The Econometrics Journal

Econometrics Journal (2014), volume **17**, pp. 241–270.
doi: 10.1111/ectj.12026

Estimation of discrete games with correlated types

HAIQING XU[†]

[†]*Department of Economics, University of Texas at Austin, BRB 1.116, Mailcode C3100, Austin, TX 78712, USA.*
E-mail: h.xu@austin.utexas.edu

First version received: January 2012; final version accepted: January 2014

Summary In this paper, I focus on the identification and estimation of static games of incomplete information with correlated types. Instead of making the independence assumption on players' types in order to simplify the equilibrium set, I propose an approach that allows me to identify subsets of the space of covariates (i.e. publicly observed state variables in payoff functions), for which there exists a unique pure strategy Bayesian Nash equilibrium (BNE) and the equilibrium strategies are monotonic functions. Moreover, I characterize the monotonic pure strategy BNE in a simple manner and propose an estimation procedure that uses observations only from the subset of the covariate space where the game admits a unique monotonic pure strategy BNE. Furthermore, I show that the proposed estimator is $\sqrt{n}$-consistent and has a limiting normal distribution.

Keywords: *Incomplete information, Maximum likelihood estimation, Monotonic pure strategy Bayesian Nash equilibrium.*

1. INTRODUCTION

In the analysis of various economic and social situations, discrete games have been used to study the strategic behaviours of firms or the interactions between individuals who are socially connected with each other. A leading example is the oligopoly entry (e.g. Berry and Tamer, 2006). The idea of these models is that an agent's utility often depends not only on her choice and covariate variables, but also on other agents' choices. Therefore, the optimal decision of each agent depends on the choices of the other agents. In this paper, I study the identification and estimation of two-player binary choice games of incomplete information where players' types are correlated. For computational tractability and to avoid the issue of multiple equilibria introduced by the type correlation, I develop an approach that exploits the information contained only in a subset of the covariate space where there exists a unique monotonic pure strategy Bayesian Nash equilibrium (BNE).

In the literature, empirical discrete games have many applications. For example, Bjorn and Vuong (1984) have studied labour force participation with strategic interactions within couples. In the last two decades, this class of games has been widely adopted in empirical industrial organization to study firms' entry behaviour (e.g. Bresnahan and Reiss, 1990, 1991a, b, Berry, 1992, Berry and Tamer, 2006, Jia, 2008, and Ciliberto and Tamer, 2009). Recently, Brock and Durlauf (2001, 2007), Xu (2011) and Kline (2012), among others, have also used discrete games to analyse social interactions. In this paper, I focus on two-player binary games, which have many practical applications (e.g. duopoly entry and household decision-making).

In this paper, I study a parametric binary game of incomplete information, which might have multiple equilibria.[1] The proposed methodology contributes to the literature in two ways.

First, I allow players' types to be correlated, which is motivated by empirical concerns. The (conditional) independence of types has been widely assumed in the recent literature (e.g. Aguirregabiria and Mira, 2002, Pesendorfer and Schmidt-Dengler, 2003, and Bajari et al., 2010). Such an assumption is crucial for their identification strategies, because it implies that each player's equilibrium beliefs about the choices of their rivals depend on observed state variables only and can be non-parametrically estimated thereof. In contrast, I assume that players' types are positively correlated; the correlation coefficient is also a parameter of interest in my setting.

The quest for correlated types in discrete games is motivated by several considerations. The (conditional) independence of types implies that players' actions should also be conditionally independent given the covariates. This is a testable restriction that is quite strong and might be rejected by data.[2] As a side note, the independence assumption also implies that all the equilibria of the game must be monotonic pure strategy BNEs, which is convenient but rules out non-monotonic strategy BNEs. Moreover, from the point of view of model specification, it is also important to allow correlation. For example, consider two firms entering a local market: one would expect the private payoff shocks on the profitability of entry to be positively correlated with each other, especially when these shocks depend on some common factors of the local market and each player observes only her own aggregated shock without knowing its components (i.e. the common factors and idiosyncratic noise).

Secondly, the proposed approach is flexible on equilibrium selection in the presence of multiple equilibria. In the literature, the multiple equilibria issue usually invokes ad hoc equilibrium selection assumptions in the data-generating process. For instance, when there are multiple equilibria, only one equilibrium is being played in data. Such an assumption can be found, e.g. in Sweeting (2009), Bajari et al. (2010), Tang (2010), Aradillas-Lopez (2010), Wan and Xu (2012) and Liu et al. (2013). If the independence assumption is dropped, then this even complicates the multiple equilibria issue. First, it is difficult to characterize all the equilibria, especially those solutions with non-monotonic strategies. Secondly, the number of equilibria is unknown and varies with the payoff-related covariates.[3] Hence, even if an equilibrium selection rule were imposed, it would be difficult to implement in practice.

In this paper, the identification and estimation strategy is reliant on finding a subset in the space of covariates (i.e. publicly observed state variables), where the game admits a unique monotonic pure strategy BNE. The advantage of this approach is to avoid all the difficulties of computing non-monotonic strategy BNEs and multiple equilibria. When covariates come from this subset, the solution derived from the game model is simple and uniquely exists. After obtaining this subset, I establish point identification of model parameters using the subpopulation

[1] Aradillas-Lopez (2010) has estimated the same game structure without making parametric restrictions on types, but under a different equilibrium solution concept.

[2] There are two other possible sources for the correlation between players' actions: unobserved heterogeneity (i.e. some payoff-relevant variables are observed by both players but not by the researcher; see Grieco, 2013) and multiple equilibria (see De Paula and Tang, 2012).

[3] In a two-player game with linear payoffs, suppose the private payoff shocks $U = (U_1, U_2)$ conform to a bivariate normal distribution and U_1 and U_2 are conditionally independent given the covariates X. Then, it is known that the number of equilibria ranges from 1 to 3 (e.g. Grieco, 2013). However, such a result is not obtained if U_1 and U_2 are positively correlated.

and then propose a two-stage maximum likelihood estimation (MLE) strategy where the first stage helps me to find the subset.[4]

The (unique) monotonic pure strategy BNE can be characterized in a simple manner. In the presence of correlation, it is costly to obtain a closed-form solution for the equilibrium in general. In the binary decision game considered in this paper, an important insight is that a monotonic pure strategy is fully characterized by a cut-off value in the real line. Therefore, a numerical solution of the monotonic pure strategy BNE can be obtained as a fixed point in the two-dimensional vector space.

This paper is organized as follows. In Section 2, I describe the game model. In Section 3, I provide a characterization of BNEs and monotonic pure strategy BNEs. I establish the existence of a unique monotonic pure strategy BNE for covariates belonging to a subset. In Sections 4 and 5, I establish the identification and estimation of the model parameters, respectively. In Section 6, I provide Monte Carlo experiment studies to illustrate the performance of the proposed estimator in finite samples. I conclude in Section 7. All proofs are in the Appendix.

2. MODEL

Consider the following two-by-two static game of incomplete information:

		Player 2	
		$Y_2 = 1$	$Y_2 = 0$
Player 1	$Y_1 = 1$	$X_1'\beta_1 - \alpha_1 - U_1,\ X_2'\beta_2 - \alpha_2 - U_2$	$X_1'\beta_1 - U_1,\ 0$
	$Y_1 = 0$	$0,\ X_2'\beta_2 - U_2$	$0,\ 0$

Here, $X = (X_1, X_2) \in \mathscr{S}_X \subseteq \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$ represents public knowledge on the state of the game, which is known to both players. The payoff shock $U_j \in \mathbb{R}$ $(j = 1, 2)$ is player j's private information (i.e. type), which is only observed by j, but not by her rival. Y_j is the choice of player j. Let $U = (U_1, U_2)$ be independent of X and let it conform to a joint normal distribution with unit variances and correlation coefficient $\rho_0 \in [0, 1)$, which is assumed to be common knowledge of both players.[5] Moreover, $\beta_j \in \mathbb{R}^{k_j}$ and $\alpha_j \in \mathbb{R}_+$ are structural parameters in the payoff functions and α_j measures the size of strategic effects.[6] Let $\theta_0 = (\alpha_1, \alpha_2, \beta_1', \beta_2', \rho_0)' \in \Theta$ be the parameter of interest, where Θ is the parameter space. Throughout this paper, I also use $\theta = (a_1, a_2, b_1', b_2', \rho)'$ to denote a generic parameter value in Θ. Furthermore, I assume that the researcher observes $Y = (Y_1, Y_2)$ and all the publicly observed state variables X, but not U.

In this setting, the unit variance of U_j $(j = 1, 2)$ is a normalization from the view of observational equivalence. Suppose the variance takes the general form σ_j^2 where $\sigma_j > 0$, then

[4] I show that the subset can be constructed/identified by applying the level-k rationality ($k = \infty$) constraints (e.g. Aradillas-Lopez and Tamer, 2008).

[5] Aradillas-Lopez (2010) have developed a semi-parametric approach without a parametric specification on the distribution of U.

[6] Here, α_j is restricted to be non-negative only for the brevity of notation.

it can be shown that such a structure with non-unit variances is observationally equivalent to the structure with $\tilde{\alpha}_j = \alpha_j/\sigma_j$, $\tilde{\beta}_j = \beta_j/\sigma_j$, $\tilde{\rho} = \rho_0$ and $\tilde{\sigma}_j^2 = 1$. In other words, the variance normalization on both types is necessary for the identification of payoff coefficients (but not for the correlation coefficient).

A game and the according equilibria with a similar set-up can also be found in, e.g. Aguirregabiria and Mira (2002) and Pesendorfer and Schmidt-Dengler (2003), and references therein. In this incomplete information game, I adopt the standard solution concept: pure strategy BNEs (e.g. Aumann, 1964, and Harsanyi, 1967, 1968a, b). In equilibrium, player j's strategy is a function $s_j^*(X, U_j)$, where $s_j^* : \mathbb{R}^{k_1+k_2} \times \mathbb{R} \to \{0, 1\}$ maps all j's information to a binary decision. Player j chooses s_j^* in such a way that it maximizes her expected payoff: choosing $s_j^* = 1$ if and only if $X_j'\beta_j - \alpha_j E[s_{-j}^*(X, U_{-j})|X, U_j] - U_j \geq 0$, where $E[s_{-j}^*|X, U_j]$ is the beliefs of her rival's move in equilibrium. In other words, fixing $X = x \in \mathscr{S}_X$, the equilibrium strategy profile $s^* = \{s_1^*(x, \cdot), s_2^*(x, \cdot)\}$ is a fixed point solving the following simultaneous equation system: for $j = 1, 2$,

$$s_j(x, u_j) = \mathbf{1}(x_j'\beta_j - \alpha_j E[s_{-j}(x, U_{-j})|U_j = u_j] - u_j \geq 0), \quad \forall u_j \in \mathbb{R}, \tag{2.1}$$

where $\mathbf{1}(\cdot)$ is the indicator function. Note that I drop the conditioning variable $X = x$ in j's belief term $E[s_{-j}(x, U_{-j})|U_j = u_j]$ because of the independence between X and U. In (2.1), player j's equilibrium strategy depends not only on j's observed state variables x_j, but also on her rival's state variables x_{-j}. This is because x_{-j} affects player j's expectation on her rival's choice.

This binary game of incomplete information can be interpreted as an entry model, where two firms simultaneously decide whether to enter a local market or not (e.g. Ciliberto and Tamer, 2009). Before they make their decisions, information X is disclosed publicly, and each player observes a private payoff shock for her own entry profit. There are interactions between the players' choices: α_j measures the magnitude of strategic impacts. Moreover, each player's entry profit is parametrized by a linear sum of the publicly observed term, the private shock and the strategic effects. Note that asymmetry in this game arises as long as $X_1'\beta_1 \neq X_2'\beta_2$, which reflects the fact that one player might have a publicly known advantage or disadvantage for entering this local market.

From the above discussion, for a given realization $x = (x_1, x_2)$ of public states X, a BNE is a fixed point in the functional space. To obtain such a solution, a convenient assumption, which is widely used in the literature, is that U_1 and U_2 are conditionally independent given X. This means that an individual's private payoff shock does not carry any additional information for the rival's choice. When U_1 and U_2 are allowed to be positively correlated, it becomes difficult to characterize any such equilibria, especially non-monotonic strategy BNEs. In general, it is quite costly to know the whole equilibrium set, or even the number of equilibria.

3. MONOTONIC PURE STRATEGY BAYESIAN NASH EQUILIBRIUM

As a special class of pure strategy BNEs, monotonic pure strategy BNEs can be characterized in a much simpler manner. Specifically, each player's strategy of such an equilibrium is a threshold function. Furthermore, I characterize the set of publicly observed state variables where the game admits a unique equilibrium and it is a monotonic strategy BNE. In the following analysis, I first characterize monotonic pure strategy BNEs.

Fix public states $X = x$. A monotonic pure strategy BNE can be characterized by a two-dimensional vector $u^*(x) \equiv (u_1^*(x), u_2^*(x))$ such that for $j = 1, 2$,

$$s_j^*(x, u_j) = \mathbf{1}(u_j \leq u_j^*(x)). \tag{3.1}$$

Here, $u^*(x)$ satisfies the following mutual consistency conditions: for $j = 1, 2$,

$$x_j'\beta_j - \alpha_j \Pr(U_{-j} \leq u_{-j}^*(x)|U_j = u_j^*(x)) - u_j^*(x) = 0. \tag{3.2}$$

Then, a monotonic pure strategy BNE is obtained by solving a fixed point $u^*(x)$ in the vector space $\mathbb{R}^2$.

In the covariate space, I now define a subset $\mathcal{U}(\theta_0)$, which depends on the underlying parameter value θ_0, such that for any $x \in \mathcal{U}(\theta_0)$ the game admits a unique monotonic pure strategy BNE.

For $j = 1, 2$, let $h_j(u; \theta_0) = u_j + \alpha_j \Pr(U_{-j} \leq u_{-j}|U_j = u_j)$ be a function of $u = (u_1, u_2)$. In Definition 3.1, I define a rectangular $\mathcal{I}(x; \theta_0)$ on the support $\mathscr{S}_U$ through a recursive scheme. This corresponds to the 'level-k rationality' of Aradillas-Lopez and Tamer (2008), which is a notion weaker than the BNE solution concept.

DEFINITION 3.1. *For any $x \in \mathscr{S}_X$, let $\mathscr{V}_{j,1}^-(x; \theta_0) = x_j'\beta_j - \alpha_j$, $\mathscr{V}_{j,1}^+(x; \theta_0) = x_j'\beta_j$ and for $k \geq 2$,*

$$\mathscr{V}_{j,k}^-(x; \theta_0) = x_j'\beta_j - \alpha_j \Pr(U_{-j} \leq \mathscr{V}_{-j,k-1}^+(x; \theta_0)|U_j = \mathscr{V}_{j,k-1}^-(x; \theta_0)),$$

$$\mathscr{V}_{j,k}^+(x; \theta_0) = x_j'\beta_j - \alpha_j \Pr(U_{-j} \leq \mathscr{V}_{-j,k-1}^-(x; \theta_0)|U_j = \mathscr{V}_{j,k-1}^+(x; \theta_0)).$$

Furthermore, let $\mathscr{V}_j^-(x; \theta_0) = \lim_{k\to\infty} \mathscr{V}_{j,k}^-(x; \theta_0)$ and $\mathscr{V}_j^+(x; \theta_0) = \lim_{k\to\infty} \mathscr{V}_{j,k}^+(x; \theta_0)$. Moreover, let $\mathcal{I}_{j,k}(x; \theta_0) = (\mathscr{V}_{j,k}^-(x; \theta_0), \mathscr{V}_{j,k}^+(x; \theta_0))$, $\mathcal{I}_j(x; \theta_0) = (\mathscr{V}_j^-(x; \theta_0), \mathscr{V}_j^+(x; \theta_0))$ and $\mathcal{I}(x; \theta_0) = \mathcal{I}_1(x; \theta_0) \times \mathcal{I}_2(x; \theta_0)$.

