Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 09-003

"Binary Regressions with Bounded Median Dependence"

by

Xun Tang

Binary Regressions with Bounded Median Dependence

Xun TANG[1]

Department of Economics

University of Pennsylvania

January 20, 2009

Abstract

In this paper we study the identification and estimation of a class of binary regressions 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 censored regressors, and binary regressions with interval data or measurement errors on regressors. We characterize the identification region of linear coefficients 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 novel, two-step smooth extreme estimator, and prove its consistency for the identification region of coefficients. We also provide encouraging Monte Carlo evidence of the estimator's performance in finite samples.

KEYWORDS: Binary response, median dependence, games with incomplete information, censored regressors, interval data, measurement error, partial identification, point identification, consistent estimation

JEL CODES: C14, C25, C51

---

[1] This paper develops from a chapter of my Ph.D. dissertation at Northwestern University. I am grateful to Elie Tamer for helpful discussions and advice. I also thank Andrés Aradillas-López, Hanming Fang, Stefan Hoderlein, Bo Honoré, Joel Horowitz, Charles Manski, Rosa Matzkin, Ulrich Müller, Aureo De Paula, Robert Porter, Kyungchul Song, Petra Todd and the econometrics seminar participants at Northwestern, Princeton, 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 relax the assumption of median independence of unobservable disturbances. Specifically, the latent outcome in the binary decision is the sum of a linear index function of regressors and a 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 are median independent of the regressors; (ii) binary response models with a censored regressor, where additive structural errors (both in the binary regression and in a latent censored regression) are symmetric around zero and independent of each other conditional on perfectly observed regressors; (iii) binary regressions with interval data on one of the regressors, where the error is median independent of all other perfectly observed regressors and the interval; (iv) binary regressions with a noisy measure of one of the regressors, where the noise in the measurement and the structural error are conditionally independent of each other and median independent of all regressors. Our approach of estimation is novel in the sense that restrictions required to formulate (and identify) these models as binary regressions 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 coefficients, and derive sufficient conditions for point identification in the motivating models above. Remarkably exact identification can be achieved under fairly general exclusion restrictions on linear indices and some conditions on richness of the support of regressors. Our third contribution is to propose a novel two-step extreme estimator of the identified set of coefficients and prove its consistency. In the first-step, we use kernel regressions to estimate choice probabilities conditional on regressors. 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 region with positive numbers. The estimator is defined as minimizers of this sample analog. The estimator is consistent for the identification region when coefficients are only partially identified. We also give some encouraging Monte Carlo evidence on the estimator's performance in finite samples in two interesting designs.

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 shows how various micro-econometric models can be included into this general class under appropriate restrictions. Section 4 characterizes the identification region of index coefficients, and proves its convexity. Section 5 defines the two-step extreme estimator of the identification region and proves its consistency. Section 6 specifies sufficient conditions on structural primitives in motivating submodels that lead to the point-identification of the coefficients. Section 7 shows 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 regressions where the latent outcome variable is 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 for identification and estimation. A most popular identifying assumption is that errors are statistically independent of regressors. Matzkin (1992) showed under this assumption that a general subutility function $u(X)$ and distribution of the additive error term F_ϵ can be uniquely recovered up to a locational normalization from choice probabilities under 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 (1993)). 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, provided the the support of regressors is rich enough, 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)). Yet another branch of the literature studied the semiparametric efficiency of binary choice models under various stochastic restrictions. Chamberlain (1986) and Cosslett (1987) derived semiparametric efficiency bounds for binary choice models under independence, and Chamberlain (1986) concluded there does not exist any root-N consistent regular estimator for linear coefficients under the median independence restriction. Chen and Khan (2003) further showed that in the presence of multiplicative heterogeneity, the semiparametric information bound is zero even when the homoskedastic component is parametrically specified.

Our contribution to the literature on binary choices is the identification and consistent estimation of coefficients where conditional medians of the error term is only known to be bounded within certain ranges. This is a weak stochastic restriction that arises naturally in various important micro-econometric models.

Our paper is also related to previous works on various motivating models listed above. Aradillas-Lopez (2005) 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 (2005), 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 (2005) 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 different from that in Aradillas-Lopez (2005), and are valid in the presence of multiple equilibria provided all players observed in data follow the same pure-strategy Bayesian Nash Equilibria. Bajari, Hong, Krainer and Nekipelov (2007) does not impose any restrictions on how players' payoffs depend on observable states or on 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 their distributions must be known to the researcher. In contrast, we are less restrictive about the unobservable distributions of private signals, while the identifying power in our approach derives from the additive form of the payoff functions. Manski and Tamer (2002) studied 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. Rigobon and Stocker (2007) studied the estimation of a linear regression model where one of the regressors are censored. They established that there is zero semiparametric information in observations with a censored regressor, and thus verified there is no fully nonparametric "fix" for estimation of a multiple regression with censored regressors. Their work suggests more structure is required to address censored regressors even for the simplest case of linear regressions. Our work fills in the gap for binary regressions with censored regressors by modelling them as outcomes in a censored regression with additive errors that satisfy weak stochastic restrictions mentioned above.

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 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 as we argue below, this framework is general enough to include several interesting 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 that δ_i is strictly negative for $i = 1, 2$, ϵ_1 is independent of ϵ_2 conditional on X, and $Med(\epsilon_i|X = x) = 0$ for $i = 1, 2$ for all x.

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 $g_i : S_X \otimes S_{\epsilon_i}(X) \rightarrow \{0, 1\}$. A *pure-strategy Bayesian Nash equilibrium* (BNE) is characterized by a pair of set-valued functions $A_i : S_X \rightarrow S_{\epsilon_i}(X)$ 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. Under conditional independence, a generic value of parameters θ, $F_{\epsilon_1,\epsilon_2|X}$ 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 processes for $i = 1, 2$ fit 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, we define and prove identification

[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 $A_i^*(x) \equiv \{\varepsilon_i : \varepsilon_i \in A_i(\varepsilon_i, x)\}$.

of (β_1, β_2) under the conditional independence and median independence, an exclusion restriction in the indices $X\beta_i$, some support conditions on X, $L(X)$ and $U(X)$, as well as some other regularity conditions on primitives.

Several recent literature have studied the estimation of such static discrete games with incomplete information, including Aradillas-Lopez (2005) 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 needed for a well-defined likelihood function. Aradillas-Lopez gives sufficient and necessary conditions for the uniqueness of the equilibrium. Bajari et.al (2007) show a general subutility function $u(.)$ 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, distribution of the disturbances is rarely known to researchers in empirical implementations.

In comparison, our approach of formulating the BNE as a system of two binary regressions with observable 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 related to observable states. Second, multiplicity of the equilibria is not an issue for estimating the model, as recoverability of parameters does not hinge on the knowledge of a well-defined likelihood function. Instead, identification is solely based on the characterization of the outcome in (2) which is shared by all equilibria. In the presence of multiple pure-strategy Bayesian equilibria, we only require players in the data to follow the same pure strategy continuous in observable states x. We remain agnostic about the equilibrium selection mechanism itself.

3.2 Binary response with imperfect data on regressors

(Binary response with censored regressors) Let $Y_1 = 1(Y_1^* \geq 0)$, where Y_1^* is an unobserved latent scalar variable

$$Y_1^* = X_1\beta_1 + Y_2\gamma_1 + \epsilon_1 \tag{3}$$

and $Y_2 = X_2\beta_2 + \epsilon_2$, where $Y_2 \in \mathbb{R}^1$, $X_1 \in \mathbb{R}^{K_1}$, $X_2 \in \mathbb{R}^{K_2}$ and $\gamma_1 > 0$. Researchers can perfectly observe Y_1 and (X_1, X_2), but can only observe censored values of Y_2, i.e. $\tilde{Y}_2 \equiv \max\{Y_2, 0\}$. There may be overlapping coordinates in X_1 and X_2, and X_2 has at least one coordinate that is not included in X_1. This type of models arise frequently in empirical applications. One example is individuals' decisions on labor participation, where Y_1^* is the net number of hours an individual can supply after subtracting certain measure of reservation hours. The covariates in X_1 may include social demographic variables and log hourly wage, and Y_2 is the log income from other non-work sources, such as endowment, welfare benefits or spouse income. Often researchers only get to observe censored data on such income. This happens when people are reluctant to reveal exact amount of income when it is above or below certain levels. X_2 may overlap with X_1, but contains at least one element excluded from X_1. For example, X_2 may include demographic variables for the individual's spouse.

