0, F^0_{\epsilon_1,\epsilon_2|\tilde{X}})$ and $F_{u|\tilde{X}}(.;\theta, F_{\epsilon_1,\epsilon_2|\tilde{X}})$ are symmetric around zero conditional on $\tilde{X}$. It then follows that the event

$$
\begin{aligned}
" \quad & \Pr(Y_1 = 1|X_1, X_2, E(Y_2|\tilde{X}); \beta_{01}, \beta_1, F_{u|\tilde{X}}(.; \theta, F_{\epsilon_1,\epsilon_2|\tilde{X}})) \\
\neq \quad & \Pr(Y_1 = 1|X_1, X_2, E(Y_2|\tilde{X}); \beta_{01}^0, \beta_1^0, F_{u|\tilde{X}}^0(.; \theta^0, F_{\epsilon_1,\epsilon_2|\tilde{X}}^0))"
\end{aligned}
$$

happens with positive probability. This contradicts our supposition that θ and θ^0 are observationally equivalent under $\mathcal{F}_{CSI}$. With β_{01}, β_1 now identified and γ_1 already normalized to 1, the other structural coefficients are recovered from reduced form coefficients as $\gamma_2 = \frac{\delta_1}{\beta_1+\delta_1}$, $\beta_2 = (1-\gamma_2)\delta_2$, $\beta_{02} = (1-\gamma_2)\delta_0 - \gamma_2\beta_{01}$.

References

[1] Aradillas-López, A. (2007): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information", Working Paper, Princeton University

[2] Bajari, P., Hong, H., Krainer, J. and D. Nekipelov (2007): "Estimating Static Models of Strategic Interactions", Working Paper, University of Minnesota

[3] Blundell, R. and J. Powell (2004): "Endogeneity in Semiparametric Binary Response Models", Review of Economic Studies 71, p655-679

[4] Buchinsky, M. and J. Hahn (1998): "An Alternative Estimator for the Censored Quantile Regression Model", Econometrica 66, p653-672.

[5] Chamberlain, G. (1986): "Asymptotic Efficiency in Semi-Parametric Models with Censoring", Journal of Econometrics, Volume 32, Issue 2, July 1986, Pages 189-218

[6] Chernozhukov, V., Tamer, E. and H. Hong, "Estimation and Inference in Partially Identified Models", Econometrica 2008

[7] Hoderlein, S. and X. Tang (2008), "Binary Response with Endogeneity: A New Perspective from Median Independence", Brown University and University of Pennsylvania

[8] Horowitz, J. (1992):"A Smoothed Maximum Score Estimator for the Binary Response Model," Econometrica, Econometric Society, 60(3), 505-31

[9] Manski, C. (1985): "Semiparametric Analysis of Random Effects Linear Models From Binary Panel Data," Econometrica, 55(2), 357-362

[10] Manski, C. (1988): "Identification of Binary Response Models ", Journal of American Statistical Association, Vol 83, No. 403, p729-738

[11] Manski, C. and Tamer, E.(2002): "Inference on Regressions with Interval Data on a Regressor or Outcome", Econometrica, 70, 519-547

[12] Newey, W.K. and McFadden, D (1994): "Large Sample Estimation and Hypothesis Testing", Handbook of Econometrics, Volume IV, Chapter 36, Elsevier Science

[13] Pagan and Ullah (1999): "Nonparametric Econometrics", Cambridge University Press.

[14] Pollard, D. (1991): "Asymptotics for Least Absolute Deviation Regression Estimators", Econometric Theory, 7, 1991, p186-199