Throughout the following analysis, I use $\mathscr{V}_{j,k}^-(x)$, $\mathscr{V}_{j,k}^+(x)$, $\mathscr{V}_j^-(x)$ and $\mathscr{V}_j^+(x)$ in lieu of $\mathscr{V}_{j,k}^-(x; \theta_0)$, $\mathscr{V}_{j,k}^+(x; \theta_0)$, $\mathscr{V}_j^-(x; \theta_0)$ and $\mathscr{V}_j^+(x; \theta_0)$, respectively, in order to simplify my notation and to emphasize their dependence on x. Note that $\mathscr{V}_j^-(x)$ and $\mathscr{V}_j^+(x)$ are well defined as the limits of the sequences, because it is possible to verify that both $\{\mathscr{V}_{j,k}^-(x)\}_{k=1}^\infty$ and $\{\mathscr{V}_{j,k}^+(x)\}_{k=1}^\infty$ are monotonic sequences. It should also be noted that $\mathscr{V}_j^-(x)$ and $\mathscr{V}_j^+(x)$ satisfy the following conditions:

$$\mathscr{V}_j^-(x) = x_j'\beta_j - \alpha_j \Pr(U_{-j} \leq \mathscr{V}_{-j}^+(x)|U_j = \mathscr{V}_j^-(x)),$$

$$\mathscr{V}_j^+(x) = x_j'\beta_j - \alpha_j \Pr(U_{-j} \leq \mathscr{V}_{-j}^-(x)|U_j = \mathscr{V}_j^+(x)).$$

By definition, there is $\mathcal{I}_{j,1}(x; \theta_0) \supseteq \cdots \supseteq \mathcal{I}_{j,k}(x; \theta_0) \supseteq \mathcal{I}_j(x; \theta_0)$ for all $k \in \mathbb{N}$.

Further, let

$$\mathcal{U}_k(\theta_0) = \left\{x \in \mathscr{S}_X : \frac{\partial h_j(u; \theta_0)}{\partial u_j} > \frac{\partial h_j(u; \theta_0)}{\partial u_{-j}} a.e., \forall u \in \mathcal{I}_{1,k}(x; \theta_0) \times \mathcal{I}_{2,k}(x; \theta_0), j = 1, 2\right\}$$

and

$$\mathcal{U}(\theta_0) \equiv \mathcal{U}_\infty(\theta_0) = \left\{x \in \mathscr{S}_X : \frac{\partial h_j(u; \theta_0)}{\partial u_j} > \frac{\partial h_j(u; \theta_0)}{\partial u_{-j}} \; a.e., \; \forall \, u \in \mathcal{I}(x; \theta_0), j = 1, 2\right\}.$$

By definition, the sequence of subsets $\{\mathcal{U}_k(\theta_0)\}_{k=1}^{\infty}$ is monotonically increasing on the support $\mathscr{S}_X$ and $\mathcal{U}(\theta_0)$ is the limit of the sequence.

THEOREM 3.1. *Suppose $X = x \in \mathcal{U}(\theta_0)$. Then, the game has a unique pure strategy equilibrium, which is a monotonic pure strategy BNE. Moreover, this monotonic pure strategy BNE can be characterized by a vector of thresholds $u^*(x) \in \mathcal{I}(x; \theta_0)$, which solves (3.2).*

By the assumption that (U_1, U_2) are bivariate normally distributed, then it can be shown that $x \in \mathcal{U}(\theta_0)$ if

$$1 - \frac{(1+\rho_0)\alpha_j}{\sqrt{2\pi\left(1-\rho_0^2\right)}} \exp\left(-\frac{t^2}{2\left(1-\rho_0^2\right)}\right) \geq 0, \tag{3.3}$$

holds for all $\mathscr{V}_{-j}^-(x) - \rho_0 \mathscr{V}_j^+(x) \leq t \leq \mathscr{V}_{-j}^+(x) - \rho_0 \mathscr{V}_j^-(x)$ and $j = 1, 2$. Thus, if the model parameters satisfy $((1+\rho_0)\alpha_j)/\sqrt{2\pi(1-\rho_0^2)} \leq 1$ for $j = 1, 2$, then (3.3) trivially holds (i.e. $\mathcal{U}(\theta_0) = \mathscr{S}_X$).

4. IDENTIFICATION

In the following analysis, I discuss the identification of the structural parameter θ_0 in the sense of Hurwicz (1950) and Koopmans and Reiersol (1950); that is, whether there is a unique structural parameter $\theta_0 \in \Theta$ to rationalize the conditional distribution of Y given X.

Let $\Theta = \mathbb{B} \times [0, \overline{\alpha}]^2 \times [0, \overline{\rho}]$ be a compact space, where $\mathbb{B} \subseteq \mathbb{R}^{k_1+k_2}$, $0 < \overline{\alpha} < \infty$ and $0 \leq \overline{\rho} < 1$. Suppose that the subset $\mathcal{U}(\theta_0)$ is known and has a strictly positive probability measure.[7] Then, θ_0 is identified under additional rank conditions; that is, conditional on $X \in \mathcal{U}(\theta_0)$, the random vector $(E[Y_1|X], E[Y_2|X])$ has a connected non-degenerate support. To see this, let us first condition on $X = x \in \mathcal{U}(\theta_0)$. Then, there is $E[Y_j|X = x] = \Phi(u_j^*(x))$, where Φ is the cumulative distribution function (CDF) of the standard normal distribution. Therefore, $u_j^*(x) = \Phi^{-1}(E[Y_j|X = x])$.

Further, arbitrarily pick $(p_1, p_2) \in \mathscr{S}_{E[Y_1|X], E[Y_2|X]|X \in \mathcal{U}(\theta_0)}$. It follows that

$$E[Y_1Y_2|E[Y_1|X] = p_1, E[Y_2|X] = p_2, X \in \mathcal{U}(\theta_0)] = \Pr\left(U_1 \leq \Phi^{-1}(p_1); U_2 \leq \Phi^{-1}(p_2)\right),$$

from which ρ_0 is identified. This is because

$$\begin{aligned}
&\frac{\partial E[Y_1Y_2|E[Y_1|X] = p_1, E[Y_2|X] = p_2, X \in \mathcal{U}(\theta_0)]}{\partial p_1} \\
&= \frac{\partial \Pr(\Phi(U_1) \leq p_1; \Phi(U_2) \leq p_2)}{\partial p_1} = \Pr(\Phi(U_2) \leq p_2 | \Phi(U_1) = p_1) \\
&= \Phi\left(\frac{\Phi^{-1}(p_2) - \rho_0 \Phi^{-1}(p_1)}{\sqrt{1-\rho_0^2}}\right),
\end{aligned}$$

[7] For example, suppose $\overline{\rho} \leq 0.5$ and $\overline{\alpha} \leq 1$. It can be shown that $\mathcal{U}(\theta) = \mathscr{S}_X$ for all $\theta \in \Theta$. Thus, $\mathcal{U}(\theta_0)$ is known to be $\mathscr{S}_X$.

in which the second equality follows Darsow et al. (1992) and the fact that $\Phi(U_i)$ is uniformly distributed in [0, 1]. Therefore,

$$\frac{\Phi^{-1}(p_2) - \rho_0 \Phi^{-1}(p_1)}{\sqrt{1-\rho_0^2}} = \Phi^{-1}\left(\frac{\partial E[Y_1 Y_2 | E[Y_1|X] = p_1, E[Y_2|X] = p_2, X \in \mathcal{U}(\theta_0)]}{\partial p_1}\right). \tag{4.1}$$

Note that the right-hand side of (4.1) is known from the conditional distribution of Y given X.[8]

It is straightforward that the term $\rho_0/\sqrt{1-\rho_0^2}$ is identified from (4.1) by taking further derivative with respect to p_1 on both sides of the equation. Because ρ_0 and $\rho_0/\sqrt{1-\rho_0^2}$ are one-to-one mapping, then ρ_0 is also identified. It should also be noted that the identification of ρ_0 allows a non-parametric set-up for the payoff functions as long as $\mathcal{U}(\theta_0)$ is known.

Moreover, given the knowledge of ρ_0 and $u_j^*(X)$, (α_j, β_j) are identified by (3.2), i.e.

$$x_j'\beta_j - \alpha_j \Phi\left(\frac{u_{-j}^*(x) - \rho_0 u_j^*(x)}{\sqrt{1-\rho_0^2}}\right) - u_j^*(x) = 0, \quad \text{for } x \in \mathcal{U}(\theta_0)$$

under an additional rank condition, i.e. the matrix $E[1(X \in \mathcal{U}(\theta_0))Z_j' Z_j]$ has a full rank where

$$Z_j = \left[X_j', \Phi\left(\frac{u_{-j}^*(X) - \rho_0 u_j^*(X)}{\sqrt{1-\rho_0^2}}\right)\right]'.$$

It should be noted that the full rank condition is a testable restriction given the identification of ρ_0 and $u_j^*(\cdot)$.

Under regularity conditions, the point identification of θ_0 implies that it uniquely maximizes the loglikelihood function from the subpopulation,

$$\theta_0 = \text{argmax}_{\theta \in \Theta} E[\mathbf{1}(X \in \mathcal{U}(\theta_0)) \times \ln \Pr_\theta(Y|X)],$$

where

$$\Pr_\theta(Y = y|X = x) = \begin{cases} \Pr_\theta(U_1 \leq u_1^*(x,\theta), U_2 \leq u_2^*(x,\theta)) & \text{if } y = (1,1), \\ \Pr_\theta(U_1 > u_1^*(x,\theta), U_2 \leq u_2^*(x,\theta)) & \text{if } y = (0,1), \\ \Pr_\theta(U_1 \leq u_1^*(x,\theta), U_2 > u_2^*(x,\theta)) & \text{if } y = (1,0), \\ \Pr_\theta(U_1 > u_1^*(x,\theta), U_2 > u_2^*(x,\theta)) & \text{if } y = (0,0). \end{cases}$$

For each given θ, $u^*(x,\theta) = (u_1^*(x,\theta), u_2^*(x,\theta))$ is obtained by the following simultaneous equations.[9] For $j = 1, 2$,

$$x_j' b_j - a_j \Phi\left(\frac{u_{-j}^*(X) - \rho u_j^*(X)}{\sqrt{1-\rho^2}}\right) - u_j^* = 0.$$

[8] Note that the differentiability of $E[Y_1 Y_2 | E[Y_1|X] = p_1, E[Y_2|X] = p_2, X \in \mathcal{U}(\theta_0)]$ requires that conditioning on $X \in \mathcal{U}(\theta_0)$, $(E[Y_1|X], E[Y_2|X])$ has a non-degenerate support in a neighbourhood of (p_1, p_2).

[9] When $\theta \neq \theta_0$, there could be multiple solutions to (3.2) even for $x \in \mathcal{U}(\theta_0)$. In this case, I choose $u_j^*(x,\theta) = x_j' b_j$ as a convention.

4.1. Unknown $\mathcal{U}(\theta_0)$

The difficulty arises when $\mathcal{U}(\theta_0)$ is unknown, which is because of the dependence of $\mathcal{U}(\theta_0)$ on the underlying parameter θ_0. However, the identification method discussed above is silent about how to obtain the knowledge of $\mathcal{U}(\theta_0)$. Therefore, I propose an alternative identification strategy when $\mathcal{U}(\theta_0)$ is unknown.

The procedure has two steps. In the first step, I construct an identified and estimable subset $\Theta^* \subseteq \Theta$, which contains θ_0 and needs to be small enough such that $\mathcal{C}(\Theta^*) \equiv \bigcap_{\theta \in \Theta^*} \mathcal{U}(\theta)$ has a strictly positive probability measure. Because $\mathcal{C}(\Theta^*)$ is a subset of $\mathcal{U}(\theta_0)$, then θ_0 is identified in the second step by replacing $\mathcal{U}(\theta_0)$ with $\mathcal{C}(\Theta^*)$ in the above analysis.[10] This discussion is summarized in the following theorem.

THEOREM 4.1. *Suppose Θ^* is a known subset of Θ such that $\theta_0 \in \Theta^* \subseteq \Theta$ and* $\Pr(X \in \mathcal{C}(\Theta^*)) > 0$. *(a) If, conditional on $X \in \mathcal{C}(\Theta^*)$, $E[Y|X]$ has a non-degenerate connected support in* $[0, 1]^2$ *and (b) if $E[\mathbf{1}(X \in \mathcal{C}(\Theta^*))Z_j' Z_j]$ has a full rank for $j = 1, 2$, then θ_0 is identified.*

Using the identification analysis at the beginning of this section, the proof of Theorem 4.1 is straightforward and is therefore omitted.

4.2. Find Θ^* using level-∞ rationality

It is crucial to construct the subset Θ^* containing θ_0 such that $\Pr(X \in \mathcal{C}(\Theta^*)) > 0$ and ideally $\mathcal{C}(\Theta^*)$ is as large as possible. Hence, here I construct the subset Θ^* as small as possible. Aradillas-Lopez and Tamer (2008) have proposed a novel approach to identify a set that contains θ_0 by using the restrictions called level-k rationality ($k \to \infty$), which are implied by the BNE solution concept.

Under the current set-up, the constraints of level-1 rationality can be derived as follows. Consider the equilibrium response for player $j = 1, 2$,

$$Y_j = \mathbf{1}(X_j' \beta_j - \alpha_j E[Y_{-j}|X, U_j] - U_j \geq 0). \tag{4.2}$$

Because the belief term satisfies $0 \leq E[Y_{-j}|X, U_j] \leq 1$ no matter how her rival behaves, player j's equilibrium response can always be bounded in the following way:

$$\mathbf{1}(\mathscr{V}_{j,1}^-(X) - U_j \geq 0) \leq Y_j \leq \mathbf{1}(\mathscr{V}_{j,1}^+(X) - U_j \geq 0). \tag{4.3}$$

Recall that $\mathscr{V}_{j,k}^-(x)$ and $\mathscr{V}_{j,k}^-(x)$ are defined in Definition 3.1. Therefore, $Y_j = 1$ if $U_j \leq \mathscr{V}_{j,1}^-(X)$, and $Y_j = 0$ if $U_j > \mathscr{V}_{j,1}^+(X)$, which are the restrictions derived from level-1 rationality. Note that level-1 rationality is silent about the rational response of Y_j when $\mathscr{V}_{j,1}^-(X) < U_j \leq \mathscr{V}_{j,1}^+(X)$.