The key restriction for point identification is that conditional on (X_1, X_2), structural errors (ϵ_1, ϵ_2) are independent of each other and are both symmetric around 0. First, note (β_1, γ_1) can only be identified up to scale under these restrictions. To see this, recall the joint distribution $F_{\tilde{Y}_2, X_2}$ is sufficient for identifying β_2 under the symmetry of $F_{\epsilon_2|X_2}$ (as in Powell(1984), Powell(1986)). Now let $p_s(t|x_1, x_2) \equiv \Pr(Y_1 = s, \tilde{Y}_2 = t|\ X_1 = x_1,\ X_2 = x_2)$ for $s = 0, 1$. For any given set of parameters $(\beta_1, \beta_2, \gamma_1, F_{\epsilon_1|X_1,X_2}, F_{\epsilon_2|X_1,X_2})$ and $t > 0$,

$$\begin{aligned} p_1(t|x_1, x_2) &= F_{\epsilon_1|X_1=x_1,X_2=x_2,\epsilon_2=t-x_2\beta_2}(-x_1\beta_1 - t\gamma_1) f_{\epsilon_2|X_1=x_1,X_2=x_2}(t - x_2\beta_2) \\ &= F_{\epsilon_1|X_1=x_1,X_2=x_2}(-x_1\beta_1 - t\gamma_1) f_{\epsilon_2|X_1=x_1,X_2=x_2}(t - x_2\beta_2) \end{aligned} \quad (4)$$

and for $t = 0$,

$$\begin{aligned} p_1(0|x_1, x_2) &= F_{\epsilon_1,\epsilon_2|X_1=x_1,X_2=x_2}(-x_1\beta_1, -x_2\beta_2) \\ &= F_{\epsilon_1|X_1=x_1,X_2=x_2}(-x_1\beta_1) F_{\epsilon_2|X_1=x_1,X_2=x_2}(-x_2\beta_2) \end{aligned} \quad (5)$$

where the second equalities in both (4) and (5) follow from conditional independence. Consider an alternative set of parameters $(c\beta_1, \beta_2, c\gamma_1, F_{\epsilon_1|X_1,X_2}(\frac{\cdot}{c}), F_{\epsilon_2|X_1,X_2})$ where c is some positive constant. By construction, this alternative set generates the same observable distributions $p_1(.|x_1, x_2)$ on $[0, +\infty)$ for all x_1, x_2 as the original set of parameters.

After normalizing γ_1 to 1, (3) can be equivalently represented as

$$Y_1^* = X_1\beta_1 + Median(Y_2|X_1, X_2) + u \quad (6)$$

where $u = \epsilon_2 + \epsilon_1$. By construction, $Median(Y_2|X_1, X_2) = X_2\beta_2$. As β_2 is exactly identified under conditional symmetry of ϵ_2 given (X_1, X_2), the binary decision in (3) with censored values of Y_2 can be equivalently represented as (6) with u symmetrically distributed around 0 conditional on (X_1, X_2). By defining $\tilde{\epsilon} = Med(Y_2|X_2) + u$, the model fits in our general class of binary responses with $L(X_1, X_2) = U(X_1, X_2) = Med(Y_2|X_2)$.

(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 this model fits in our class of binary regressions with bounded conditional medians. The identifying restrictions here are 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 true (but unobservable) regressor V, the disturbance ϵ is statistically independent from the random bounds (V_0, V_1). Among other things, this conditional independence 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, and allows for such relations.

(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. $X_2 = X_2^* - \eta$ is observed instead of X_2^*. Suppose 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 individuals' 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 which affect the net payoffs from labour participation are mutually independent and both symmetric around 0 given 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 $\Pr(d=1|x;\beta_0,F^0_{\epsilon|X})$. Below 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 in the parameter space that are observationally equivalent to β_0 under BCQ.*

By construction, the size of the identification region decreases as stronger restrictions are imposed on the distribution of unobserved disturbances $F_{\epsilon|X}$. *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*

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

That $F_{\epsilon|X}$ has continuous conditional densities is only a regularity condition in BCQ for proof of asymptotic properties of our estimators proposed below. It is not necessary for the identification lemma. Instead, *Lemma 1* is 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 can be used to define 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 Θ_I' will not be a singleton. The size of identification regions become smaller as stronger restrictions are imposed on $F_{\epsilon|X}$.

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

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

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 sizes of these sets is determined by F_X and linear coefficients b considered. The difference between Θ_I and Θ_I' does not exist if $Pr(p^*(X) = \frac{1}{2}) = 0$.

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

Convexity of identification regions is a desirable property that brings computational advantages in the 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 $Q(.)$) is equal to 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 $\hat{Q}_n(.)$ converge to Θ_I in probability (denoted $\xrightarrow{p}$) under certain metrics between sets, 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.*

In practice, 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 still characterized as minimizers of the smoothed version of $Q(b)$. The additional regularity condition necessary for identification under a smooth Λ is $\Pr\{p^*(X) = \frac{1}{2}\} = 0$.

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* $\Pr\{-Xb = U(X) \text{ or } -Xb = L(X)\} = 0\ \forall b \in \Theta$*, 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 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 the estimation, and is omitted.[3]

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

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

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 $Q(b)$:

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

The two-step extreme estimator is defined as:

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

When L and U are not perfectly observed (as in the case with interval data on regressors) but directly identified in data (as in the case with simultaneous games with incomplete information, we replace them with first-stage estimates $\hat{L}$ and $\hat{U}$. In general, the true parameter β may not be point identified (i.e. $Q(b) = 0$ on a non-singleton set). Therefore we need to choose a metric for differences between sets prior to the definition and proof of consistency of our extreme estimator. We choose the Hausdorff metric as in Manski and Tamer (2002). This metric between two sets A and B in $\mathbb{R}^K$ is defined as

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

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 this 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 $f_0(x_1, ., x_K)$, and both $f_0(x)$ and $f_0(x)p^*(x)$ are 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* $\Pr\{p^*(X) = \frac{1}{2}\} = 0$.

[3] 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.

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) (i) $\sup_{x \in S_X} |\hat{L}(x) - L(x)| = o_p(1)$ and $\sup_{x \in S_X} |\hat{U}(x) - U(x)| = o_p(1)$; and (ii) $\frac{1}{n}\sum_{i=1}^n (-\hat{L}(x_i) - x_i b)_-^2 = O_P(1)$ and $\frac{1}{n}\sum_{i=1}^n (-\hat{U}(x_i) - x_i b)_+^2 = O_P(1)$ for all $b \in \Theta$.

Conditions in *RD(iii)* are regularity conditions for identification. 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 technical conditions that 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, note on most occasions, such as in our examples of discrete games with incomplete information and binary regressions with imperfect data on regressors, the bounding functions $L(x)$ and $U(x)$ are either directly observed, or consistently estimated along with $p^*(x)$ from the sample data in the first step. Conditions in $BF(i)$ require such estimates to converge to the truth uniformly over the support of X at an appropriate rate, and $BF(ii)$ requires sample averages of the difference between the indices and bounds to be stochastically bounded. These (rather weak) conditions ensure the sample analog $\hat{Q}_n$ converges in probability to Q pointwise.

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) Suppose $\sup_{b \in \Theta_I} |\hat{Q}_n(b)| = O_p(a_n^{-1})$ for some sequence of normalizing constants $a_n \to \infty$ and let $\tilde{\Theta}_n = \{b \in \Theta : \hat{Q}_n(b) \leq \hat{c}/a_n\}$, where $\hat{c} \geq a_n\hat{Q}_n(b)$ with probability approaching 1 and $\hat{c}/a_n \xrightarrow{p} 0$. 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 on 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 part *(i)* and 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. The introduction of the sequence $\hat{c}/a_n$ in the definition of $\tilde{\Theta}_n$ in part *(ii)* is needed to account for general cases where Θ_I is not a singleton. The perturbed estimator $\tilde{\Theta}_n$ is consistent for non-singleton Θ_I in Hausdorff metric. We do not derive results on the rate of convergence, or asymptotic distribution of the estimator. A direction of future research will be to find regularity conditions on the joint distribution of (X, ϵ), and functions Λ, L and U, so that $\hat{Q}_n$ satisfy conditions for existence of polynomial minorant in Chernozhukov, Hong and Tamer (2007) and the rate of convergence can be derived.[4]

6 Point identification of β_0

Despite generality in the characterization of Θ_I, exact identification of β_0 is possible under fairly weak conditions on the parameter space, the support of regressors, and the bounding functions. In this section we first specify conditions for point-identification under the general framework of binary response with bounded median dependence. Then we discuss in greater detail how these conditions are satisfied by more primitive restrictions on structural parameters 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}.

The condition *EX* requires researchers to know which coordinates are included in the index and restrict the parameter space accordingly. Along with support conditions in *SX*, these

[4] The condition on existence of polynomical minorant requires there exist positive constants (δ, κ, γ) such that for any $\varepsilon \in (0, 1)$ there are $(\kappa_\varepsilon, n_\varepsilon)$ such that for all $n \geq n_\varepsilon$, $Q_n(\theta) \geq \kappa[d(\theta, \Theta_I) \wedge \delta]^\gamma$ uniformly on $\{\theta \in \Theta : d(\theta, \Theta_I) \geq (\kappa_\varepsilon/a_n)^{1/\gamma}\}$ with probability at least $1 - \varepsilon$.