The restrictions of the level-2 rationality can be derived similarly. From (4.3), we have

$$\Pr(\mathscr{V}_{-j,1}^-(X) - U_{-j} \geq 0|X, U_j) \leq E[Y_{-j}|X, U_j] \leq \Pr(\mathscr{V}_{-j,1}^+(X) - U_{-j} \geq 0|X, U_j).$$

Thus for $\mathscr{V}_{j,1}^-(X) \leq U_j \leq \mathscr{V}_{j,1}^+(X)$, it follows that

$$\begin{aligned} &\Pr(\mathscr{V}_{-j,1}^-(X) - U_{-j} \geq 0|X, U_j = \mathscr{V}_{j,1}^+(X)) \\ &\quad \leq E[Y_{-j}|X, U_j] \leq \Pr(\mathscr{V}_{-j,1}^+(X) - U_{-j} \geq 0|X, U_j = \mathscr{V}_{j,1}^-(X)). \end{aligned}$$

[10] Aradillas-Lopez (2010) have also suggested focusing a subset of observations where the BNE is likely to be unique.

By (4.2) and the fact that $\alpha_j \geq 0$, it follows that

$$\mathbf{1}(\mathscr{V}^-_{j,2}(X) - U_j \geq 0) \leq Y_j \leq \mathbf{1}(\mathscr{V}^+_{j,2}(X) - U_j \geq 0). \tag{4.4}$$

Therefore, $Y_j = 1$ if $U_j < \mathscr{V}^-_{j,2}(X)$, and $Y_j = 0$ if $U_j > \mathscr{V}^+_{j,2}(X)$. Note that $\mathscr{V}^-_{j,1}(X) \leq \mathscr{V}^-_{j,2}(X) \leq \mathscr{V}^+_{j,2}(X) \leq \mathscr{V}^+_{j,1}(X)$, which means that higher level of rationality provides more stringent restrictions. Moreover, applying level-k rationality for $k \in \mathbb{N} \cup \{\infty\}$ recursively, there is

$$\mathbf{1}(\mathscr{V}^-_{j,k}(X) - U_j \geq 0) \leq Y_j \leq \mathbf{1}(\mathscr{V}^+_{j,k}(X) - U_j \geq 0). \tag{4.5}$$

Let $\theta = (a_1, a_2, b_1', b_2', \rho)'$ be a generic parameter value in Θ. Furthermore, for any $k \in \mathbb{N}$, let

$$\Theta_k = \{\theta \in \Theta : \Phi(\mathscr{V}^-_{j,k}(x;\theta)) \leq E[Y_j|X = x] \leq \Phi(\mathscr{V}^+_{j,k}(x;\theta)),\ \forall x \in \mathscr{S}_X,\ \ j = 1, 2\}.$$

By definition, Θ_k is monotonically increasing in k and $\Theta_\infty = \cup_{k=1}^{\infty} \Theta_k$.

To choose a proper Θ^* to satisfy the conditions in Theorem 4.1, Θ_k with any natural number k could be a specific candidate for Θ^*. Because the probability $\Pr(X \in \mathcal{C}(\Theta_k))$ is increasing in k and because $\Pr(X \in \mathcal{C}(\Theta_\infty))$ is the largest, I choose $\Theta^* = \Theta_\infty$ to exploit information as much as possible for the identification and estimation.

4.3. Rank condition and the support of covariates

Essentially, $\Pr(X \in \mathcal{C}(\Theta^*)) > 0$ is a rank condition, which requires the support of X to be rich enough. Note that Θ^* is defined by a recursive scheme, which makes it complicated to evaluate (e.g. its size and shape). It is also difficult to provide a numerical example. Moreover, both theoretical and computational difficulties arise when it comes to estimating the subset Θ^*.

Therefore, I define a slightly smaller subset Π of $\mathcal{C}(\Theta^*)$, which can be characterized in a much simpler manner and, more importantly, can be estimated consistently. The characterization of Π also helps to answer the important question of how large the set $\mathcal{C}(\Theta^*)$ is.

Recall that $\overline{\alpha}$ and $\overline{\rho}$ are the upper bounds for the structural parameters α_j and ρ_0, respectively, which therefore defines the compact parameter space $\Theta \equiv \mathbb{B} \times [0, \overline{\alpha}]^2 \times [0, \overline{\rho}]$. In my setting, because the level-k rationalization constraints do not provide any restrictions for an upper bound of α_j, $\mathcal{C}(\Theta^*)$ could be an empty set if Θ is set unbounded. Hence, it is necessary to introduce a finite $\overline{\alpha}$ to ensure the rank condition. The larger $\overline{\alpha}$ is, the more stringent support conditions are required for the covariates X to achieve identification.[11]

Let

$$\gamma_0^* = -\overline{\Delta} + \overline{\alpha} \cdot \Phi\left(\sqrt{\frac{1+\overline{\rho}}{1-\overline{\rho}}} \cdot \overline{\Delta}\right),$$

where

$$\overline{\Delta} = \sqrt{2\frac{1-\overline{\rho}}{1+\overline{\rho}} \ln \max\left\{\frac{(1+\overline{\rho})\overline{\alpha}}{\sqrt{2\pi(1-\overline{\rho}^2)}}, 1\right\}}$$

[11] If $\overline{\alpha}$ is set to be arbitrarily large, then we can assume the full support of $(X_1'\beta_1, X_2'\beta_2)$ on $\mathbb{R}^2$ for identification, by which the proposed identification strategy is similar to the identification-at-infinity argument; see, e.g. Tamer (2003) and Bajari et al. (2010) in the context of complete information games.

Table 1. γ_0^* for different values of $\overline{\alpha}$.

	$\overline{\alpha}=1$	$\overline{\alpha}=1.5$	$\overline{\alpha}=2$	$\overline{\alpha}=2.5$	$\overline{\alpha}=3$	$\overline{\alpha}=4$	$\cdots$
$\overline{\rho}=0$	–	–	–	–	1.5772	2.3659	$\cdots$
$\overline{\rho}=0.4$	–	–	1.0589	1.4500	1.8729	2.7623	$\cdots$
$\overline{\rho}=0.5$	–	0.7537	1.1144	1.5266	1.9620	2.8690	$\cdots$
$\overline{\rho}=0.6$	–	0.7886	1.1830	1.6125	2.0597	2.9830	$\cdots$
$\overline{\rho}=0.7$	–	0.8464	1.2668	1.7120	2.1703	3.1091	$\cdots$
$\overline{\rho}=0.8$	0.5257	0.9299	1.3732	1.8331	2.3023	3.2565	$\cdots$
$\overline{\rho}=0.9$	0.6123	1.0577	1.5234	1.9986	2.4793	3.4504	$\cdots$
$\overline{\rho}\simeq 1$	1.0000	1.5000	2.0000	2.5000	3.0000	4.0000	$\cdots$

Note: The '–' refers to the degenerate case, $((1+\overline{\rho})/(1-\overline{\rho}))\cdot(\overline{\alpha}^2/2\pi)\le 1$, which implies $\mathcal{U}(\theta)=\mathscr{S}_X$ for all $\theta\in\Theta$.

and

$$\Pi=\left\{x\in\mathscr{S}_X: E[Y_1|X=x]\ge\Phi(\gamma_0^*);\ E[Y_2|X=x]\le\Phi(-\gamma_0^*)\right\}$$
$$\bigcup\left\{x\in\mathscr{S}_X: E[Y_1|X=x]\le\Phi(-\gamma_0^*);\ E[Y_2|X=x]\ge\Phi(\gamma_0^*)\right\}.$$

By definition, Π depends on $\overline{\alpha}$ and $\overline{\rho}$ through γ_0^*, which is (weakly) monotonically increasing in $\overline{\alpha}$ and $\overline{\rho}$. For $\overline{\rho}=1$, γ_0^* is defined by its limit (i.e. $\gamma_0^*\to\overline{\alpha}$ as $\overline{\rho}\to 1$). It should be noted that Π can be computed directly using $F_{Y|X}$. In the folowing theorem, I establish the relationship between Π and $\mathcal{C}(\Theta^*)$.

THEOREM 4.2. *By definition,* $\Pi\subseteq\mathcal{C}(\Theta^*)$.

By Theorem 4.2, the rank condition for $\mathcal{C}(\Theta^*)$ will be satisfied if we have $\Pr(X\in\Pi)>0$. Suppose $x_1'\beta_1-\alpha_1\ge\gamma_0^*$ and $x_2'\beta_2\le-\gamma_0^*$. Because $x_j'\beta_j-\alpha_j\le\mathscr{V}_j^-(x)\le\mathscr{V}_j^+(x)\le x_j'\beta_j$ and because (4.5) holds, it follows that $E[Y_1|X=x]\ge\Phi(\gamma_0^*)$ and $E[Y_2|X=x]\le\Phi(-\gamma_0^*)$. Thus, $x\in\Pi$. This means that a large support of $(X_1'\beta_1, X_2'\beta_2)$ is sufficient for $\Pr(X\in\Pi)>0$. Figure 2 (more details are provided later) provides a numerical example in which Π is described by the shaded areas.

Note that γ_0^* is defined in such a way as to ensure that for any $x\in\Pi$, the condition 3.3 holds for all $\theta\in\Theta^*$. By Theorem 3.1, it follows that $x\in\mathcal{C}(\Theta^*)$.

Now, I discuss numerically the choice of γ_0^* for some given $\overline{\alpha}$. First, I show that $\overline{\alpha}/2\le\gamma_0^*\le\overline{\alpha}$. It is also understood that

$$\frac{1+\overline{\rho}}{1-\overline{\rho}}\cdot\frac{\overline{\alpha}^2}{2\pi}>1,$$

otherwise $\mathcal{C}(\Theta^*)$ is known to have the full support. Table 1 provides γ_0^* for different combinations of $\overline{\alpha}$ and $\overline{\rho}$. It should also be noted that the standard deviation of U_i has been normalized to be 1. Hence, the value of $\overline{\alpha}$ imposes an upper bound for the strategic component at the scale of the error's standard deviation.

In Figure 1, the shaded area shows where the marginal choice probabilities $(E[Y_1|X], E[Y_2|X])$ satisfy the restrictions for Π; the marginal choice probability profiles in the area become more and more asymmetric as I relax the restrictions on the parameter space, i.e. increasing $\overline{\alpha}$ or $\overline{\rho}$ (hence increasing γ_0^*). Moreover, Figure 2 illustrates the size of Π in the space of covariates (X_1, X_2) in a simple set-up for $\overline{\alpha}=1.5$ and 2, respectively, and $\overline{\rho}=0.6$.

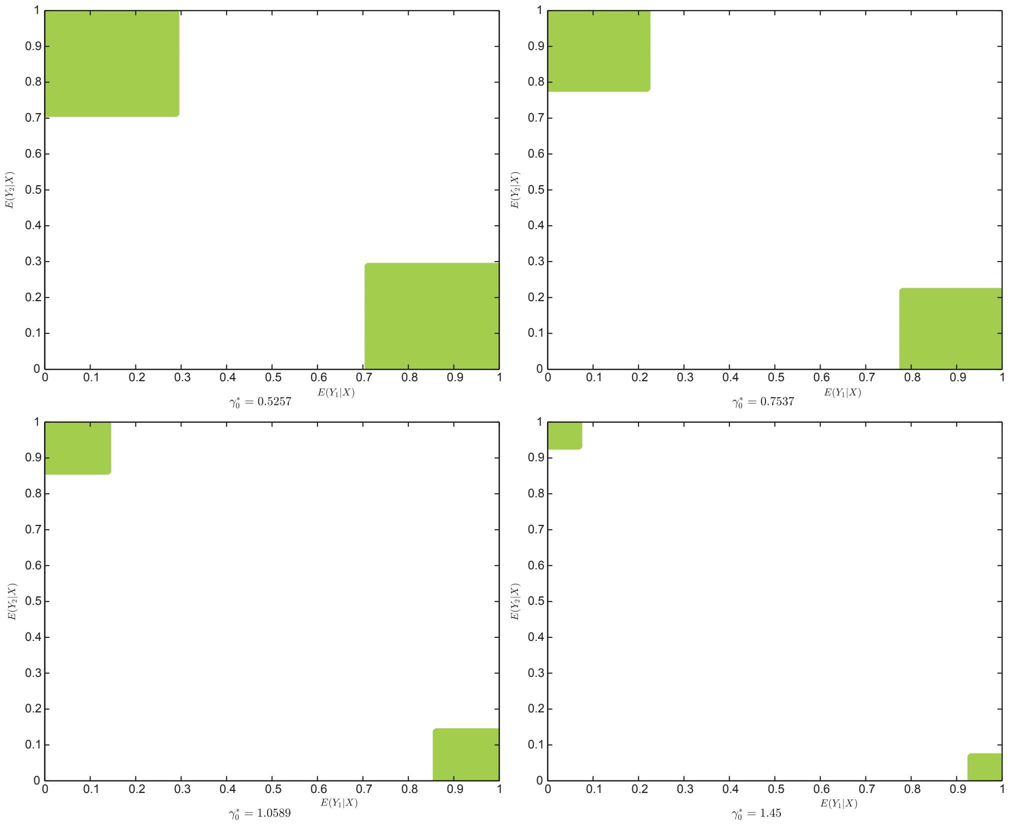

Figure 1. Projections of Π on $(E[Y_1|X], E[Y_2|X])$ for different γ_0^*.

Because the set Π depends on the marginal choice probability $E[Y_j|X]$, then we need to simulate the distribution of Y given X from a particular game structure. In particular, the payoff functions for both players are identical, $X_j\beta - \alpha Y_{-j} - U_j$, in which $\beta = 1$ and $\alpha = 1.5$ and $\mathscr{S}_X = [-1, 5] \times [-1, 5]$; moreover, the correlation coefficient parameter ρ_0 is 0.3 and 0.5, respectively (see Section 6 for more details on generating the distribution of Y given X). The subsets Π are given by the shaded areas.

From Figures 1 and 2, setting a large value of γ_0^* (e.g. $\gamma_0^* = 2$) would clearly cause problems for estimation in practice. Not only would the amount of data included in Π be small, but also these data would contain little information about the structural parameter θ_0, because the expected marginal choice probabilities of those observations would be quite close to the boundaries 0 or 1. The Monte Carlo experiments in Section 6 also confirm this.

5. OUTLINE OF ESTIMATION STRATEGY

The estimation approach is naturally suggested by the identification strategy in Section 4. Suppose that $\{X_i, Y_i\}_{i=1}^n$ is an i.i.d. random sample of size n, where $X_i = (X_{1i}', X_{2i}')'$ and

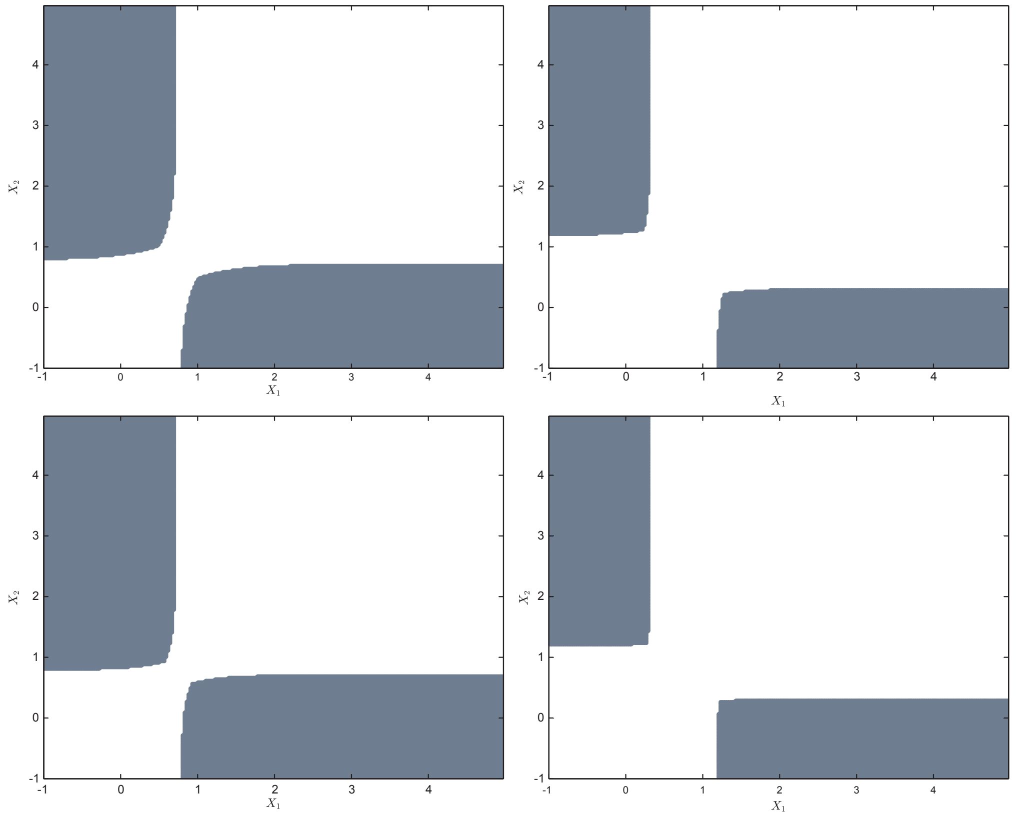

Figure 2. Examples of Π with $\overline{\alpha} = 1.5$ (left) and 2 (right) and $\rho_0 = 0.3$ (upper) and 0.5 (lower).

$Y_i = (Y_{1i}, Y_{2i})'$. The estimation has two steps. I now proceed by introducing my first-step estimator:

$$\tilde{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} 1(X_i \in \widetilde{\Pi}) \log \Pr_\theta(Y_i | X_i). \tag{5.1}$$

Here, $\Pr_\theta(Y|X)$ is the conditional probability of Y given X defined in Section 4, and $\widetilde{\Pi}$ is a consistent estimator of Π such that $\mathbf{1}(X \in \widetilde{\Pi}) - \mathbf{1}(X \in \Pi) \xrightarrow{p} 0$. Note that a uniformly consistent estimator of $E[Y_j|X]$ is sufficient to define $\mathbf{1}(X \in \widetilde{\Pi})$. Suppose X is continuously distributed. Then, let

$$\mathbf{1}(X_i \in \widetilde{\Pi}) \equiv \mathbf{1}\left(\sum_{\ell \neq i} (Y_{1\ell} - \Phi(\gamma_0^*)) K\left(\frac{X_\ell - X_i}{h}\right) \geq 0 \right)$$
$$\times \mathbf{1}\left(\sum_{\ell \neq i} (Y_{2\ell} - \Phi(-\gamma_0^*)) K\left(\frac{X_\ell - X_i}{h}\right) \leq 0 \right)$$

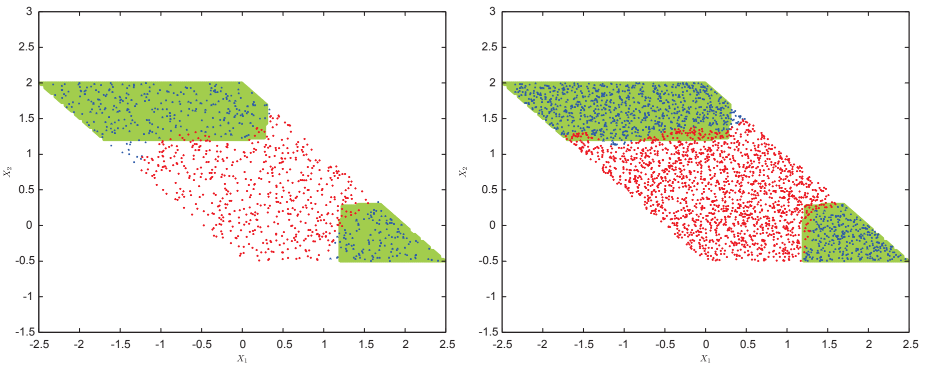

Figure 3. Estimates $\widetilde{\Pi}$ of Π in one simulation.

$$+\mathbf{1}\left(\sum_{\ell\neq i}(Y_{1\ell}-\Phi(-\gamma_0^*))K\left(\frac{X_\ell-X_i}{h}\right)\leq 0\right)$$
$$\times\mathbf{1}\left(\sum_{\ell\neq i}(Y_{2\ell}-\Phi(\gamma_0^*))K\left(\frac{X_\ell-X_i}{h}\right)\geq 0\right),$$

where K and h are the kernel function and the smoothing bandwidth, respectively. Under additional conditions, which are standard in the literature, it could be shown that $\mathbf{1}(X\in\widetilde{\Pi})-\mathbf{1}(X\in\Pi)\xrightarrow{p}0$. If X is discrete, $\mathbf{1}(X\in\widetilde{\Pi})$ can also be defined similarly by plugging into a non-parametric estimator of $E[Y_j|X]$, but it is necessary to rule out the case that the distribution of X has a mass point on the boundary of Π.

ASSUMPTION 5.1. *Let* $\mathbf{1}(X\in\widetilde{\Pi})-\mathbf{1}(X\in\Pi)\xrightarrow{p}0$.

ASSUMPTION 5.2. *Let* $\mathscr{S}_X$ *be compact and* $\Pr(X\in\Pi)>0$.

ASSUMPTION 5.3. *Let* Θ *be compact and* $E[\sup_{\theta\in\Theta}|\ln\Pr_\theta(Y|X)|^{1+\epsilon}]<\infty$ *for some* $\epsilon>0$.

Assumption 5.1 is a high-level condition for the brevity of presentation. It has been well studied in the non-parametric estimation literature. The first half of Assumption 5.2 is standard in the literature and the second half constitutes a rank condition, as discussed in Section 4.3 Assumption 5.3 is slightly stronger than the condition $E[\sup_{\theta\in\Theta}|\ln\Pr_\theta\,(Y|X)|]<\infty$, which is a standard assumption in the MLE literature (e.g. Newey and McFadden, 1986).

THEOREM 5.1. *Suppose Assumptions 5.1–5.3 hold. Then,* $\tilde{\theta}\xrightarrow{p}\theta_0$.

By definition, Π depends on $\overline{\alpha}$ and $\overline{\rho}$, which are specified ad hoc. The larger $\overline{\alpha}$ (or $\overline{\rho}$) I choose, the smaller proportion of data are used for estimation. An alternative way is to apply the commonly used sample-size dependent method for choosing $\overline{\alpha}$ (or $\overline{\rho}$). For example, consider a sequence of $\overline{\alpha}_n$ such that $\overline{\alpha}_n$ increases with the sample size n slowly, for which it is still possible to obtain a consistent estimator of θ_0. However, any meaningful gains of such a procedure in empirical work are phantasmic, because the finite sample behaviour of the estimator could be quite different from its asymptotic distribution.

Table 2. Finite sample behaviour of $\tilde{\beta}_1$ and $\tilde{\alpha}_1$.

ρ_0	n	TRUE	MEAN	MED	SD	RMSE
				$\tilde{\beta}_1$		
0.00	1000	1.00	1.0638	1.0051	0.2495	0.2576
0.00	3000	1.00	1.0003	1.0101	0.0714	0.0714
0.00	5000	1.00	1.0039	0.9946	0.0617	0.0619
0.30	1000	1.00	1.0551	1.0257	0.2095	0.2167
0.30	3000	1.00	1.0223	1.0249	0.0665	0.0702
0.30	5000	1.00	1.0114	1.0099	0.0552	0.0564
0.50	1000	1.00	1.0899	1.0382	0.4129	0.4226
0.50	3000	1.00	1.0023	1.0265	0.0672	0.0711
0.50	5000	1.00	1.0118	1.0055	0.0530	0.0543
				$\tilde{\alpha}_1$		
0.00	1000	1.50	1.4714	1.4684	0.2340	0.2358
0.00	3000	1.50	1.4714	1.4583	0.1169	0.1204
0.00	5000	1.50	1.4888	1.4905	0.0922	0.0929
0.30	1000	1.50	1.5349	1.5070	0.2058	0.2088
0.30	3000	1.50	1.4991	1.4878	0.1202	0.0702
0.30	5000	1.50	1.5122	1.5160	0.0907	0.0915
0.50	1000	1.50	1.5367	1.5027	0.1786	0.1823
0.50	3000	1.50	1.5053	1.5097	0.1166	0.1167
0.50	5000	1.50	1.5119	1.5077	0.0861	0.0869

The consistent estimator $\tilde{\theta}$ allows me to exploit information further in a larger subset of the data (i.e. $\mathcal{V}(\tilde{\theta}, \delta)$), which is a subset of $\mathcal{U}(\tilde{\theta})$ and satisfies regularity conditions. For each $\theta \in \Theta$, let

$$\Delta(\theta) = \sqrt{\frac{1-\rho}{1+\rho} \ln \max\left\{\frac{1+\rho}{1-\rho} \cdot \frac{\alpha_{\max}^2}{2\pi}, 1\right\}},$$

where $a_{\max} = \max\{a_1, a_2\}$. Further, for fixed $\delta > 0$, let $\gamma(\theta) = -\Delta(\theta) + a_{\max} \times \Phi(\sqrt{(1+\rho)/(1-\rho)}\Delta(\theta))$ and

$$\mathcal{V}(\theta, \delta) = \left\{x \in \mathscr{S}_X : x_1'b_1 \geq \gamma(\theta) + \delta(1 + \|x\|); x_2'b_2 - a_2 \leq -\gamma(\theta) - \delta(1 + \|x\|)\right\}$$
$$\bigcup \left\{x \in \mathscr{S}_X : x_1'b_1 - a_1 \leq -\gamma(\theta) - \delta(1 + \|x\|); x_2'b_2 \geq \gamma(\theta) + \delta(1 + \|x\|)\right\}.$$

It can be shown that $\mathcal{V}(\theta, \delta) \subseteq \mathcal{U}(\theta)$ for all $\theta \in \Theta$ and moreover, for any fixed $\delta \in \mathbb{R}^+$, $\{\mathcal{V}(\theta, \delta) : \theta \in \Theta\}$ is a Vapnik–Chervonenkis (VC) class of sets.

LEMMA 5.1. *Fix $\delta > 0$. By definition, $\mathcal{V}(\theta, \delta) \subseteq \mathcal{U}(\theta)$.*

LEMMA 5.2. *Fix $\delta > 0$. The collection $\{\mathcal{V}(\theta, \delta) : \theta \in \Theta\}$ is a VC class of sets.*

Fix $\delta > 0$. By definition, there exists $\epsilon_\delta > 0$ such that for any $\|\theta - \theta_0\| \le \epsilon_\delta$, there is $\mathcal{V}(\theta, \delta) \subseteq \mathcal{V}(\theta_0, 0) \subseteq \mathcal{U}(\theta_0)$. Thus, by consistency of $\tilde{\theta}$, $\Pr(\mathcal{V}(\tilde{\theta}, \delta) \subseteq \mathcal{U}(\theta_0)) \to 1$ as n goes to infinity. Thus, my second-step estimator is defined as

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} 1(X_i \in \mathcal{V}(\tilde{\theta}, \delta)) \log \Pr_\theta(Y_i | X_i). \tag{5.2}$$

To establish the asymptotic properties of $\hat{\theta}$, I make further assumptions. Let $X_j^{[k]}$ be the kth variable in regressors X_j. Similar notation is used for $\beta_j^{[k]}$. Further, let $s(y, x; \theta)$ be the score function, i.e. $s(y, x, \theta) = \partial \log \Pr_\theta(y|x)/\partial \theta$.

ASSUMPTION 5.4. *Let θ_0 be an interior point of Θ.*

ASSUMPTION 5.5. *For $j = 1, 2$, $X_j^{[1]}$ is a continuous argument and $\beta_j^{[1]} \neq 0$. Let $\overline{X}_j$ be all the X variables without $X_j^{[1]}$, i.e. $\overline{X}_j = (X_j^{[2]}, \ldots, X_j^{[k_j]}; X_{-j})$. Assume further $E[\sup_t f_{X_j^{[1]}|\overline{X}_j}(t|\overline{X}_j) \times \|\overline{X}_j\|] < \infty$, where $f_{X_j^{[1]}|\overline{X}_j}$ is the conditional probability density function of $X_j^{[1]}$ given $\overline{X}_j$.*

Assumption 5.4 is standard in the MLE literature; see, e.g. Newey and McFadden (1986). Assumption 5.5 is also used by Manski (1985), which ensures $E[|1(X_i \in \mathcal{V}(\theta, 0)) - 1(X_i \in \mathcal{V}(\theta_0, 0))|] = O(\|\theta - \theta_0\|)$ for θ in a small neighbourhood of θ_0.

THEOREM 5.2. *Suppose Assumptions 5.1–5.5 hold and $\tilde{\theta} \xrightarrow{p} \theta_0$. Then, $\hat{\theta} \xrightarrow{p} \theta_0$. Moreover,*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V_\delta^{-1}),$$

where $V_\delta = E[1\,(X \in \mathcal{V}(\theta_0, \delta)) \cdot s(Y, X; \theta_0) \cdot s'(Y, X; \theta_0)]$.

Similar to Chernozhukov and Hong (2002), we can repeat the above second-step estimation procedure one or more times, using sample $\mathcal{V}(\hat{\theta}, \delta_n)$ in the place of $\mathcal{V}(\tilde{\theta}, \delta)$, where δ_n is a deterministic sequence with $\delta_n \downarrow 0$ (slower than $n^{-1/2}$). The updated estimator will achieve greater efficiency.[12]

6. MONTE CARLO STUDIES

In this section, I use numerical experiments to examine the performance of the proposed estimator and also to illustrate that ignoring the correlation between the private information results in inconsistent estimates and possibly misleading inference. In particular, I investigate the performance of the usual two-step MLE where the players' types are misspecified to be independent.

6.1. *Comparison with the (misspecified) two-step MLE*

If U_1 and U_2 are independent, the two-step MLE comes from the following equation,

$$Y_j = \mathbf{1}(X'_j \beta_j - \alpha_j \Pr(Y_{-j} = 1|X) - U_j \ge 0),$$

[12] Such a result and other details for the asymptotic properties are available from the author upon request.

where $\Pr(Y_{-j} = 1|X)$ can be non-parametrically estimated in the first stage.[13] It is a misspecified model, because $\Pr(Y_{-j}|X) \neq \Pr(Y_{-j}|X, U_j)$ in general.