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

$$\Pr\left(\begin{array}{c} \text{"} -X\beta_0 \leq L(X) \leq U(X) < -Xb\text{"} \\ or \text{ "} -Xb \leq L(X) \leq U(X) < -X\beta_0\text{"} \end{array}\right) > 0$$

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 can be satisfied when: *(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, 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. Proof of Proposition 2 can be interpreted as an extension of Manski's identifying arguments in our current framework with bounded median dependence.

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

6.1 Simultaneous games with incomplete information revisited

Consider the 2-by-2 discrete game with incomplete information in Section 3. We need to extend the definition of identification of parameters 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 may be a correspondence (rather than a function). Let Θ denote the parameter space for $\theta \equiv (\beta_i, \delta_i)_{i=1,2}$ and let $\mathcal{F}_{CMI}$ denote the set of distributions of (ϵ_1, ϵ_2) conditional on X (denoted $F_{\epsilon_1,\epsilon_2|X}$) 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 $\theta^0 \equiv (\theta^0_1, \theta^0_2) \in \Theta$ 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 Brouwer'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 equilibrium outcome 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. Obviously, the conditional independence restriction "$\epsilon_1 \perp \epsilon_2$ given X" has the testable implication

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

Our focus in this paper is on identification rather than testability. We are interested in finding out what can be learned about θ, given that the model is already known (or assumed) to be correctly specified (or equivalently, p^* observed are known to be rationalizable by certain 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 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 for $i = 1, 2$. 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 both 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^0_i 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 in CMI, outcomes of Bayesian Nash Equilibria are characterized by profiles 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. Such exclusion restrictions arise naturally in lots of empirical applications. For example, consider static entry/exit games between firms located in different geographical regions. The indices $X\beta_i$ are interpreted as conditional medians of monopoly profits. There can be commonly observed geographical features (such as local demographics of the workforce with in a region, etc) that only affect the profitability of the local firms but not that of others. This restriction, along with median independence in CMI, enable 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 model structures, and therefore cannot be simply assumed. Below we specify several primitive conditions on support of X and the true 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) For i=1,2, (i) for all $x \in S(X)$, $F^0_{\varepsilon_i|x}$ are Lipschitz continuous on the support of ϵ_i with an unknown constant C_{F_i}; (ii) there exists an unknown constant $K^i_{F_j} > 0$ such that

$$\sup_{t\in\mathbb{R}^1} |F^0_{\varepsilon_j|\bar{x}_{-h_i},\tilde{x}_{h_i}}(t) - F^0_{\varepsilon_j|\bar{x}_{-h_i},x_{h_i}}(t)| \leq K^i_{F_j}|\tilde{x}_{h_i} - x_{h_i}|$$

for all $\bar{x}_{-h_i} \in S(X_{-h_i})$ and $x_{h_i}, \tilde{x}_{h_i} \in S(X_{h_i}|\bar{x}_{-h_i})$; (iii) For all $\eta \in (0,1)$, there exists a finite constant C_η (independent of x) such that $F_{\epsilon_i|X}(C_\eta|x) < \eta$ for all $x \in S(X)$.

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. *DDF-(i)* is satisfied if the true probability densities are bounded above uniformly for all x, and *DDF-(ii)* can be satisfied when ϵ_j is independent of X_{h_i} conditional on X_{-h_i}. These two restrictions enable 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 $b = (b_1, b_2)$ in the parameter space, there exist solutions $\{p_i(.;b, F^0_{\epsilon_1,\epsilon_2|X})\}_{i=1,2}$ 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 $\bar{x}_{-h_i} \in S(X_{-h_i})$ and $x_{h_i} \in S(X_{h_i}|\bar{x}_{-h_i})$.*

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 $b = (b_1, b_2)$ in the parameter space, there exist solutions $\{p_i(.; b, F^0_{\epsilon_1,\epsilon_2|X})\}_{i=1,2}$ 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 $\bar{x}_{-h_i} \in S(X_{-h_i})$.*[5]

Finally, we need restrictions on the support of X_{-h_i} for $i = 1, 2$ so 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 $\alpha S \equiv \{\tilde{x}_{-h_i} : \tilde{x}_{-h_i} = \alpha x_{-h_i}\ for\ some\ x_{-h_i} \in S\}$.

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)* ensure part *(b)* in *SX* is satisfied.

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

---

[5] In fact, results in Lemma 4 and 5 are stronger than necessary for proof of identification of (β_1^0, β_2^0), which only requires results hold for the truth (β_1^0, β_2^0).

Hence exact identification of (β_1^0, β_2^0) is possible under fairly general restrictions on distribution of private signals and the support of regressors. The regressors that are included in both indices can be discrete. Furthermore, the unboundedness of support of X_{h_i} given $\bar{x}_{-h_i}$ can also be relaxed if (1) β_{-i,h_i}^0 is known to be bounded away from zero, and (2) *DDF-(iii)* is strengthened so that ϵ_i have bounded support given all x.

6.2 Binary regressions with imperfect data on regressors revisited

(Binary response with censored regressors) Let $\mathcal{F}_{CSI}$ denote the set of conditional distributions of structural errors $F_{\epsilon_1,\epsilon_2|(X_1,X_2)}$ that are symmetric around 0 and satisfy mutual independence between ϵ_1 and ϵ_2 conditional on (X_1, X_2). Let $S_{(X_1,X_2)}$ denote the support of (X_1, X_2) and $F_{Y_1,\tilde{Y}_2|X_1,X_2}$ denote observable distribution of $Y_1, \tilde{Y}_2$ conditional on $(X_1, X_2) \in S_{(X_1,X_2)}$. Let Θ denote the parameter space for structural coefficients, with its generic element denoted by $\theta \equiv (\beta_1, \beta_2, \gamma_1)$. The set Θ could reflect any exogenous restrictions on the coefficients, such as knowledge of the sign of some coefficients or 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_1) or any real- or vector-valued function of θ.

Definition 4 *θ is observationally equivalent to θ' under $\mathcal{F}_{CSI}$ if $\exists F_{\epsilon_1,\epsilon_2|(X_1,X_2)}, F'_{\epsilon_1,\epsilon_2|(X_1,X_2)} \in \mathcal{F}_{CSI}$ such that*

$$F_{Y_1,\tilde{Y}_2|(X_1,X_2)}(.,.;\theta, F_{\epsilon_1,\epsilon_2|(X_1,X_2)}) = F_{Y_1,\tilde{Y}_2|(X_1,X_2)}(.,.;\theta', F'_{\epsilon_1,\epsilon_2|(X_1,X_2)})$$

almost everywhere on $S_{(X_1,X_2)}$. A feature of the truth $\theta^0 \in \Theta$, 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 $\mathcal{F}_{CSI}$.

Obviously Γ can be the identity function, in which case the identification of $\Gamma(\theta^0)$ refers to point-identification of θ^0. Any $\theta \equiv (\beta_1, \beta_2, \gamma_1)$ must be observationally equivalent to $\theta^c \equiv (c\beta_1, \beta_2, c\gamma_1)$ under $\mathcal{F}_{CSI}$ for some constant $c > 0$.[6] We normalize $|\gamma_1| = 1$. Let θ^0 denote the truth. We maintain the following identifying restrictions.

(CDR) (i) The supports of X_1 and X_2 are both not contained in any linear subspace in the corresponding Euclidean spaces; (ii) there exists a subvector in X_2 (denoted X_{2J}) such

[6] To see this, consider two sets of parameters $(\theta, F_{\epsilon_1,\epsilon_2|(X_1,X_2)})$ and $(\theta^c, F^c_{\epsilon_1,\epsilon_2|(X_1,X_2)})$, where $F^c_{\epsilon_1,\epsilon_2|(X_1,X_2)} = F_{\epsilon_1,\epsilon_2|(X_1,X_2)}(\frac{\cdot}{c}, .)$.

that $\beta_{1J} = 0$ *for all* $\theta \in \Theta$ *and* $\beta^0_{2J} \neq 0$*; (iii) For all* $x_1 \in S_{(X_1)}$, $\Pr(X_2\beta^0_2 \in (c_0, c_1)|x_1) > 0$ *where* (c_0, c_1) *is any non-degenerate open interval in* $\mathbb{R}^1$.

Corollary 4 *Suppose* $\gamma_1 > 0$ *and is normalized to* 1*. Under CDR,* (β^0_1, β^0_2) *are identified under* $\mathcal{F}_{CSI}$.

The scale normalization of γ_1 is necessary for identification of β_1 as argued in Section 3. Note that $\gamma_1 = 1$ is more than a scale normalization, for it requires the sign of γ_1 to be known to researchers. Such a priori knowledge is often possible in empirical applications. This allows us to hold $Median(Y_2|X_1, X_2)$ as known while trying to identify β_1 in the semi-reduced form (6) that defines Y_1. There are no restrictions on the parameter space Θ.

(Binary regressions with interval data 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 suggests 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 finite sample performances of two-step extreme estimators through Monte Carlo simulations. We consider two designs where linear coefficients are only partially identified. The first design is a binary response model with interval data on one regressor. The second is a 2-by-2 discrete game with incomplete information. We find positive evidence that our estimator works well in finite samples under the Hausdorff metric.