To begin with, I evaluate the performance of my proposed estimator and the two-step pseudo-MLE in the following setting. I specify the distribution of $X = (X_1, X_2) \in \mathbb{R}^2$ on a compact support as follows. Let Z_1 and Z_2 be two independent random variables with uniform distribution on [0, 2.5]; further, let $X_1 = Z_2 - Z_1$ and $X_2 = 2 - Z_2$. Note that the rank condition in Theorem 4.1 is satisfied under such a specification. Let $\beta_1 = \beta_2 = 1$ and $\alpha_1 = \alpha_2 = 1.5$. In this experiment, I set $\rho_0 = 0, 0.3$ and 0.5, respectively. This helps us to investigate the robustness of the proposed approach, especially when the correlation does not exist. Note that for the setting with $\rho_0 = 0$, the two-step pseudo-MLE is not misspecified, and therefore it produces consistent estimates. Moreover, I vary the sample size by $n = 1000$, 3000 and 5000.

To generate observables $\{(X_i, Y_i) : i = 1, \ldots, n\}$, I need to solve the equilibrium for each observation. Note that it is difficult to compute an equilibrium when its strategies are not monotonic. Therefore, it is hard to simulate the equilibrium distribution of observables when a monotonic pure strategy BNE does not exist for some realizations of X_i. For a similar reason, it is computationally complicated to obtain the whole equilibrium set for those $x_i \notin \mathcal{U}(\theta_0)$. Instead, I mimic the data generated from multiple equilibria or non-monotonic pure strategy BNEs by the following process. Fix $x \in \mathscr{S}_X$. Let $\{(u^*_{1,1}, u^*_{2,1}), \ldots, (u^*_{1,K}, u^*_{2,K})\}$ be the set of solutions to the following equation system[14]

$$x_1\beta_1 - \alpha_1\Phi\left(\frac{u^*_2 - \rho_0 u^*_1}{\sqrt{1-\rho_0^2}}\right) = u^*_1, \qquad x_2\beta_2 - \alpha_2\Phi\left(\frac{u^*_1 - \rho_0 u^*_2}{\sqrt{1-\rho_0^2}}\right) = u^*_2.$$

Let $\bar{u}^*_j(x, \theta_0) = \sum_{k=1}^{K} u^*_{j,k}/K$. Further, I use $Y_j = \mathbf{1}(U_j \leq \bar{u}^*_j(X; \theta_0))$ to generate data.

For the estimation, I choose a compact parameter space: $\Theta = [-5, 5]^2 \times [0, 2]^2 \times [0, 0.6]$, for which $\overline{\alpha} = 2$ and $\overline{\rho} = 0.6$. From Table 1, $\gamma^*_0 = 1.1830$. For each design, I simulate $R = 100$ samples and calculate summary statistics from empirical distributions of estimators from these simulations, including mean (MEAN), median (MED), standard deviation (SD) and root of mean squared error (RMSE). Note that RMSE is estimated using the empirical distribution of estimators and the knowledge of the true parameters in the designs.

In the first-stage estimator, $E[Y_j|X]$ is estimated using a kernel method in which I employ a standard second-order normal kernel with bandwidth $h = 1.06 \times n^{-1/6}$. Figure 3 reports both Π and its estimate $\widetilde{\Pi}$ in one setting with $\rho_0 = 0.5$. The sample size is $n = 1000$ (left panel) and $n = 3000$ (right panel). The shaded area shows Π. The blue crosses represent the draws of X_i that belong to $\widetilde{\Pi}$ while the red circles are the draws outside $\widetilde{\Pi}$. From Figure 3, it is straightforward to see that a considerable amount of data (about 40% of the sample) have been used in the first-stage estimation under this setting.[15] Table 2 reports summary statistics for the first-stage estimator $\tilde{\beta}_1$ and $\tilde{\alpha}_1$. The column of RMSE shows that the mean squared error (MSE) diminishes as the sample size increases from 1000 to 5000, which suggests the consistency of the proposed estimator.

Tables 3 and 4 compare the performance of the proposed estimator with the usual two-step (misspecified) MLE using summary statistics in the three settings with $\rho_0 = 0$, 0.3

[13] In my experiments, I actually compute the term $\Pr(Y_{-j} = 1|X)$ in the first stage, instead of estimating it.

[14] Note that there exists at least one solution in the equation system and the number of solutions K depends on x and θ_0. If the solution is unique, then it characterizes the unique monotonic pure strategy BNE.

[15] In the same experiment, about 70% of the sample is used in the second-stage estimation.

Table 3. Proposed estimator $\hat{\beta}_1$ and misspecified MLE for β_1.

ρ_0	n	TRUE	MEAN	MED	SD	RMSE
		Proposed estimator $\hat{\beta}_1$				
0.00	1000	1.00	0.9921	0.9976	0.0757	0.0761
0.00	3000	1.00	0.9947	0.9984	0.0477	0.0480
0.00	5000	1.00	0.9994	0.9974	0.0331	0.0331
0.30	1000	1.00	1.0049	1.0037	0.0801	0.0803
0.30	3000	1.00	1.0028	0.9992	0.0521	0.0522
0.30	5000	1.00	1.0062	1.0076	0.0357	0.0362
0.50	1000	1.00	1.0019	0.9939	0.0842	0.0842
0.50	3000	1.00	1.0032	0.9976	0.0533	0.0534
0.50	5000	1.00	1.0041	1.0059	0.0363	0.0365
		Misspecified MLE				
0.00	1000	1.00	1.0173	1.0193	0.0665	0.0687
0.00	3000	1.00	1.0024	1.0008	0.0426	0.0427
0.00	5000	1.00	1.0042	1.0016	0.0272	0.0275
0.30	1000	1.00	1.0948	1.0945	0.0740	0.1206
0.30	3000	1.00	1.0796	1.0859	0.0437	0.0912
0.30	5000	1.00	1.0825	1.0812	0.0299	0.0882
0.50	1000	1.00	1.1276	1.1327	0.0796	0.1510
0.50	3000	1.00	1.1156	1.1119	0.0462	0.1250
0.50	5000	1.00	1.1164	1.1135	0.0351	0.1222

and 0.5. For the usual two-step MLE approach, the correlation between private information is falsely assumed away when $\rho_0 \neq 0$. In this approach, instead of non-parametrically estimating the beliefs $E[Y_{-j}|X]$ in the first stage, I use its true value for the second-stage Probit to avoid the finite sample bias from the non-parametric estimation. This should conceivably improve the performance of the final estimator of (α_j, β_j) in the two-step (misspecified) MLE. The summary statistics suggest that misspecified MLEs are inconsistent, except for the case that $\rho_0 = 0$. In contrast, the proposed estimator converges in terms of both bias and variance in all the specifications (including $\rho_0 = 0$) as the sample size increases.

When $\rho_0 = 0$, the two-step MLE is applied to the model that is correctly specified. From Tables 3 and 4, it should be noted that the two-step MLE performs slightly better than the proposed method. The efficiency comes from the fact that in the two-step MLE, the correlation coefficient $\rho_0 = 0$ is assumed to be known and, hence, need not be estimated.

The proposed method also estimates the correlation coefficient parameter ρ_0. Table 5 reports summary statistics for $\hat{\rho}$ in all the settings. There is also evidence of the improvement of the proposed estimator $\hat{\rho}$ in terms of each summary statistic as the sample size increases.

Table 4. Proposed estimator $\hat{\alpha}_1$ and misspecified MLE for α_1.

ρ_0	n	TRUE	MEAN	MED	SD	RMSE
		Proposed estimator $\hat{\alpha}_1$				
0.00	1000	1.50	1.5293	1.5104	0.1504	0.1523
0.00	3000	1.50	1.4901	1.4867	0.0834	0.0840
0.00	5000	1.50	1.4944	1.4932	0.0577	0.0580
0.30	1000	1.50	1.5482	1.5235	0.1747	0.1812
0.30	3000	1.50	1.5039	1.4939	0.0899	0.0900
0.30	5000	1.50	1.5090	1.5013	0.0665	0.0671
0.50	1000	1.50	1.5444	1.5214	0.1567	0.1629
0.50	3000	1.50	1.5099	1.4998	0.0853	0.0859
0.50	5000	1.50	1.5074	1.4990	0.0616	0.0621
		Misspecified MLE				
0.00	1000	1.50	1.5156	1.5128	0.1054	0.1066
0.00	3000	1.50	1.5031	1.4996	0.0666	0.0667
0.00	5000	1.50	1.5013	1.5009	0.0481	0.0481
0.30	1000	1.50	1.6637	1.6628	0.1152	0.2009
0.30	3000	1.50	1.6549	1.6513	0.0718	0.1714
0.30	5000	1.50	1.6549	1.6530	0.0502	0.1635
0.50	1000	1.50	1.7509	1.7338	0.1157	0.2774
0.50	3000	1.50	1.7453	1.7514	0.0734	0.2572
0.50	5000	1.50	1.7472	1.7466	0.0546	0.2544

6.2. Robustness checks

To check the robustness of the proposed approach, it is useful to compare and contrast different specifications, especially by varying the value of γ_0^* as well as the degree of correlation in X to see how these affect the precision of the proposed estimator.

I denote the design in Section 6.1 with $(\overline{\alpha}, \overline{\rho}) = (2, 0.6)$ as 'Setting 0' in the following discussion. Note that in Setting 0, $X_1 = Z_2 - Z_1$ and $X_2 = 2 - Z_2$, where Z_1 and Z_2 are two independent random variables with uniform distribution on [0, 2.5]. Setting 0 represents the case that X are negatively correlated. I also design another two experiments by varying the degree of correlation in X as follows.

SETTING 1. Let $X_1 = Z_1 - Z_2$ and $X_2 = 2 - Z_2$.

SETTING 2. Let $X_1 = Z_1 - 0.5$ and $X_2 = Z_2 - 0.5$.

In Settings 1 and 2, I adjust the correlation in X to be positive and independent, respectively. For all three settings, I set the sample size to be $n = 3000$.

Table 6 reports the performance of the proposed estimators in Settings 0, 1 and 2, respectively. Settings 1 and 2 exhibit finite sample performances similar to that in Setting 0. Our proposed

Table 5. Proposed estimator $\hat{\rho}$.

TRUE	n	MEAN	MED	SD	RMSE
0.00	1000	0.0765	0.0300	0.1293	0.1497
0.00	3000	0.0293	0.0050	0.0461	0.0545
0.00	5000	0.0192	0.0000	0.0284	0.0341
0.30	1000	0.3644	0.3500	0.1797	0.1916
0.30	3000	0.3122	0.3050	0.1002	0.1009
0.30	5000	0.2975	0.3000	0.0700	0.0701
0.50	1000	0.5101	0.6000	0.1396	0.1399
0.50	3000	0.5063	0.5100	0.0906	0.0908
0.50	5000	0.5048	0.5000	0.0711	0.0713

Table 6. Finite sample performance in different settings.

		$\hat{\beta}_1$ ($\beta_0 = 1$)			$\hat{\alpha}_1$ ($\alpha_0 = 1.5$)			$\hat{\rho}$		
Setting	ρ_0	MEAN	SD	RMSE	MEAN	SD	RMSE	MEAN	SD	RMSE
0	0.0	0.9947	0.0477	0.0480	1.4901	0.0834	0.0840	0.0293	0.0461	0.0545
1	0.0	1.0149	0.2332	0.2337	1.4965	0.1034	0.1035	0.0648	0.1232	0.1394
2	0.0	0.9898	0.0491	0.0502	1.4871	0.0992	0.1000	0.0371	0.0722	0.0812
0	0.3	1.0028	0.0521	0.0522	1.5039	0.0899	0.0900	0.3122	0.1002	0.1009
1	0.3	1.0385	0.2344	0.2376	1.5239	0.1113	0.1139	0.3230	0.1310	0.1330
2	0.3	0.9978	0.0518	0.0518	1.5050	0.0917	0.0918	0.3037	0.0927	0.0928
0	0.5	1.0032	0.0533	0.0534	1.5099	0.0853	0.0859	0.5063	0.0906	0.0908
1	0.5	1.0455	0.2448	0.2490	1.5191	0.0945	0.0965	0.4887	0.1147	0.1152
2	0.5	1.0038	0.0550	0.0551	1.5086	0.0909	0.0914	0.5017	0.1104	0.1104

Note: $(\overline{\alpha}, \overline{\rho}) = (2, 0.6)$ and $n = 3000$.

estimator behaves slightly better when X are negatively correlated. This is simply because more observations belongs to Π in Setting 0. Another interesting thing for robustness checks is to see how the estimator behaves under different choices of γ_0^*, which come from different choices of $(\overline{\alpha}, \overline{\rho})$ in empirical applications. As γ_0^* increases, one might drop too much data for estimation. In a simple design with $(\alpha_j, \beta_j) = (1.5, 1)$ and $\rho_0 = 0.5$, I investigate how far one can go for γ_0^* using simulations under different sample sizes.

Table 7 reports the finite sample performance of the proposed estimators under different values of γ_0^*. In Table 7, $\#\widetilde{\Pi}/n$ and $\#\mathcal{V}/n$ denote the proportion of observations used in the first- and second-stage estimations, respectively. Clearly, the proposed estimator behaves badly in the current setting as γ_0^* reaches 1.8. The reason for this is that only a few data points (less than 4%) have been used in the first-stage estimation and all those observations involve quite extreme marginal choice probabilities. Therefore, the first-stage estimator $(\tilde{\alpha}_j, \tilde{\beta}_j)$ behaves poorly and the induced bias carries on to the second-stage estimation.

Table 7. Finite sample performance under different values of γ_0^*.

				$\hat{\beta}_1$			$\hat{\alpha}_1$			$\hat{\rho}$		
n	γ_0^*	$\#\widetilde{\Pi}/n$	$\#\mathcal{V}/n$	MEAN	SD	RMSE	MEAN	SD	RMSE	MEAN	SD	RMSE
1000	1.4	25.2%	56.7%	1.0106	0.0959	0.0960	1.5354	0.1614	0.1653	0.5278	0.1213	0.1244
	1.6	11.7%	62.9%	1.0131	0.1038	0.1046	1.5496	0.1586	0.1662	0.4951	0.1268	0.1269
	1.8	4.2%	60.7%	0.9295	0.2927	0.3012	1.4323	0.4285	0.4579	0.4409	0.1744	0.1842
	2.0	0.9%	39.5%	0.7903	0.4261	0.4754	1.2227	0.6637	0.7199	0.3850	0.2304	0.2578
3000	1.4	27.4%	61.9%	1.0030	0.0537	0.0538	1.5113	0.0866	0.0874	0.5029	0.1044	0.1045
	1.6	13.7%	55.8%	1.0134	0.0599	0.0614	1.5012	0.0873	0.0873	0.5055	0.1208	0.1209
	1.8	3.6%	66.4%	0.9681	0.2042	0.2067	1.4435	0.3042	0.3095	0.4766	0.1236	0.1258
	2.0	0.7%	41.0%	0.6731	0.4797	0.5824	0.9998	0.7181	0.8765	0.3295	0.2453	0.2992
5000	1.4	27.7%	61.7%	1.0038	0.0355	0.0357	1.5114	0.0643	0.0653	0.5035	0.0788	0.0789
	1.6	14.5%	58.8%	1.0039	0.0371	0.0373	1.5094	0.0724	0.0730	0.5028	0.0881	0.0881
	1.8	3.9%	63.9%	0.9636	0.1994	0.2027	1.4459	0.3020	0.3069	0.4861	0.1226	0.1234
	2.0	0.5%	44.3%	0.5794	0.4985	0.6536	0.8594	0.7433	0.9834	0.2916	0.2591	0.3332

Note: $(\beta_1, \alpha_1, \rho_0) = (1, 1.5, 0.5)$.