7.1 Binary response with interval data on a regressor

In this design, $Y = 1\{\beta_0 X_0 + \beta_1 X_1 + V + \epsilon \geq 0\}$ with $\beta_0 = 1$, $\beta_1 = -3/2$. Support of X_0 is $\{-1, 1, 2, 3\}$ and that of X_1 is $\{\frac{1}{2}, 1, \frac{3}{2}, 2\}$. The support of V consists of 18 grid points scattered evenly on $[0, \frac{17}{3}]$. Each of the three regressors $\{X_0, X_1, V\}$ are independent of the other two, and the probability masses are evenly distributed over the finite, discrete support. Researchers can not observe V perfectly, but do observe $V_0 \equiv int(V)$ and $V_1 \equiv int(V) + 1$ (where $int(a)$ is the largest integer smaller than or equal to a). Conditional on observables (x_0, x_1, v_0, v_1), the disturbance ϵ is normally distributed with mean 0 and standard deviation $x_1 v_1$. Unlike the design considered in Manski and Tamer (2002), we allow ϵ to depend on the interval (v_0, v_1) conditional on v in this design so as to highlight the fact that our estimator can be consistent under weaker restrictions than conditional independence. The median independence restriction $Median(\epsilon|x, v_0, v_1)$ is satisfied. Note all regressors have discrete supports and the rich support condition in Section 6 is not satisfied. Given the simple data generating process, the identification regions of β_0 and β_1 are characterized by systems of linear inequalities whose coefficients are known and can be calculated from conditional choice probabilities $\Pr(Y = 1|x_0, x_1, v_0, v_1)$. We solve analytically for these identification regions, and plot them in the figures below. The identified sets turn out to be not very big, suggesting a lot can be learned about the true parameters even though supports of regressors are discrete and far from satisfying sufficient conditions for point identification in Section 6.

Below we report the estimator's performance under different sample sizes $N = 1000$ and $N = 3000$. We simulate 200 samples by drawing N observations from population randomly with replacement. We calculate two-step estimates for each of the samples. In the first step, we use standard Naradaya-Watson estimates for choice probabilities conditional on discrete regressors. We choose $\Lambda(c) \equiv [\max(c, 0)]^d$ as the smoothing function in $\hat{Q}_n$, where d is a positive integer controlling the smoothness of Λ. We let $\hat{c} = \log n$ and experiment with different choices of sequences of normalizing constants (with $a_n = n^{1/2}$ and n respectively) to check the sensitivity of the performance of our estimator.[7] The objective function is convex in coefficients, thus enabling us to calculate the set estimates in the second step through a simple two-dimensional grid search. The grid points are evenly scattered in the parameter space in $\mathbb{R}^2$ with grid width $1/10$ along both dimensions. As a measure of distance between the identification region Θ_I and the set estimates $\tilde{\Theta}_I$, we record (for each of the 200 set estimates) the proportion of Θ_I covered by $\tilde{\Theta}_I$ (denoted $P1$), and the proportion of $\tilde{\Theta}_I$ covered by Θ_I (denoted $P2$). This is simply calculated in each sample by dividing the number of grid points in $\Theta_I \cap \tilde{\Theta}_I$ by the number of grid points in Θ_I and the number of grid points in $\tilde{\Theta}_I$ respectively.

[7] That $\sup_{b \in \Theta_I} |\hat{Q}_n(b)| = O_p(n^{-1/2})$ follows from applying a Taylor approximation of $\hat{Q}_n$ around the true conditional choice probabilities $p^*(x_i)$. Then the characterization of Θ_I, the uniform boundedness of the higher order derivatives of Λ, $(Xb)^2_-$ and $(Xb)^2_+$ over the support of X and a compact parameter space Θ implies the result.

Table 1 reports performance of the estimator in 200 simulated samples with $\hat{c}/a_n = n^{-\frac{1}{2}}\log(n)$. The parameter spaces considered in the minimization are square areas centered at true coefficients with lengths equal to 20 on each side. In our simulated samples, all set estimates are contained in the interior of the parameter space. To measure discrepancies between the estimates and Θ_I, we report percentiles of (P_1, P_2), percentiles of maximum distances from (β_0, β_1) in $\tilde{\Theta}_I$, and the area of estimates out of the 200 simulations.[8]

Table 1: $a_n = n^{1/2}$ (with grid length=1/10)

(a) $N = 1000$, $d = 2$

prctile	$P1$	$P2$	$\sup_{b\in\tilde{\Theta}_I} \|\|b-\beta\|\|$	*Area (10^3 pts)*
5%	1.000	0.0085	5.037	3.962
10%	1.000	0.0092	5.281	4.122
25%	1.000	0.0100	5.684	4.555
50%	1.000	0.0113	6.196	5.136
75%	1.000	0.0127	6.928	5.799
90%	1.000	0.0141	7.448	6.317
95%	1.000	0.0146	7.878	6.856

(b) $N = 3000$, $d = 2$

prctile	$P1$	$P2$	$\sup_{b\in\tilde{\Theta}_I} \|\|b-\beta\|\|$	*Area (10^3 pts)*
5%	1.000	0.0076	5.876	4.912
10%	1.000	0.0081	6.111	5.149
25%	1.000	0.0087	6.365	5.513
50%	1.000	0.0097	6.826	5.983
75%	1.000	0.0105	7.339	6.673
90%	1.000	0.0113	7.651	7.148
95%	1.000	0.0118	7.914	7.585

(c) $N = 1000$, $d = 3$

[8] An alternative measure of the estimation error of $\tilde{\Theta}_I$ should be its distance from Θ_I in the Hausdorff metric. It is computationally intensive to implement through grid searches in each of the simulated samples. We argue that (1) $\sup_{b\in\tilde{\Theta}_I} ||b-\beta||$ is an interesting measure of discrepancies between estimates and truth in its own right, and (2) given identification regions are small in our designs, $\sup_{b\in\tilde{\Theta}_I} ||b-\beta||$ can be interpreted as a good proxy for the Hausdorff distance.

prctile	$P1$	$P2$	$\sup_{b\in\tilde{\Theta}_I} \|\|b-\beta\|\|$	*Area (10^3 pts)*
5%	1.000	0.0108	4.259	2.832
10%	1.000	0.0117	4.511	3.089
25%	1.000	0.0128	4.940	3.410
50%	1.000	0.0145	5.566	3.997
75%	1.000	0.0170	6.090	4.524
90%	1.000	0.0188	6.605	4.974
95%	1.000	0.0205	7.085	5.358

(d) $N = 3000$, $d = 3$

prctile	$P1$	$P2$	$\sup_{b\in\tilde{\Theta}_I} \|\|b-\beta\|\|$	*Area (10^3 pts)*
5%	1.000	0.0089	5.661	4.226
10%	1.000	0.0092	5.763	4.447
25%	1.000	0.0100	6.124	4.828
50%	1.000	0.0111	6.530	5.235
75%	1.000	0.0120	6.972	5.818
90%	1.000	0.0130	7.341	6.273
95%	1.000	0.0137	7.631	6.538

Table 1 shows that $\tilde{\Theta}_I$ covers Θ_I in all simulated samples regardless of the choice of parameters and sample sizes. On the other hand, the area of $\tilde{\Theta}_I$ is large relative to Θ_I in all simulated samples (including those with bigger sizes). These suggest defining $\tilde{\Theta}_n$ as $\{b \in \Theta : \hat{Q}_n(b) \leq n^{-\frac{1}{2}} \log n\}$ may not be optimal in the sense that the sequence of cutoffs does not vanish faster enough. The maximum distance between truth and set estimates varies mostly between 5 and 7. The choice of the smoothing parameter d has little impact on the coverage probabilities and the maximum estimation errors. Holding other parameters fixed, the size of set estimates tends to be greater and P_2 tends to be smaller when $N = 3000$. At first sight, this pattern seems puzzling, with the cutoff $\hat{c}/a_n$ in the definition of $\hat{\Theta}_I$ decreasing in n. However, recall sample sizes also affect precision of first-stage kernel estimates $\hat{p}(x_i)$. In smaller samples, it is more likely for the sign of $\hat{p}(x_i) - 1/2$ to be different from the truth $p(x_i) - 1/2$ due to estimation errors. As a result, a coefficient b may not be penalized with a positive number in the estimand when $\hat{p}$ is estimated with greater precision in larger samples, but may be penalized in smaller samples. Thus for a fixed level of cutoffs, $\tilde{\Theta}_I$ may exclude more candidates in Θ_I in smaller samples. The disadvantage of a smaller sample should show up in lower values for $\Pr(\Theta_I \subseteq \hat{\Theta}_I)$. In Table 1, this effect is subsumed as $P1$ is literally degenerate at 1, due to our choice of $\hat{c}/a_n$. However, we note evidence of such a disadvantage of smaller samples exists in the first columns in panels (a) and (c) in Table 2 below, as the sequence of cutoff values vanishes at a faster rate.

To improve the estimator's performance in terms of coverage probability $P2$, we consider choosing a sequence $\hat{c}/a_n$ that vanishes faster than $n^{-\frac{1}{2}}\log n$. Choosing such a sequence clearly involves a trade-off between $\Pr(\Theta_I \subseteq \hat{\Theta}_I)$ and $\Pr(\hat{\Theta}_I \subseteq \Theta_I)$. However, Table 1 shows that $\Pr(\Theta_I \subseteq \hat{\Theta}_I)$ is literally degenerate at 1 while $\Pr(\hat{\Theta}_I \subseteq \Theta_I)$ remains low in all simulations, suggesting enough room for trading the estimator's performance in P_1 for P_2. Table 2 reports the performance of the estimator when $\hat{c}/a_n = n^{-1}\log(n)$. As in the previous case, our estimates $\tilde{\Theta}_I$ are contained in the interior of the parameter space in all simulated samples.

Table 2: $a_n = n$ (with grid length=1/10)

(a) $N = 1000$, $d = 2$

prctile	$P1$	$P2$	$\sup_{b\in\tilde{\Theta}_I} \lvert\lvert b-\beta\rvert\rvert$	*Area (10^3 pts)*
5%	0.035	0.0198	1.208	0.112
10%	0.138	0.0562	1.360	0.154
25%	0.647	0.0883	1.565	0.258
50%	1.000	0.1179	1.803	0.374
75%	1.000	0.1514	2.081	0.522
90%	1.000	0.1850	2.406	0.656
95%	1.000	0.1974	2.704	0.704

(b) $N = 3000$, $d = 2$

prctile	$P1$	$P2$	$\sup_{b\in\tilde{\Theta}_I} \lvert\lvert b-\beta\rvert\rvert$	*Area (10^3 pts)*
5%	1.000	0.070	1.572	0.353
10%	1.000	0.074	1.652	0.430
25%	1.000	0.082	1.856	0.519
50%	1.000	0.095	2.140	0.603
75%	1.000	0.111	2.401	0.699
90%	1.000	0.132	2.625	0.773
95%	1.000	0.151	2.711	0.820

(c) $N = 1000$, $d = 3$

prctile	$P1$	$P2$	$\sup_{b\in\tilde{\Theta}_I} \|\|b-\beta\|\|$	*Area (10^3 pts)*
5%	0.1810	0.0348	1.709	0.287
10%	0.3707	0.0446	1.803	0.329
25%	0.9483	0.0599	1.989	0.473
50%	1.0000	0.0784	2.265	0.652
75%	1.0000	0.0995	2.642	0.843
90%	1.0000	0.1195	3.091	1.009
95%	1.0000	0.1310	3.394	1.101

(d) $N = 3000$, $d = 3$

prctile	$P1$	$P2$	$\sup_{b\in\tilde{\Theta}_I} \|\|b-\beta\|\|$	*Area (10^3 pts)*
5%	1.0000	0.0418	2.0522	0.679
10%	1.0000	0.0448	2.2791	0.767
25%	1.0000	0.0493	2.5465	0.889
50%	1.0000	0.0571	2.8723	1.017
75%	1.0000	0.0650	3.1623	1.176
90%	1.0000	0.0751	3.3727	1.295
95%	1.0000	0.0854	3.5078	1.388

Table 2 shows choosing a faster sequence of cutoffs substantially reduces maximum estimation errors in $\tilde{\Theta}_I$ as well as the size of $\tilde{\Theta}_I$ for both sample sizes 1000 and 3000. There is also a remarkable improvement in performance measured by $P2$ relative to Table 1. When the sample size is small at 1000, the new estimator can have a lower coverage probability $P1$ in the worst cases, as is shown by the 5th, 10th and 25th percentiles reported in panels (a) and (c) in Table 2. However, this is an acceptable costs for vastly improving performance in terms of $P2$, as is shown by comparing with all panels in Table 1. Also note the first quartile of $P1$ is still high at 65% when $d = 2$ and 95% when $d = 3$. Furthermore, panels (b) and (d) in Table 2 also show that $P1$ is again degenerate at 1 when $N = 3000$, which suggest increasing the sample size to 3000 provides a quick remedy for performance measured by $\Pr(\Theta_I \subseteq \hat{\Theta}_I)$ as $\hat{c}/a_n$ vanishes faster at $n^{-1}\log(n)$. Choosing $d = 2$ seems to be associated with slightly better performance both in terms of sizes of estimates and the maximum estimation errors. We include four figures to visualize the performance of our estimator. For the case with $N = 3000$, $a_n = n$ and $d = 2$, Figure 1 and Figure 2 depict the best estimates as measured by $P1$ and $P2$ respectively, and Figure 3 and Figure 4 plot the corresponding worst estimates.

7.2 Discrete Games with incomplete information

In this design, we consider a simple 2-by-2 discrete game with incomplete information. There are three observable state variables $X \equiv [X_1, X_2, X_3]$, each with a finite support $S_1 \equiv [1 : 1/3 : 3]$, $S_2 \equiv [-1 : 1/2 : 2]$, and $S_3 \equiv [-2 : 1/2 : 1]$ respectively. The states are independently distributed with the same probability masses for all 343 points in support. The linear coefficients for players' payoffs are $\beta_1 = [1/2, -1, 0]$, $\beta_2 = [3/5, 0, 5/4]$ respectively. The linear coefficients satisfy the exclusion restrictions necessary for point identification discussed in Section 6. Conditional on observable states, private signals ϵ_1, ϵ_2 are normally distributed with mean 0 and standard deviation cX_1 (where c is a dispersion parameter to be controlled for in simulation experiments). As discussed, δ_1 and δ_2 are normalized to be -1 for identification of other parameters. This specification satisfies the conditional independence and median independence of private signals conditional on observed states, as well as technical conditions in *DDF*. In general, β_1, β_2 can not be exactly identified due to the finite support of observable states. By focusing on this design, we are able to gauge the contribution of rich support to the exact identification of coefficients when all other conditions in Section 6 are satisfied.

For each possible value for states, we analytically solve for the Bayesian Nash Equilibria of the game. In general the Bayesian Nash equilibria may not be unique for a given state. We treat the solution of fixed points calculated from Matlab's FMINSEARCH algorithm as outcome from randomly chosen Bayesian Nash equilibrium. Figure 5 and Figure 6 plot the induced choice probabilities in equilibrium conditional all possible states with $c = 1/2$ and 1 respectively. The identification region for (β_1, β_2) are analytically solved for, and are plotted in Figure 7-8 and Figure 9-10 for both $c = 1/2$ and 1. By construction, they are convex regions around true parameters, with boundaries defined by a finite system of linear inequalities whose coefficients depend on equilibrium choice probabilities and support of the states. The magnitude of identification regions of (β_1, β_2) are small relative to the support of regressors and the variance of ϵ_1, ϵ_2. This is positive evidence that a lot can be learned about the true parameter even under finite support of observable states. For different sample sizes $N = 1000$, and $N = 3000$, we simulate 200 samples, each containing N 2-by2 games with states drawn from multinomial distributions above and player i's choices are determined as $d_{i,n} = 1(X_n\beta_i - p^*_{-i}(X_n) - \epsilon_{i,n} \geq 0)$. For each sample, we calculate two-step extreme estimates. The parameter spaces are square areas centered at true parameters with lengths equal to 10 on each side. We also record whether set estimates for identification region hit the boundaries of the parameter space. As before, we use standard Nadaraya-Watson kernel estimates for discrete regressors in the first step. In the second step, $\hat{c}/a_n \equiv \log n/n$, and minimization is done by a simple grid search over compact parameter spaces. The grid length is .01 along all four dimensions of $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{23})$. We use the number of grid points as a proxy measure for the size of the identification region and our set estimates.

The identification regions in designs in (a) and (b) are both very small and are close to being point-identified sizewise. (There are 7 and 56 grid points in (Θ_I^1, Θ_I^2) under (a) and 6 and 6 under (b) respectively.) The coverage probability $\Pr(\Theta_I \subseteq \tilde{\Theta}_I)$ is degenerate at 1 in both (a) and (b). We report in Table 3 the percentiles of maximum estimation errors and the sizes of $\tilde{\Theta}_I$. In panels (a) and (b), we experiment with different scale parameter $c = 1/2$ and 1 which determines variances of ϵ conditional on X, and sample sizes $N = 1000$ and 3000. The set estimates in all 200 simulated samples under both designs are contained in the interior of the parameter space. The sizes of estimates in all 200 simulations are also small relative to the support of latent variables. The maximum estimation errors in both (a) and (b) are comparable and are both small given the support of regressor as well as the variance of the errors. There is evidence that the higher estimation errors in the first stage when $c = 1$ can be offset by increasing the sample size to $N = 3000$.