7. CONCLUSION

It is worth emphasizing that the approach established in this paper hinges crucially on two features of the game model. First, the structure error terms are the private information of the players. The entire results would clearly break down if the underlying game being played is of complete information. As pointed out by Morris and Shin (2003, Section 5), there is no continuity for the equilibrium set because the information structures in a sequence of incomplete information games converge to complete information. Usually, the limit of the equilibrium set depends on the path that the information structures converges; see Table 3.6 of Morris and Shin (2003) for a detailed example. Moreover, because the games of complete and incomplete information form two non-nested models, then the likelihood ratio test of Vuong (1989) can be applied for the model selection issue.

Another possibility is to add unobserved complete information structural terms to the payoffs in the current setting. A model featured with both unobserved heterogeneity and independent private information also generates dependence among players' choices conditional on covariates (see Grieco, 2013). When there are payoff variables (V) observed by both players but not by the researcher, the proposed method does not work. Additional model restrictions would be necessary such that one could obtain $E[Y|X, V]$ from inverting $E[Y|X]$. In some applications, it might not be possible to decompose each player's payoff shock into an unobserved heterogeneity component V_i and a private information component U_i that is i.i.d. across players. However, it is not clear how to extend the framework of Grieco (2013) by allowing type correlation.

Secondly, this paper focuses exclusively on two-player games, which have many practical applications, especially on household decision-making. For example, one could use this model to study the marriage decisions of a couple in an intimate relationship. However, it should also be noted that the proposed approach does not naturally extend to binary games with more than two players. This is because of the issue of multiple equilibria, which generally exists in a large subset of the covariate space when there are more than two players.

It should also be noted that the proposed method could be generalized to a discrete game with ordered multiple choices, but not multinomial games; for an illustration of multinomial games, see, e.g. Bajari et al. (2010). When the error term is a multidimensional random vector rather than a scale, difficulties arise in defining and characterizing monotonic pure strategy BNEs.

Finally, the joint normal distribution of private information is not essential to the proposed method, especially for the marginal normal distribution. See Appendix B for a detailed discussion.

ACKNOWLEDGEMENTS

This paper is a revision of Chapter 2 of my PhD thesis. I am indebted to Joris Pinkse and Sung Jae Jun for their guidance and support. I am grateful to Quang Vuong, to a Co-Editor and to three anonymous referees, whose comments and suggestions have helped to generate a substantially improved paper. I thank Jason Abrevaya, Herman Bierens, Frank Erickson, Edward Green, Han Hong, Hiroyuki Kasahara, Nianqing Liu, Xun Tang, Yuanyuan Wan, and seminar participants at the Far East and South Asia Meeting of the Econometric Society for helpful comments. I also thank Vivian Goldman-Leffler for proofreading this paper. Any and all errors are my own.

REFERENCES

Aguirregabiria, V. and P. Mira (2002). Swapping the nested fixed point algorithm: a class of estimators for discrete Markov decision models. *Econometrica 70*, 1519–43.

Aradillas-Lopez, A. (2010). Semiparametric estimation of a simultaneous game with incomplete information. *Journal of Econometrics 157*, 409–31.

Aradillas-Lopez, A. and E. Tamer (2008). The identification power of equilibrium in simple games. *Journal of Business and Economic Statistics 26*, 261–310.

Aumann, R. J. (1964). Mixed vs. behavior strategies in infinite extensive games. *Annals of Mathematics Studies 52*, 627–30.

Bajari, P., H. Hong, J. Krainer and D. Nekipelov (2010). Estimating static models of strategic interactions. *Journal of Business and Economic Statistics 28*, 469–82.

Berry, S. and E. Tamer (2006). Identification in models of oligopoly entry. In R. Blundell, W. K. Newey and T. Persson (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Volume 2, Econometric Society Monographs 42*, 46–85. Cambridge: Cambidge University Press.

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica 60*, 889–917.

Bjorn, P. A. and Q. H. Vuong (1984). Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation. Working Paper No. 537, Division of the Humanities and Social Sciences, California Institute of Technology.

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

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

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

Brock, W. and S. Durlauf (2001). Discrete choice with social interactions. *Review of Economic Studies 68*, 235–60.

Brock, W. and S. Durlauf (2007). Identification of binary choice models with social interactions. *Journal of Econometrics 140*, 52–75.

Chernozhukov, V. and H. Hong (2002). Three-step censored quantile regression and extramarital affairs. *Journal of the American Statistical Association 97*(459), 872–82.

Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in airline markets. *Econometrica 77*, 1791–1928.

Darsow, W., B. Nguyen and E. Olsen (1992). Copulas and Markov processes. *Illinois Journal of Mathematics 36*, 600–42.

de Castro, L. I. (2007). Affiliation, equilibrium existence and the revenue ranking of auctions. Working paper, University of Illinois at Urbana-Champaign.

De Paula, A. and X. Tang (2012). Inference of signs of interaction effects in simultaneous games with incomplete information. *Econometrica 80*, 143–72.

Grieco, P. (2013). Discrete games with flexible information structures: an application to local grocery markets. Working paper, Pennsylvania State University.

Harsanyi, J. C. (1967). Games with incomplete information played by 'Bayesian' players. Part I. The basic model. *Management Science 14*, 159–82.

Harsanyi, J. C. (1968a). Games with incomplete information played by 'Bayesian' players. Part II. Bayesian equilibrium points. *Management Science 14*, 320–34.

Harsanyi, J. C. (1968b). Games with incomplete information played by 'Bayesian' players. Part III. The basic probability distribution of the game. *Management Science 14*, 486–502.

Hurwicz, L. (1950). Generalization of the concept of identification. In T. Koopmans (Ed.), *Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph 10*, 245–57. New York, NY: John Wiley.

Jia, P. (2008). What happens when Wal-Mart comes to town: an empirical analysis of the discount retailing industry. *Econometrica 76*, 1263–316.

Kline, B. (2012). Identification of complete information games. Working paper, University of Texas at Austin.

Koopmans, T. and O. Reiersol (1950). The identification of structural characteristics. *Annals of Mathematical Statistics 21*, 165–81.

Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Berlin: Springer.

Liu, N., Q. Vuong and H. Xu (2013). Rationalization and identification of discrete games with correlated types. Working paper, Pennsylvania State University.

Manski, C. F. (1985). Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator. *Journal of Econometrics 27*, 313–33.

Morris, S. and H. S. Shin (2003). Global games: theory and applications. In M. Dewatripont, L. P. Hansen and S. J. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Volume 1, Econometric Society Monographs 35*, 56–114. Cambridge: Cambidge University Press.

Newey, W. K. and D. McFadden (1986). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume 4*, 2111–245. Amsterdam: Elsevier.

Pesendorfer, M. and P. Schmidt-Dengler (2003). Identification and estimation of dynamic games. Working Paper No. 9726, National Bureau of Economic Research.

Pollard, D. (1989). Asymptotics via empirical processes. *Statistical Science 4*, 341–54.

Sweeting, A. (2009). The strategic timing of radio commercials: an empirical analysis using multiple equilibria. *RAND Journal of Economics 40*, 710–42.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies 70*, 147–65.

Tang, X. (2010). Estimating simultaneous games with incomplete information under median restrictions. *Economics Letters 108*, 273–76.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*, 307–33.

Wan, Y. and H. Xu (2012). Semiparametric estimation of binary decision of games of incomplete information with correlated private signals. Working paper, University of Toronto.

Xu, H. (2011). Social interactions: a game theoretic approach. Working paper, University of Texas.

APPENDIX A: PROOFS OF RESULTS

I begin with two preliminary lemmata, Lemmata A.1 and A.2. Let $\mathcal{B}$ be the collection of Boreal subsets in $\mathbb{R}$. Further, for any $x \in \mathscr{S}_X$, let

$$\mathcal{K}_j(x) = \{B \in \mathcal{B} : (-\infty, \mathscr{V}_j^-(x)] \subseteq B \quad \text{and} \quad [\mathscr{V}_j^+(x), +\infty) \cap B = \emptyset\}.$$

Note that by level-k rationality with $k = \infty$, player j's equilibrium response must satisfy $Y_j = 1$ for $U_j \leq \mathscr{V}_j^-(x)$ and $Y_j = 0$ for $U_j \geq \mathscr{V}_j^+(x)$ (for a detailed argument, see the discussion in Section 4.2) Hence, I can restrict my attention to the strategy profiles, which take the form

$$s_1(x, u_1) = \mathbf{1}(u_1 \in \mathcal{A}_1), \qquad s_2(x, u_2) = \mathbf{1}(u_2 \in \mathcal{A}_2)$$

where $(\mathcal{A}_1, \mathcal{A}_2) \in \mathcal{K}_1(x) \times \mathcal{K}_2(x)$.

LEMMA A.1. *Suppose $X = x$. Suppose for any given $(\mathcal{A}_1, \mathcal{A}_2) \in \mathcal{K}_1(x) \times \mathcal{K}_2(x)$ and for $j = 1, 2$, the function $u_j + \alpha_j \Pr(U_{-j} \in \mathcal{A}_{-j} | U_j = u_j)$ is an increasing function of $u_j \in \mathcal{I}_j(x_j; \theta_0)$. Then, conditional on $X = x$, all pure strategy BNEs in this game are monotonic strategy BNEs.*

Proof: Fix x. Suppose a strategy profile $\{s_1^*(x, \cdot), s_2^*(x, \cdot)\}$ is a pure strategy BNE. Then, there exists $(\mathcal{A}_1^*, \mathcal{A}_2^*) \in \mathcal{K}_1(x) \times \mathcal{K}_2(x)$, such that $s_j^*(x, u_j) = \mathbf{1}(u_j \in \mathcal{A}_j^*)$ and $\{s_1^*(x, \cdot), s_2^*(x, \cdot)\}$ satisfies the best-response equations (2.1). Because

$$x_j' \beta_j - \alpha_j \Pr(s_{-j}^*(x, U_{-j}) = 1 | U_j = u_j) - u_j = x_j' \beta_j - \alpha_j \Pr(U_{-j} \in \mathcal{A}_{-j}^* | U_j = u_j) - u_j,$$

which is a decreasing function of u_j, there exists $u_j^*(x)$ such that (2.1) can be represented as $s_j^*(x, u_j) = \mathbf{1}(u_j \leq u_j^*(x))$. This implies that the equilibrium strategies have to be monotonic functions. □

Next, I define a subset $\mathcal{M}(\theta_0)$, which contains $\mathcal{U}(\theta_0)$, and I show that all BNEs are monotonic pure strategy BNEs when $X = x \in \mathcal{M}(\theta_0)$. For each $k \in \mathbb{N}$, let

$$\mathcal{M}_k(\theta_0) = \{x \in \mathscr{S}_X : \frac{\partial h_j(u; \theta_0)}{\partial u_j} \geq 0, \ \forall u \in \mathcal{I}_{1,k}(x; \theta_0) \times \mathcal{I}_{2,k}(x; \theta_0), \ j = 1, 2\}$$

and

$$\mathcal{M}(\theta_0) \equiv \mathcal{M}_\infty(\theta_0) = \{x \in \mathscr{S}_X : \frac{\partial h_j(u; \theta_0)}{\partial u_j} \geq 0, \ \forall u \in \mathcal{I}(x; \theta_0), \ j = 1, 2\}.$$

By definition, $\{\mathcal{M}_k(\theta_0)\}_{k=1}^\infty$ is a monotonically increasing sequence of subsets on the support $\mathscr{S}_X$ and $\mathcal{M}(\theta_0)$ is the limit of the sequence. Note that $\mathcal{U}(\theta_0) \subseteq \mathcal{M}(\theta_0)$.

LEMMA A.2. *Suppose $X = x \in \mathcal{M}(\theta_0)$. All pure strategy BNEs in the game with $X = x$ are monotonic pure strategy BNEs.*

Proof: By Lemma A.1, it suffices to show that for any $(\mathcal{A}_1, \mathcal{A}_2) \in \mathcal{K}_1(x) \times \mathcal{K}_2(x)$, $u_j + \alpha_j \Pr(U_{-j} \in \mathcal{A}_{-j} | U_j = u_j)$ is an increasing function of u_j in $\mathcal{I}_j(x_j; \theta_0)$. Without loss of generality, I take $j = 1$. Let ϕ be the probability density function of the standard normal distribution. Because

$$u_1 + \alpha_1 \Pr(U_2 \in \mathcal{A}_2 | U_1 = u_1) = u_1 + \frac{\alpha_1}{\sqrt{1-\rho_0^2}} \int_{\mathcal{A}_2} \phi\Big(\frac{t - \rho_0 u_1}{\sqrt{1-\rho_0^2}}\Big) dt,$$

which is differentiable in u_1, then it is equivalent to show that for all $u_1 \in \mathcal{I}_1(x; \theta_0)$

$$1 - \frac{\rho_0 \alpha_1}{1-\rho_0^2} \int_{\mathcal{A}_2} \phi'\Big(\frac{t - \rho_0 u_1}{\sqrt{1-\rho_0^2}}\Big) dt \geq 0. \tag{A.1}$$

Because $\phi'(t) = -t\phi(t)$ for any $t \in \mathbb{R}$, then

$$1 - \frac{\rho_0 \alpha_1}{1-\rho_0^2} \int_{\mathcal{A}_2} \phi'\Big(\frac{t - \rho_0 u_1}{\sqrt{1-\rho_0^2}}\Big) dt = 1 + \frac{\rho_0 \alpha_1}{\sqrt{1-\rho_0^2}} \int_{\overline{\mathcal{A}}_2(u_1)} s\phi(s) ds,$$

where $\overline{\mathcal{A}}_2(u_1)$ is a linear transformation of the set $\mathcal{A}_2$, i.e.

$$\overline{\mathcal{A}}_2(u_1) = \Big\{\frac{t - \rho_0 u_1}{\sqrt{1-\rho_0^2}} : t \in \mathcal{A}_2\Big\}.$$

Therefore, I need to show, for all $u_1 \in \mathcal{I}_1(x)$

$$1 + \frac{\rho_0 \alpha_1}{\sqrt{2\pi(1-\rho_0^2)}} \int_{\overline{\mathcal{A}}_2(u_1)} s \cdot e^{-s^2/2} ds \geq 0.$$

Note that the left-hand side is minimized by choosing $\mathcal{A}_2$ in $\mathcal{K}_2(x)$ such that $\overline{\mathcal{A}}_2(u_1)$ contains all possible negative elements, i.e. $\mathcal{A}_2^*(u_1) = (-\infty, \mathscr{V}_2^-(x)] \cup \{t \in [\mathscr{V}_2^-(x), \mathscr{V}_2^+(x)] : t - \rho_0 u_1 \leq 0\}$. It is straightforward to see that there exists $\overline{u}_2(u_1) \in \mathcal{I}_2(x; \theta_0)$ such that $\mathcal{A}_2^*(u_1) = (-\infty, \overline{u}_2(u_1)]$.