Table 3 (a) : $c = 1/2$, $N = 1000$ *(with grid length=1/100)*

prctile	$\sup_{b\in\tilde{\Theta}_I^1} \|\|b-\beta_1\|\|$	*Area* of $\tilde{\Theta}_I^1$ *(10^3 pts)*	$\sup_{b\in\tilde{\Theta}_I^2} \|\|b-\beta_2\|\|$	*Area* of $\tilde{\Theta}_I^2$ *(10^3 pts)*
5%	0.982	3.659	1.250	4.934
10%	0.997	4.033	1.261	5.403
25%	1.017	4.615	1.280	6.932
50%	1.035	5.819	1.305	8.801
75%	1.055	7.287	1.332	11.324
90%	1.089	9.296	1.450	13.253
95%	1.191	10.161	1.715	15.015

Table 3 (b) : $c = 1,\ N = 3000$ *(with grid length=1/100)*

prctile	$\sup_{b\in\tilde{\Theta}_I^1} \|\|b-\beta\|\|$	*Area* of $\tilde{\Theta}_I^1$ *(10^3 pts)*	$\sup_{b\in\tilde{\Theta}_I^2} \|\|b-\beta_2\|\|$	*Area* of $\tilde{\Theta}_I^2$ *(10^3 pts)*
5%	1.067	4.345	1.328	5.996
10%	1.077	4.754	1.337	6.504
25%	1.095	5.803	1.351	7.916
50%	1.111	6.938	1.370	9.889
75%	1.133	8.632	1.389	12.320
90%	1.150	9.984	1.414	14.281
95%	1.164	10.506	1.554	15.839

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 a disturbance term. 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 have not been studied in the literature so far. We prove point-identification of linear coefficients in these motivating submodels under fairly general restrictions on 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, and the derivation of asymptotic distribution of the estimator under point identification. Another interesting direction is the estimation when 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 $p(x; b, F_{\varepsilon|X}) = p^*(x)$ $\forall x \in S_X \backslash \xi_b$ (and therefore $a.e. F_X$ 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 $b_\alpha \equiv \alpha b_1 + (1-\alpha) b_2$ 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 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*,

$$
\sup_{x\in S_X}|\hat{p}(x)-p^*(x)| = o_p(n^{-1/4}) \tag{7}
$$

Apply a mean value expansion of $\hat{Q}_n(b)$ around $p^*(x_i)$, $L(x_i)$ and $U(x_i)$ respectively:

$$\begin{aligned}\hat{Q}_n(b) &= \frac{1}{n}\sum_{i=1}^n\left[\Lambda(p_i^*-\frac{1}{2})(-x_ib-U_i)_+^2+\Lambda(\frac{1}{2}-p_i^*)(-x_ib-L_i)_-^2\right]+...\\ &\quad +\frac{1}{n}\sum_{i=1}^n\left\{\left[\begin{array}{c}\tilde{\Lambda}'_{i+}(-x_ib-\tilde{U}_i)_+^2-\tilde{\Lambda}'_{i-}(-x_ib-\tilde{L}_i)_-^2\\ -2\tilde{\Lambda}_{i+}(-x_{ib}-\tilde{U}_i)_+\\ -2\tilde{\Lambda}_{i-}(-x_ib-\tilde{L}_i)_-\end{array}\right]'\left[\begin{array}{c}\hat{p}_i-p_i^*\\ \hat{L}_i-L_i\\ \hat{U}_i-U_i\end{array}\right]\right\}\end{aligned}$$

where subscripts i for p, U, L denotes values of these functions at x_i, $\tilde{p}_i$, $\tilde{U}_i$, $\tilde{L}_i$ are shorthands for points on line segments between $(p_i, \hat{p}_i)$, $(U_i, \hat{U}_i)$ and $(L_i, \hat{L}_i)$ respectively, and $\tilde{\Lambda}_{i+} \equiv \Lambda(\tilde{p}_i - \frac{1}{2})$, $\tilde{\Lambda}_{i-} \equiv \Lambda(\frac{1}{2} - \tilde{p}_i)$. Let

$$\bar{Q}_n(b) \equiv \frac{1}{n}\sum_{i=1}^n\left[\Lambda(p_i^*-\frac{1}{2})(-x_ib-U_i)_+^2+\Lambda(\frac{1}{2}-p_i^*)(-x_ib-L_i)_-^2\right]$$

Then for all $b \in \Theta$, $|\hat{Q}_n(b) - \bar{Q}_n(b)|$ is bounded between

$$\left[\begin{array}{c}\frac{1}{n}\sum_{i=1}^n|\tilde{\Lambda}'_{i+}(-x_ib-\tilde{U}_i)_+^2-\tilde{\Lambda}'_{i-}(-x_ib-\tilde{L}_i)_-^2|\\ \frac{2}{n}\sum_{i=1}^n|\tilde{\Lambda}_{i+}(-x_ib-\tilde{U}_i)_+|\\ \frac{2}{n}\sum_{i=1}^n|\tilde{\Lambda}_{i-}(-x_ib-\tilde{L}_i)_-|\end{array}\right]'\left[\begin{array}{c}\sup_{x\in S_X}|\hat{p}-p^*|\\ \sup_{x\in S_X}|\hat{L}-L|\\ \sup_{x\in S_X}|\hat{U}-U|\end{array}\right]$$

Recall $\hat{p} \xrightarrow{p} p^*$ uniformly over S_X as shown in (7) and $\hat{L} \xrightarrow{p} L, \hat{U} \xrightarrow{p} U$ uniformly over S_X by $BF(i)$. Then the boundedness of Λ' over $[-1/2, 1/2]$, the condition $BF(ii)$ and the Law of Large Numbers together imply $\hat{Q}_n(b) \xrightarrow{p} Q(b)\ \forall b \in \Theta$. 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 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 Theorem 3.1 in Chernozhukov, Hong and Tamer (2007), and is omitted for brevity.

Proof of Proposition 2. By *BCQ-2* and *Lemma 1* and *Corollary 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,

$$\begin{bmatrix} p_1(\bar{x}_{-h_1}, x_{h_1}) \\ p_2(\bar{x}_{-h_1}, x_{h_1}) \end{bmatrix} = \begin{bmatrix} F_{\epsilon_1|\bar{x}_{-h_1},x_{h_1}}(\bar{x}_{-h_1}b_{1,-h_1} + x_{h_1}b_{1,h_1} - p_2(\bar{x}_{-h_1}, x_{h_1})) \\ F_{\epsilon_2|\bar{x}_{-h_1},x_{h_1}}(\bar{x}_{-h_1}b_{2,-h_1} + x_{h_1}b_{2,h_1} - p_1(\bar{x}_{-h_1}, x_{h_1})) \end{bmatrix} \tag{8}$$

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 $k \le K_1$. Then $C^{K_1}(\bar{x}_{-h_1})$ is bounded in the sup-norm and equi-continuous 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 $\exists x_{h_1}^a, x_{h_1}^b \in S(X_{h_1}|\bar{x}_{-h_1})$ such that $|f_0(x_{h_1}^a) - f_0(x_{h_1}^b)| > K_1'|x_{h_1}^a - x_{h_1}^b|$ for some $K_1' > K_1$. By convergence of f_n, for all $\varepsilon > 0$, $|f_n(x_{h_1}^j) - f_0(x_{h_1}^j)| \le \frac{\varepsilon}{2}|x_{h_1}^a - x_{h_1}^b|$ for $j=a,b$ for n big enough. Hence $\frac{|f_n(x_{h_1}^a) - f_n(x_{h_1}^b)|}{|x_{h_1}^a - x_{h_1}^b|} > K_1' - \varepsilon$ for n big enough. For any $\varepsilon < K_1' - K_1$, this implies for n big enough, f_n is not Lipschitz continuous with $k \le K_1$. Contradiction. Hence $C^{K_1}(\bar{x}_{-h_1})$ is bounded, equi-continuous, 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 (8) into the first one, and we have

$$\bar{p}_1(x_{h_1}) = \bar{F}_{\epsilon_1|x_{h_1}}\{\bar{x}_{-h_1}b_{1,-h_1} - \bar{F}_{\epsilon_2|x_{h_1}}[\bar{x}_{-h_1}b_{2,-h_1} + x_{h_1}b_{2,h_1} - \bar{p}_1(x_{h_1})]\} \tag{9}$$

where $\bar{p}_1$ and $\bar{F}_{\epsilon_i|x_{h_1}}$ are shorthand notations for conditioning on $\bar{x}_{-h_1}$. Let $\bar{\tau}(x_{h_1})$ denote the right-hand side of (9). Suppose $\bar{p}_1(x_{h_1})$ is Lipschitz continuous with constant $k \le K_1$ for some $K_1 > 0$. Then by the definition of the Lipschitz constants in *DDF (i)-(ii)*, for all x_{h_1}, $\tilde{x}_{h1} \in S(X_{h_1}|\bar{x}_{-h_1})$, $|\bar{\tau}(x_{h_1}) - \bar{\tau}(\tilde{x}_{h_1})| \le D(K_1)|x_{h_1} - \tilde{x}_{h_1}|$, where

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

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 (9) 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. We first prove the case from player 1's perspective. Fix $\bar{x}_{-h_1} \in S(X_{-h_1})$. then for any $b \equiv (b_1, b_2)$ in the parameter space $\Theta \equiv \Theta_1 \otimes \Theta_2$, and all $x_{h_1} \in S(X_{h_1}|x_{-h_1})$,

$$\bar{p}_2(x_{h_1}) = \bar{F}_{\epsilon_2|x_{h_1}}[\bar{x}_{-h_1} b_{2,-h_1} + x_{h_1} b_{2,h_1} - \bar{p}_1(x_{h_1})] \tag{10}$$

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 $b \in \Theta$. Then *REG, DDF* and that $b_{2,h_1} \neq 0$ implies the image of $\bar{p}_2(x_{h_1})$ is the connected interval $(0, 1)$ for all $b \in \Theta$ given $\bar{x}_{-h_1}$. By similar arguments, it follows from *REG* and *DDF* that $\bar{p}_2(x_{h_1})$ defined in (10) is also Lipschitz continuous on $S(X_{h_1}|\bar{x}_{-h_1})$ for all $b \in \Theta$. And symmetric arguments complete the proof for player 2.

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

$$\Pr\{X'_{-h_1}(b_{1,-h_1} - \beta^0_{1,-h_1}) \neq 0, sgn(X'_{-h_1} b_{1,-h_1}) \neq sgn(X'_{-h_1}\beta^0_{1,-h_1})\} > 0$$

then *SX-(b)* is automatically 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 $F^0_{\epsilon_1,\epsilon_2|X}$, and let $\bar{p}^*_1(x_{h_1})$ be a shorthand for conditioning on $\bar{x}_{-h_1}$. 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*. That is, $\Pr\{\bar{p}^*_2(X_{h_1}) \in I_{b_1,\beta^0_1}(X)|\bar{x}_{-h_i}\} > 0$ for all $\bar{x}_{-h_i} \in S(X_{-h_i})$ and $b_1 \neq \beta^0_1$ in Θ_1, where $\bar{p}^*_2$ is the equilibrium outcome under the truth, and $I_{b_1,\beta^0_1}(X)$ is the intersection of $(0, 1)$ with the random interval between Xb_1 and $X\beta^0_1$. It follows immediately that *SX-(c)* is also satisfied. Proof for the case with player 2 follows from similar arguments.

Proof of Corollary 4. Note *CDR-(i)* is the standard full rank condition so that (i) $\Pr(X_1\beta^0_1 \neq X_1 b) > 0$ for all $b \neq \beta^0_1$ in the parameter space and (ii) β^0_2 is exactly identified under the conditional symmetry assumption. The exclusion restriction in *CDR-(ii)* and the rich support condition in *CDR-(iii)* then ensures

$$\Pr\left(\begin{array}{c} \text{"} -X_1\beta_1^0 \leq Med(Y_2|X_1,X_2) < -X_1 b\text{"} \\ \vee \text{"} -X_1 b \leq Med(Y_2|X_1,X_2) < -X_1\beta_1^0\text{"} \end{array}\right) > 0.$$

for all $b \neq \beta_1^0$.

References

[1] Aradillas-López, A. (2005): "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] Chamberlain, G. (1986): "Asymptotic Efficiency in Semiparametric Models with Censoring," Journal of Econometrics, 32, 189-218

[4] Chen, S., and S. Khan (2003):

[5] Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and Inference in Partially Identified Models", Econometrica, Vol. 70, No. 5, 1243-1284

[6] Cosslett, S. (1987): "Efficient Bounds for Distribution -Free Estimators of the Binary Choice and the Censored Regression Models," Econometrica, 55, 559-585

[7] Han, A. (1987): "Nonparametric Analysis of Generalized Regression Model," Journal of Econometrics, 35, 303-316

[8] Horowitz, J. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," Econometrica, 60, 505-531

[9] Ichimura, H. (1993): "Semiparametric Least Squares and Weighted SLS Estimation of Single-Index Models," Journal of Econometrics, 58, 71-120

[10] Klein, R. and R. Spady (1993): "An Efficient Semiparametric Estimator of Binary Response Models," Econometrica, 61, 387-421

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

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

[13] Matzkin, R. (1992): "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models," Econometrica, 60, 239-270

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

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

[16] Rigobon, R. and T. Stocker (2005): "Estimation with Censored Regressors: Basic Issues", International Economic Review, forthcoming

[17] Powell, J. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," Journal of Econometrics, 25, 303-325

[18] Powell, J. (1986): "Symmetrically Trimmed Least Squares Estimation for Tobit Models," Econometrica, 54, 1435-1460

[19] Sherman, R. (1993): "The Limiting Distribution of the Maximum Rank Correlation Estimator," Econometrica, 61, 123-137

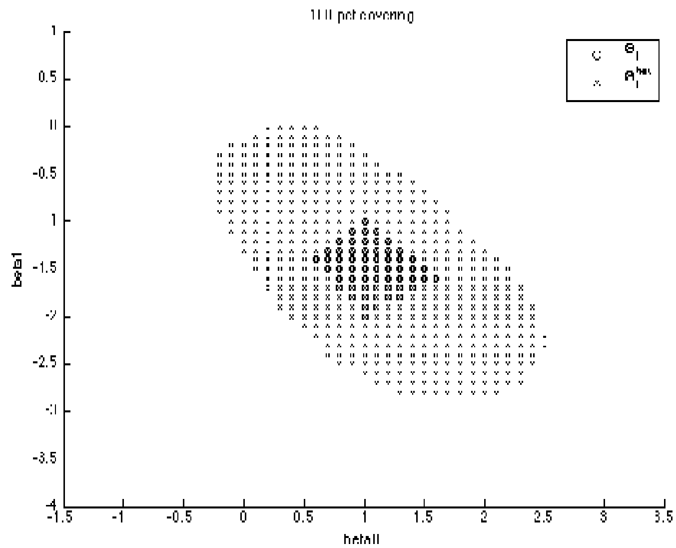

Figure 1: 100% of Θ_I covered by $\tilde{\Theta}_I$ (with $N = 3000$, $a_n = n$, $d = 2$)

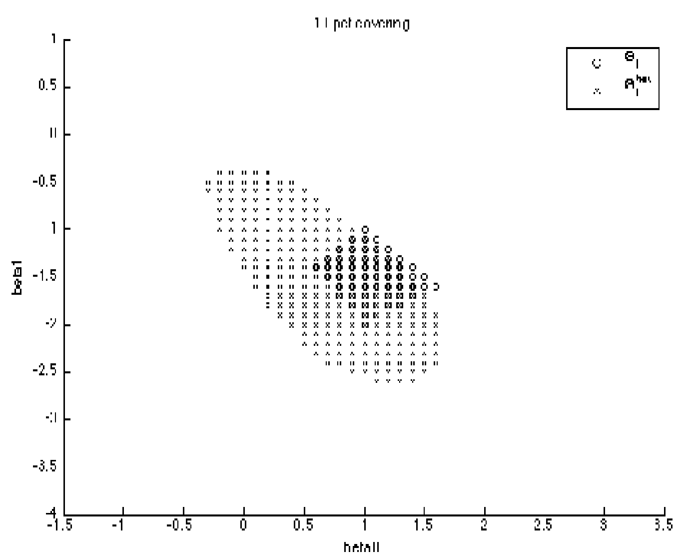

Figure 2: 19% of $\tilde{\Theta}_I$ covered by Θ_I (with $N = 3000$, $a_n = n$, $d = 2$)

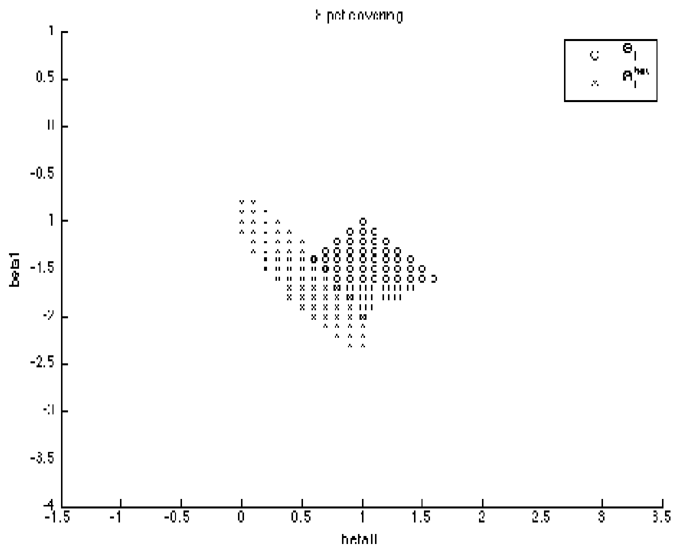

Figure 3: 6% of Θ_I covered by $\tilde{\Theta}_I$ (with $N = 3000$, $a_n = n$, $d = 2$)