Hence, to prove that (A.1) holds for all $u_1 \in \mathcal{I}_1(x; \theta)$ and $\mathcal{A}_2 \in \mathcal{K}_2(x)$, it suffices to show that for all $(u_1, \overline{u}_2) \in \mathcal{I}(x; \theta_0)$, there is

$$1 + \frac{\rho_0 \alpha_1}{\sqrt{2\pi}\sqrt{1-\rho_0^2}} \int_{-\infty}^{(\overline{u}_2 - \rho_0 u_1)/(\sqrt{1-\rho_0^2})} s \cdot e^{-s^2/2} ds \geq 0. \tag{A.2}$$

By the definition of $\mathcal{M}(\theta_0)$, it is straightforward that (A.2) holds. □

Proof of Theorem 3.1: Note that $\partial h_j(u; \theta_0)/\partial u_{-j} \geq 0$ for all u and $j = 1, 2$. It follows that $\mathcal{U}(\theta_0) \subseteq \mathcal{M}(\theta_0)$. By Lemma A.2, all equilibria are monotonic pure strategy BNEs when $X = x \in \mathcal{U}(\theta_0)$. Therefore, for the first half of Theorem 3.1, it suffices to show that there is a unique monotonic pure strategy BNE. This statement is proved by contradiction.

Fix $x \in \mathcal{U}(\theta_0)$. Suppose $u^*(x) = (u_1^*(x), u_2^*(x))$ and $v^*(x) = (v_1^*(x), v_2^*(x))$ are the cut-off values that define two different monotonic strategy BNEs. Here, for notational brevity, I suppress the dependence on x of u^* and v^*. Using the level-k rationality argument, both u^* and v^* belong to $\mathcal{I}(x, \theta_0)$. Define $T(\cdot) : \mathcal{I}(x; \theta_0) \to \mathcal{I}(x; \theta_0)$ as follows:

$$x_1'\beta_1 - \alpha_1 \Pr(U_2 \leq u_2 | U_1 = T_1(u)) - T_1(u) = 0;$$
$$x_2'\beta_2 - \alpha_2 \Pr(U_1 \leq u_1 | U_2 = T_2(u)) - T_2(u) = 0. \tag{A.3}$$

Note that $T(\cdot)$ is well defined, i.e. for any fixed $u \in \mathcal{I}(x;\theta_0)$, there exists a unique $T(u)$ satisfying A.3, because of the monotonicity of $\alpha_j \Pr(U_{-j} \leq u_{-j}|U_j = u_j) + u_j$ in u_j on $\mathcal{I}(x;\theta_0)$. Hence, $T(u^*) = u^*$, $T(v^*) = v^*$.

Define a continuously differentiable function $\varphi(t)$ by

$$\varphi(t) = \frac{\langle T(u^*) - T(v^*), T(v^* + t(u^* - v^*))\rangle}{\|T(u^*) - T(v^*)\|}.$$

Note that $\varphi(1) - \varphi(0) = \|T(u^*) - T(v^*)\| = \|u^* - v^*\|$, and also $\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt$. Moreover, $\forall t \in (0, 1)$, we have

$$\begin{aligned}\varphi'(t) &= \frac{\langle u^* - v^*, T'(v^* + t(u^* - v^*))\,(u^* - v^*)\rangle}{\|u^* - v^*\|} \\ &\leq \frac{\|u^* - v^*\| \times \|T'(v^* + t(u^* - v^*))\,(u^* - v^*)\|}{\|u^* - v^*\|} < \|u^* - v^*\| \quad a.e.\end{aligned}$$

The first inequality comes from the Cauchy–Schwartz inequality and the last inequality is based on the fact that $T'_{jj} = 0$ and the condition $x \in \mathcal{U}(\theta_0)$ implies that $|T'_{12}|, |T'_{21}| < 1$ for all $t \in (0, 1)$. Hence, $\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt < \|u^* - v^*\|$, contradiction.

For the second half of Theorem 3.1, note that for $x \in \mathcal{U}(\theta_0)$, $x_j'\beta_j - \alpha_j \Pr(U_{-j} \leq u^*_{-j}|U_j = u_j) - u_j$ is a continuously decreasing function of u_j on the support $\mathcal{I}(x;\theta_0)$. When $U_j = u_j^*(x)$, player j would feel indifferent between choosing $Y_j = 0$ and 1, which implies (3.2). □

Proof of Theorem 4.2: Without loss of generality, let $X = x$ satisfy that $E[Y_1|X = x] \geq \Phi(\gamma_0^*)$ and $\mathbb{E}(Y_2|X = x) \leq \Phi(-\gamma_0^*)$. It suffices to show that for any $\theta \in \Theta^*$, there is $x \in \mathcal{U}(\theta)$.

Fix $\theta \in \Theta^*$. Without loss of generality, let $((1 + \rho)/(1 - \rho)) \cdot (\overline{\alpha}^2/2\pi) > 1$; otherwise, $\mathcal{U}(\theta) = \mathscr{S}_X$ and $x \in \mathcal{U}(\theta)$ would be a trivial statement. By the definition of Θ^*, we have $\Phi(\mathscr{V}_1^+(x;\theta)) \geq E[Y_1|X = x] \geq \Phi(\gamma_0^*)$ and $\Phi(\mathscr{V}_2^-(x;\theta)) \leq \mathbb{E}(Y_2|X = x) \leq \Phi(-\gamma_0^*)$.

Let

$$\Delta^*(\rho) = \sqrt{\frac{1-\rho}{1+\rho} \ln \max\left\{\frac{1+\rho}{1-\rho} \cdot \frac{\overline{\alpha}^2}{2\pi}, 1\right\}}$$

and

$$\gamma^*(\rho) = -\Delta^*(\rho) + \overline{\alpha} \times \Phi\left(\frac{1+\rho}{1-\rho}\Delta^*(\rho)\right).$$

Because the function

$$g(t) \equiv -\sqrt{t \ln \frac{\overline{\alpha}^2}{2\pi t}} + \overline{\alpha} \times \Phi\left(\sqrt{\ln \frac{\overline{\alpha}^2}{2\pi t}}\right)$$

is (weakly) monotonically decreasing in $t \in (0, \overline{\alpha}^2/2\pi]$, $\gamma^*(\rho)$ is monotonically increasing in ρ. Therefore, $\gamma_0^* \equiv \gamma^*(\overline{\rho}) \geq \gamma^*(\rho)$. It follows that

$$\mathscr{V}_1^+(x;\theta) \geq \gamma^*(\rho), \qquad \mathscr{V}_2^-(x;\theta) \leq -\gamma^*(\rho). \tag{A.4}$$

Next, I show that $\mathscr{V}_1^-(x;\theta) \geq \Delta^*(\rho)$ and $\mathscr{V}_2^+(x;\theta) \leq -\Delta^*(\rho)$ by using (A.4).

By the definition of $\mathscr{V}_j^-(x;\theta)$ and $\mathscr{V}_j^+(x;\theta)$, there are

$$\mathscr{V}_1^-(x;\theta) = x_1'b_1 - a_1\Phi\left(\frac{\mathscr{V}_2^+(x;\theta) - \rho\mathscr{V}_1^-(x;\theta)}{\sqrt{1-\rho^2}}\right),$$

$$\mathscr{V}_1^+(x;\theta) = x_1'b_1 - a_1\Phi\bigg(\frac{\mathscr{V}_2^-(x;\theta) - \rho\mathscr{V}_1^+(x;\theta)}{\sqrt{1-\rho^2}}\bigg),$$

$$\mathscr{V}_2^-(x;\theta) = x_2'b_2 - a_2\Phi\bigg(\frac{\mathscr{V}_1^+(x;\theta) - \rho\mathscr{V}_2^-(x;\theta)}{\sqrt{1-\rho^2}}\bigg),$$

$$\mathscr{V}_2^+(x;\theta) = x_2'b_2 - a_2\Phi\bigg(\frac{\mathscr{V}_1^-(x;\theta) - \rho\mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}}\bigg).$$

It follows that

$$\mathscr{V}_1^+(x;\theta) - \mathscr{V}_1^-(x;\theta) = a_1\bigg(\Phi\bigg(\frac{\mathscr{V}_2^+(x;\theta) - \rho\mathscr{V}_1^-(x;\theta)}{\sqrt{1-\rho^2}}\bigg) - \Phi\bigg(\frac{\mathscr{V}_2^-(x;\theta) - \rho\mathscr{V}_1^+(x;\theta)}{\sqrt{1-\rho^2}}\bigg)\bigg),$$

$$\mathscr{V}_2^+(x;\theta) - \mathscr{V}_2^-(x;\theta) = a_2\bigg(\Phi\bigg(\frac{\mathscr{V}_1^+(x;\theta) - \rho\mathscr{V}_2^-(x;\theta)}{\sqrt{1-\rho^2}}\bigg) - \Phi\bigg(\frac{\mathscr{V}_1^-(x;\theta) - \rho\mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}}\bigg)\bigg).$$

Therefore, by (A.4),

$$\mathscr{V}_1^-(x;\theta) + a_1\bigg(\Phi\bigg(\frac{\mathscr{V}_2^+(x;\theta) - \rho\mathscr{V}_1^-(x;\theta)}{\sqrt{1-\rho^2}}\bigg) - \Phi\bigg(\frac{\mathscr{V}_2^-(x;\theta) - \rho\mathscr{V}_1^+(x;\theta)}{\sqrt{1-\rho^2}}\bigg)\bigg) \geq \gamma^*(\rho)$$

and

$$\mathscr{V}_2^+(x;\theta) - a_2\bigg(\Phi\bigg(\frac{\mathscr{V}_1^+(x;\theta) - \rho\mathscr{V}_2^-(x;\theta)}{\sqrt{1-\rho^2}}\bigg) - \Phi\bigg(\frac{\mathscr{V}_1^-(x;\theta) - \rho\mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}}\bigg)\bigg) \leq -\gamma^*(\rho),$$

which implies that

$$\mathscr{V}_1^-(x;\theta) + \overline{\alpha} \times \Phi\bigg(\frac{\mathscr{V}_2^+(x;\theta) - \rho\mathscr{V}_1^-(x;\theta)}{\sqrt{1-\rho^2}}\bigg) > \gamma^*(\rho), \tag{A.5}$$

$$\mathscr{V}_2^+(x;\theta) - \overline{\alpha} \times \bigg(1 - \Phi\bigg(\frac{\mathscr{V}_1^-(x;\theta) - \rho\mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}}\bigg)\bigg) < -\gamma^*(\rho). \tag{A.6}$$

Thus, there exists some $\epsilon > 0$ such that for $\gamma_\epsilon^*(\rho) = \gamma^*(\rho) + \epsilon$,

$$\mathscr{V}_1^-(x;\theta) + \overline{\alpha} \times \Phi\bigg(\frac{\mathscr{V}_2^+(x;\theta) - \rho\mathscr{V}_1^-(x;\theta)}{\sqrt{1-\rho^2}}\bigg) \geq \gamma_\epsilon^*(\rho),$$

$$\mathscr{V}_2^+(x;\theta) - \overline{\alpha} \times \bigg(1 - \Phi\bigg(\frac{\mathscr{V}_1^-(x;\theta) - \rho\mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}}\bigg)\bigg) \leq -\gamma_\epsilon^*(\rho).$$

Moreover, I use a recursive approach to obtain bounds for $\mathscr{V}_1^-(x;\theta)$ and $\mathscr{V}_2^+(x;\theta)$. Let $\ell_{1,1}^-(\theta) \equiv \gamma_\epsilon^*(\rho) - \overline{\alpha}$ and $\ell_{2,1}^+(\theta) \equiv -\gamma_\epsilon^*(\rho) + \overline{\alpha}$. Further, for $k \geq 2$, let

$$\ell_{1,k}^-(\theta) = \gamma_\epsilon^*(\rho) - \overline{\alpha} \times \Phi\bigg(\frac{\ell_{2,k-1}^+(\theta) - \rho\ell_{1,k-1}^-(\theta)}{\sqrt{1-\rho^2}}\bigg),$$

$$\ell_{2,k}^+(\theta) = -\gamma_\epsilon^*(\rho) + \overline{\alpha} \times \bigg(1 - \Phi\bigg(\frac{\ell_{1,k-1}^-(\theta) - \rho\ell_{2,k-1}^+(\theta)}{\sqrt{1-\rho^2}}\bigg)\bigg).$$

Note that $\{\ell_{1,k}^-(\theta)\}_{k\geq 1}$ is a decreasing sequence and $\{\ell_{2,k}^+(\theta)\}_{k\geq 1}$ is increasing. Define $\ell_1^-(\theta) = \lim_k \ell_{1,k}^-(\theta)$ and $\ell_2^+(\theta) = \lim_k \ell_{2,k}^+(\theta)$. By (A.5), $\mathscr{V}_1^-(x;\theta) \geq \ell_{1,1}^-(\theta)$ and $\mathscr{V}_2^+(x;\theta) \leq \ell_{2,1}^+(\theta)$, which further implies that

$\mathcal{V}_1^-(x;\theta) \geq \ell_{1,2}^-(\theta)$ and $\mathcal{V}_2^+(x;\theta) \leq \ell_{2,2}^+(\theta)$, etc. In the limit, there is $\mathcal{V}_1^-(x;\theta) \geq \ell_1^-(\theta)$ and $\mathcal{V}_2^+(x;\theta) \leq \ell_2^+(\theta)$.

Next, I show that $\ell_1^-(\theta) \geq \Delta^*(\rho)$ and $\ell_2^+(\theta) \leq -\Delta^*(\rho)$. Note that $\ell_{1,1}^-(\theta) = -\ell_{2,1}^+(\theta)$, which implies $\ell_{1,2}^-(\theta) = -\ell_{2,2}^+(\theta)$, etc. Thus, $\ell_1^-(\theta) = -\ell_2^+(\theta)$. Therefore, $\ell_1^-(\theta)$ solves the following equation:

$$t + \overline{\alpha}\Phi\Big(-\sqrt{\frac{1+\rho}{1-\rho}} \times t\Big) = \gamma_\epsilon^*(\rho).$$

This is the smallest solution if there are multiple solutions. Because

$$\frac{1+\rho}{1-\rho} \cdot \frac{\overline{\alpha}^2}{2\pi} > 1,$$

then

$$g(t) \equiv t + \overline{\alpha}\Phi\Big(-\sqrt{\frac{1+\rho}{1-\rho}} \times t\Big)$$

is (locally) maximized and minimized at $t = \Delta^*(\rho)$ and $t = -\Delta^*(\rho)$, respectively. Further, by the shape of $g(\cdot)$, the equation $t + \overline{\alpha}\Phi(-\sqrt{(1+\rho/1-\rho)} \times t) = \gamma_\epsilon^*(\rho)$ has a unique solution, which is larger than $\Delta^*(\rho)$, i.e. $\ell_1^-(\theta) \geq \Delta^*(\rho)$.