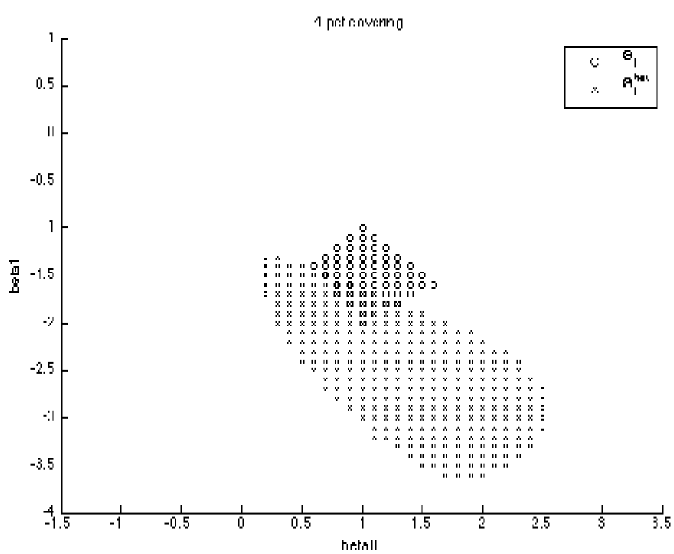

Figure 4: 4% of $\tilde{\Theta}_I$ covered by Θ_I (with $N = 3000$, $a_n = n$, $d = 2$)

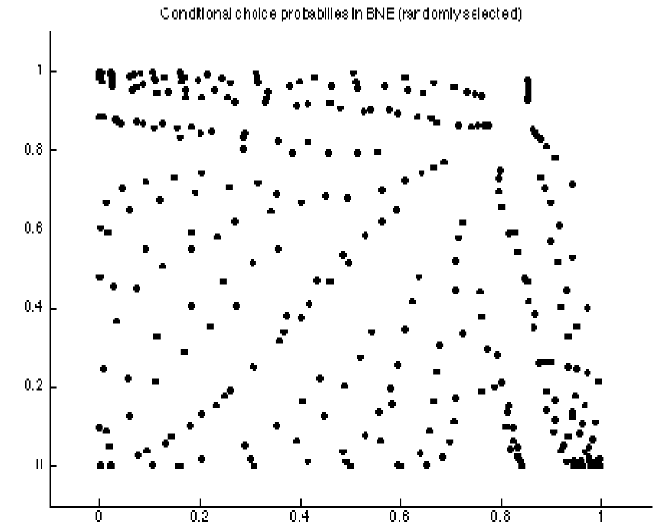

Figure 5: Equilibrium outcome for $c = 1/2$ (randomly selected BNE)

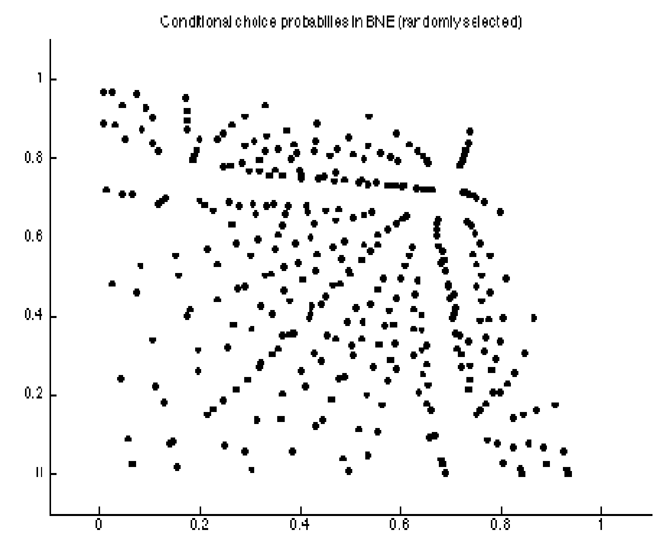

Figure 6: Equilibrium outcome for $c = 1$ (randomly selected BNE)

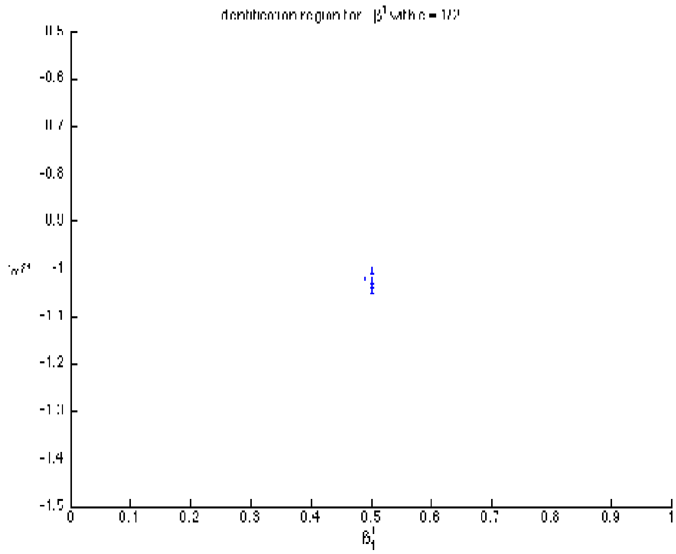

Figure 7: Identification region for β_1 with $c = 1/2$

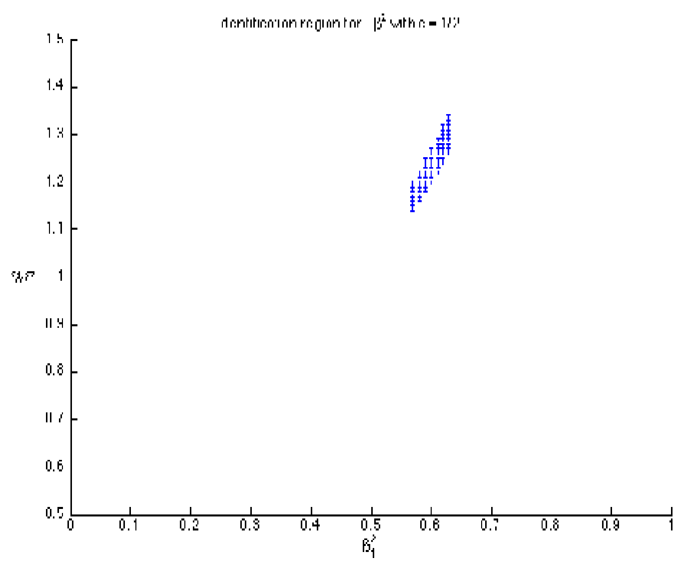

Figure 8: Identification region for β_2 with $c = 1/2$

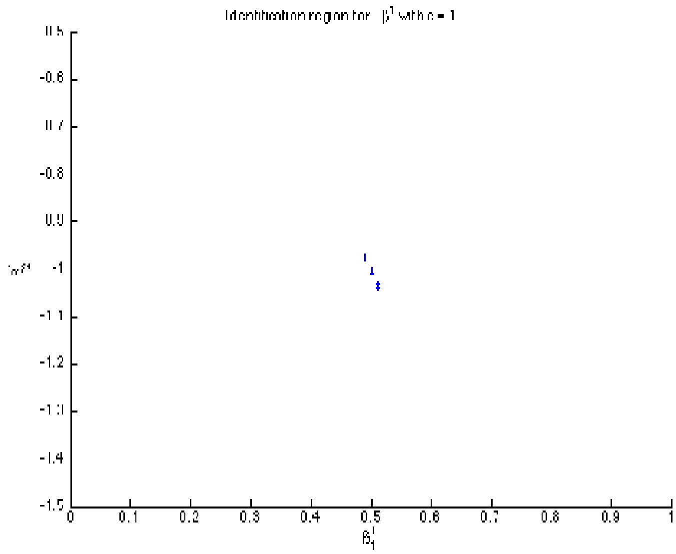

Figure 9: Identification region for β_1 with $c = 1$

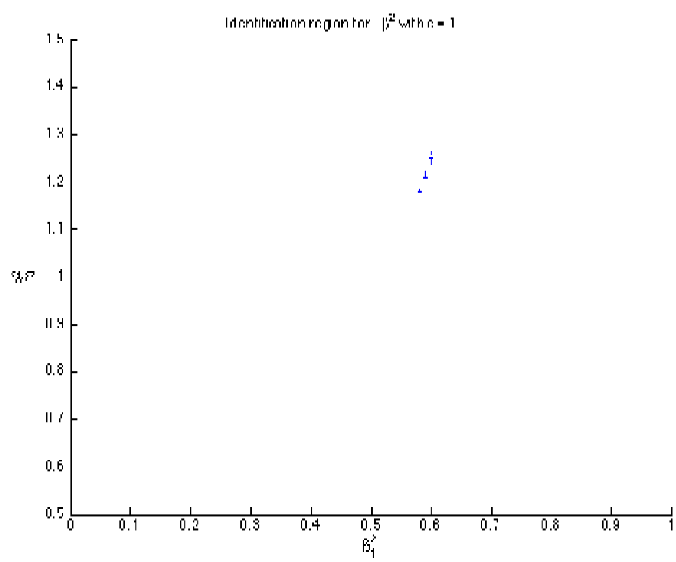

Figure 10: Identification region for β_2 with $c = 1$