Therefore, we obtain bounds for $\mathcal{V}_1^-(x;\theta)$ and $\mathcal{V}_2^+(x;\theta)$:

$$\mathcal{V}_1^-(x;\theta) \geq \Delta^*(\rho), \qquad \mathcal{V}_2^+(x;\theta) \leq -\Delta^*(\rho).$$

Thus, $\mathcal{V}_1^-(x;\theta) - \rho\mathcal{V}_2^+(x;\theta) \geq (1+\rho)\Delta^*(\rho)$ and $\mathcal{V}_2^+(x;\theta) - \rho\mathcal{V}_1^-(x;\theta) \leq -(1+\rho)\Delta^*(\rho)$. It follows that

$$1 - \frac{(1+\rho)a_j}{\sqrt{2\pi(1-\rho^2)}} \cdot \exp\left(-\frac{t^2}{2(1-\rho^2)}\right) \geq 0,$$

holds for all $\mathcal{V}_{-j}^-(x;\theta) - \rho\mathcal{V}_j^+(x;\theta) \leq t \leq \mathcal{V}_{-j}^+(x;\theta) - \rho\mathcal{V}_j^-(x;\theta)$ and $j = 1, 2$, which is a sufficient condition for $x \in \mathcal{U}(\theta)$. Because θ is arbitrarily chosen, then $x \in \mathcal{C}(\Theta^*)$. □

Proof of Theorem 5.1: Let $L_n(\theta) = (1/n)\sum_{i=1}^n 1(X_i \in \Pi))\log \Pr_\theta(Y_i|X_i)$ and $G_n(\theta) = (1/n)\sum_{i=1}^n 1(X_i \in \widetilde{\Pi})) \times \log \Pr_\theta(Y_i|X_i)$. By Newey and McFadden (1986) (Theorem 2.5), it suffices to show $L_n(\tilde{\theta}) = \sup_{\theta\in\Theta} L_n(\theta) + o_p(1)$. By the definition of $\tilde{\theta}$, it suffices to show

$$\sup_{\theta\in\Theta} |L_n(\theta) - G_n(\theta)| = o_p(1).$$

Note that

$$\sup_{\theta\in\Theta} |L_n(\theta) - G_n(\theta)| \leq \frac{1}{n}\sum_{i=1}^n |\mathbf{1}(X_i \in \Pi) - \mathbf{1}(X_i \in \widetilde{\Pi})| \times \sup_{\theta\in\Theta} |\ln \Pr_\theta(Y_i|X_i)|.$$

Then, it suffices to show

$$E[|\mathbf{1}(X_i \in \Pi) - \mathbf{1}(X_i \in \widetilde{\Pi})| \times \sup_{\theta\in\Theta} |\ln \Pr_\theta(Y_i|X_i)|] \to 0. \tag{A.7}$$

By Assumptions 5.1 and 5.3 and the Hölder inequality, condition (A.7) holds. □

Proof of Lemma 5.1: Fix θ and δ. Without loss of generality, let $x \in \mathcal{V}(\theta,\delta)$ satisfy $x_1'b_1 \geq \gamma(\theta) + \delta(1 + \|x\|); x_2'b_2 - a_2 \leq -\gamma(\theta) - \delta(1 + \|x\|)$.

Note that

$$\mathscr{V}_1^-(x;\theta) = x_1'b_1 - \alpha_1\Phi\left(\frac{\mathscr{V}_1^+(x;\theta) - \rho\mathscr{V}_2^-(x;\theta)}{\sqrt{1-\rho^2}}\right),$$

$$\mathscr{V}_1^+(x;\theta) = x_2'b_2 - \alpha_2\Phi\left(\frac{\mathscr{V}_1^-(x;\theta) - \rho\mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}}\right).$$

Therefore,

$$\mathscr{V}_1^-(x;\theta) + \alpha_{\max}\Phi\left(\frac{\mathscr{V}_1^+(x;\theta) - \rho\mathscr{V}_2^-(x;\theta)}{\sqrt{1-\rho^2}}\right) \geq x_1'b_1 > \gamma(\theta),$$

$$\mathscr{V}_1^+(x;\theta) - \alpha_{\max}\left(1 - \Phi\left(\frac{\mathscr{V}_1^-(x;\theta) - \rho\mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}}\right)\right) \leq x_2'b_2 - \alpha_2 < -\gamma(\theta)$$

where $\gamma(\theta) = -\Delta(\theta) + \alpha_{\max} \times \Phi(\sqrt{(1+\rho/1-\rho)}\Delta(\theta))$.

By a similar argument as that for Theorem 4.2, there is

$$\mathscr{V}_1^-(x;\theta) \geq \Delta(\theta), \qquad \mathscr{V}_2^+(x;\theta) \leq -\Delta(\theta),$$

which implies that $x \in \mathcal{U}(\theta)$. □

Proof of Lemma 5.2: By Lemma 9.12 in Kosorok (2008), the class $\mathcal{G}_0$ of functions with the form $x_1'c_1 + c_0$ with (c_0, c_1) ranging over $\mathbb{R} \times \mathbb{R}^{k_1}$ is a VC class of functions. The class $\mathcal{G}_1$ of functions $x_1'b_1 - \gamma(\theta)$ with b_1 ranging over $\mathbb{R}^{k_1}$ and $\gamma(\theta) \in \mathbb{R}$ is also a VC class of functions. This is because for any $\theta \in \Theta$, $x_1'b_1 - \gamma(\theta)$ can be written as $x_1'c_1 + c_0$ for some (c_0, c_1). Then, $\mathcal{G}_1$ is a subclass of $\mathcal{G}_0$, and therefore $\mathcal{G}_1$ is also a VC class of functions with no greater index. Moreover, by Lemma 9.9 (v) of Kosorok (2008), the class of functions with the form $x_1'b_1 - \gamma(\theta) - \delta(1 + \|x\|)$ is a VC class of functions for fixed $\delta \in \mathbb{R}^+$. Therefore, the class of sets $\{x \in \mathscr{S}_X : x_1'b_1 \geq \gamma(\theta) + \delta(1 + \|x\|)\}$ is a VC class of subsets. By Lemma 9.7 (ii) of Kosorok (2008), $\{\mathcal{V}(\theta, \delta) : \theta \in \Theta\}$ is a VC class of subsets. □

Proof of Theorem 5.2: For the consistency of $\hat{\theta}$, all the proofs similarly follow that for Theorem 5.1. For the second part of this theorem, by definition of $\hat{\theta}$, there is

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{1}(X_i \in \mathcal{V}(\tilde{\theta}, \delta))s(Y_i, X_i; \hat{\theta}) = 0.$$

By Taylor expansion

$$\frac{1}{n}\sum_{i=1}^{n} 1(X_i \in \mathcal{V}(\tilde{\theta}, \delta))\, s(Y_i, X_i; \theta_0) + \frac{1}{n}\sum_{i=1}^{n} 1(X_i \in \mathcal{V}(\tilde{\theta}, \delta))\left(\frac{\partial}{\partial\theta}s(Y_i, X_i; \theta^\dagger\right)'\left(\hat{\theta} - \theta_0\right) = 0$$

where $\theta^\dagger$ is between $\hat{\theta}$ and θ_0. Hence

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\left(\frac{1}{n}\sum_{i=1}^{n} 1(X_i \in \mathcal{V}(\tilde{\theta}, \delta))\left(\frac{\partial}{\partial\theta}s(Y_i, X_i; \theta^\dagger)\right)'\right)^{-1}$$

$$\times \frac{1}{\sqrt{n}}\sum_{i=1}^{n} 1(X_i \in \mathcal{V}(\tilde{\theta}, \delta))s(Y_i, X_i; \theta_0).$$

By the uniform law of large numbers, Assumption 5.5 and the fact that $1(X_i \in \mathcal{V}(\theta, \delta))$ belongs to a VC class of functions indexed by $\theta \in \mathcal{N}_\epsilon(\theta_0)$, there is

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \in \mathcal{V}(\tilde{\theta}, \delta))\left(\frac{\partial}{\partial\theta}s(Y_i, X_i; \theta^\dagger)\right)' \xrightarrow{p} E\left[\mathbf{1}(X_i \in \mathcal{V}(\theta_0, \delta))\left(\frac{\partial}{\partial\theta}s(Y_i, X_i; \theta_0)\right)'\right].$$

Hence, it suffices to show

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{1}(X_i \in \mathcal{V}(\tilde{\theta},\delta))\, s(Y_i,X_i;\theta_0) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{1}\,(X_i \in \mathcal{V}(\theta_0,\delta))\, s(Y_i,X_i,\theta_0) = o_p(1).$$

Let $h(Y,X;\theta,\delta) = 1(X \in \mathcal{V}(\theta,\delta))s(Y,X;\theta_0)$ and $\mathbb{G}_n(\theta) = n^{-1}\sum_{i=1}^{n} h(Y_i,X_i;\theta,\delta) - E[h(Y,X;\theta,\delta)]$. Because $1(x \in \mathcal{V}(\theta,\delta))$ indexed by θ is a VC class of functions, then by the empirical processes method (see Pollard, 1989), for every sequence of positive numbers $\{\epsilon_n\}$ converging to zero

$$\sup\{n^{1/2}|\mathbb{G}_n(\theta) - \mathbb{G}_n(\theta_0)| : \|\theta - \theta_0\| \le \epsilon_n\} = o_p(1).$$

This implies that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(Y_i,X_i;\tilde{\theta},\delta) &= n^{1/2}\mathbb{G}_n(\widetilde{\theta}) + n^{1/2}E[h(Y,X;\widetilde{\theta},\delta)]\\ &= n^{1/2}(\mathbb{G}_n(\widetilde{\theta}) - \mathbb{G}_n(\theta_0)) + n^{1/2}\mathbb{G}_n(\theta_0) + n^{1/2}E[h(Y,X;\widetilde{\theta},\delta)]\\ &= o_p(1) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(Y_i,X_i;\theta_0,\delta) + n^{1/2}(E[h(Y,X;\widetilde{\theta},\delta)] - E[h(Y,X;\theta_0,\delta)]).\end{aligned}$$

Because (1)$E[h(Y,X;\theta_0,\delta)] = 0$; (2) $\tilde{\theta} \xrightarrow{p} \theta_0$, then $\Pr(\mathcal{V}(\tilde{\theta},\delta) \subseteq \mathcal{V}(\theta_0,0)) \to 1$. Thus $E[h(Y,X;\widetilde{\theta},\delta)] = 0$ with probability approaching to one. Then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(Y_i,X_i;\tilde{\theta},\delta) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(Y_i,X_i;\theta_0,\delta) = o_p(1). \qquad \square$$

APPENDIX B: GENERALIZATION USING COPULA

Let $C(v_1,v_2;\rho_0)$ be the copula function of the joint distribution of U that is known up to a finite-dimensional parameter ρ_0. Further, let F and f be the marginal CDF and probability density function of U_j, respectively. Then, the conditions to define $\mathcal{U}(\theta_0)$ can be written as $x \in \mathcal{U}(\theta_0)$ if and only if

$$1 + \alpha_j \cdot \frac{\partial^2 C(F(u_1),F(u_2);\rho_0)}{\partial v_j^2} \cdot f(u_j) \ge \alpha_j \cdot \frac{\partial^2 C(F(u_1),F(u_2);\rho_0)}{\partial v_1 \partial v_2} \cdot f(u_{-j})$$

for all $u \in \mathcal{I}(x;\theta_0)$ and $j = 1, 2$.

ASSUMPTION B.1. *Let $f(u) = f(-u)$ for all $u \in \mathbb{R}$ and let $f(F^{-1}(\tau))$ be increasing in $\tau \in (0,1/2]$.*

ASSUMPTION B.2. *The copula function C satisfies: (a) $\partial C^2(v_1,v_2;\rho_0)/\partial v_j^2 \le 0$; (b) $\partial C^2(v_1,v_2;\rho_0)/\partial v_1\partial v_2$ is monotonically increasing in v_j and monotonically decreasing in v_{-j} on the support $(v_j,v_{-j}) \in (0,1/2] \times [1/2,1)$; (c) $\partial C^2(v_1,v_2;\rho_0)/\partial v_j^2$ is monotonically decreasing in v_j and monotonically increasing in v_{-j} on the support $(v_j,v_{-j}) \in (0,1/2] \times [1/2,1)$.*

Assumption B.1 imposes weak restrictions on the shape of F, which can be satisfied by, for example, the standard normal distribution, and also implies that $F^{-1}(\tau) = -F^{-1}(1-\tau)$. Assumption B.2 essentially restricts the dependence structure of the joint distribution of types. Assumption B.2(a) is equivalent to the positive regression dependence condition (e.g. de Castro, 2007). Note that

$$\frac{\partial C^2(v_1,v_2;\rho_0)}{\partial v_1 \partial v_2} = \frac{f_U(F^{-1}(v_1),F^{-1}(v_2))}{f(F^{-1}(v_1)) \cdot f(F^{-1}(v_2))}.$$

Therefore, $\partial C^2(v_1, v_2; \rho_0)/\partial v_1 \partial v_2$ is always positive. Assumption B.2 can be satisfied by, for example, a Farlie–Gumbel–Morgenstern (FGM) copula $\mathcal{C}(\nu_1, \nu_2; \rho_0) = \nu_1 \nu_2 (1 + \rho_0 (1 - \nu_1)(1 - \nu_2))$ with $0 \leq \rho_0 \leq 1$. It is straightforward to verify Assumption B.2(a)–(c), because

$$\frac{\partial C^2(v_1, v_2; \rho_0)}{\partial v_1 \partial v_2} = 1 + \rho_0(-v_1 - v_2 + 2v_1 v_2), \qquad \frac{\partial C^2(v_1, v_2; \rho_0)}{\partial v_j^2} = 2\rho_0 v_{-j}(v_{-j} - 1).$$

LEMMA B.1. *Suppose Assumptions B.1 and B.2 hold. Let* $\tau(\theta_0) \in (0, 1/2]$ *solve*

$$1 + \alpha_{\max} \cdot \frac{\partial^2 C(\tau, 1 - \tau; \rho_0)}{\partial v_j^2} \cdot f\Big(F^{-1}(\tau)\Big) \geq \alpha_{\max} \cdot \frac{\partial^2 C(\tau, 1 - \tau; \rho_0)}{\partial v_1 \partial v_2} \cdot f\Big(F^{-1}(1 - \tau)\Big).$$

By Assumptions B.1 and B.2, there is at most one solution. It is understood that if there is no such solution, it corresponds to the degenerated case, i.e. $\mathcal{U}(\theta_0)$ *is the whole support in the covariate space. As a convention, let* $\tau(\theta_0) = 1/2$ *when such a solution does not exist. Then, a sufficient condition for* $x \in \mathcal{U}(\theta_0)$ *is either* $\mathscr{V}_1^-(x) \geq F^{-1}(1 - \tau(\theta_0))$; $\mathscr{V}_2^+(x) \leq F^{-1}(\tau(\theta_0))$, *or* $\mathscr{V}_1^-(x) \leq F^{-1}(\tau(\theta_0))$; $\mathscr{V}_2^-(x) \geq F^{-1}(1 - \tau(\theta_0))$.

Proof: It directly follows from Assumptions B.1 and B.2. □

Further, I define Π as follows. Let $\Pi \equiv \{x \in \mathscr{S}_X : E[Y_1|X] \geq F(\tilde{\gamma}_0^*), E[Y_2|X] \leq 1 - F(\tilde{\gamma}_0^*)\} \cup \{x \in \mathscr{S}_X : E[Y_1|X] \leq 1 - F(\tilde{\gamma}_0^*), E[Y_2|X] \geq F(\tilde{\gamma}_0^*)\}$, where $\tilde{\gamma}^*(\theta) \equiv F^{-1}(\tau(\theta)) + \overline{\alpha} \times (\partial C(\tau(\theta), 1 - \tau(\theta); \rho)/\partial v_1)$ and $\tilde{\gamma}_0^* = \sup_{\theta \in \Theta} \tilde{\gamma}^*(\theta)$. By a similar argument as that in the proof of Theorem 4.2, it can be shown that $\Pi \subseteq \mathcal{C}(\Theta^*)$.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Replication Files