Information Structure and Statistical Information in Discrete Response Models

Shakeeb Khan[a] and **Denis Nekipelov**[b*]

This version: January 2012

Abstract

Strategic interaction parameters characterize the impact of actions of one economic agent on the payoff of another economic agent. In this paper we study how the information available to economic agents regarding other economic agents can influence the ability of an econometrician to recover strategic information parameters from the observed actions. We consider two extreme cases: the complete information case where the information sets of participating economic agents coinside, and the incomplete information case where each agent has a privately observable type. We find that in models with complete information the statistical (Fisher) information for the interaction parameters is zero, implying that estimation and inference become nonstandard. In contrast, in the incomplete information models with any non-zero variance of player types, the statistical information is positive, implying the existence of regular estimators for these parameters converging at the parametric rate. This finding is illustrated in two cases: treatment effect models (expressed as a triangular system of equations), and the static game models. In both types of models the observed discrete outcomes are driven by continuously distributed errors with an unknown distribution (unobserved heterogeneity). We find that the key factor driving the result in these models is the relative tail behavior of the distribution of unobserved heterogeneity of economic agents' payoffs and the distribution of covariates. Our result has important implications for experimental design in economic systems where unobserved heterogeneity plays a major role: a failure to provide sufficient independent randomization of the payoffs of participating agents may lead to non-robustness of estimated parameters to the distribution of unobserved heterogeneity.

JEL Classification: C35, C14, C25, C13.
Keywords: Endogenous discrete response, optimal convergence rate, static game, strategic interaction.

1 Introduction

Endogenous regressors are frequently encountered in econometric models, and failure to correct for endogeneity can result in incorrect inference. Correcting for endogeneity can be particularly difficult in nonlinear models and in the models where endogenous variables are discrete. While two-stage least squares (2SLS) yields consistent estimates in linear models without the need for parametric assumptions on the error disturbances that is not the case for nonlinear models, as the consistency of 2SLS depends critically upon the orthogonality conditions that arise in the linear regression context. For nonlinear models with continuous endogenous regressors, a "control function approach" has been proposed by Blundell and Powell (2004) and Imbens and Newey (2009) provide a counterpart for the models without linear-index and separability restrictions.

The control function approach, however, requires the endogenous regressors to be continuously distributed, and in models where they are discretely distributed, controlling for endogeneity can be difficult. Consequently, existing approaches are not applicable to the models we study in this paper, which focuses on simultaneous discrete response models with discrete endogenous variables. Identification and inference in these models becomes much more complicated than in the continuous case, which is exhibited in the important work in Chesher (2003), Chesher (2007), and Chesher (2010), who considers general classes of nonlinear, nonseparable models. He finds that the discrete model is not point identified under endogeneity and adopts a partial identification approach. It turns out, that with more support points the endogenous variable one can construct tighter bounds for the parameter of interest.

The class of models we consider in this paper includes many important special cases that have received a great deal of attention in both theoretical and empirical work. Important examples include strategic compliance models, models of social interactions, and the simultaneous move discrete game models. In these models we are specifically interested in estimation of the coefficient of the discrete endogenous variable(s). In the case of the triangular binary response model, this parameter is directly related to the treatment effect. In the other class we study, which nests social interaction and static game models, such a parameter characterizes the degree of strategic interaction or a peer effect. In this paper we are interested in identification of these parameters and their *Fisher information*. The Fisher information can be considered an important indicator of the "quality" of identification.

We find that analysis of the endogenous discrete response model as a model of strategic response of an economic agent leads to insights regarding the quality of identification of coefficients of discrete endogenous regressors. There is a fundamental relationship between the choice-theoretic information of the economic agents (reflecting their knowledge regarding

their opponents' types) and the Fisher information of the strategic interaction parameters. To demonstrate our finding, we consider a complete information model where the agents have perfect knowledge regarding payoffs and then we consider a incomplete information model by introducing economic agents' types as additive random shocks to their payoffs. The payoff shocks in this setting are the private information (types) of the agents and cannot be observed by other agents. In this setting the incomplete information model embeds[1] the complete information model, given that the payoffs of agents in the incomplete information model converge to their payoffs in the complete information model as the variance of their types approaches zero. The payoff components in the complete information model that are commonly observed by the participating agents but not observed by the econometrician can be associated with the unobserved heterogeneity. In this case the complete information model is a limiting case of the incomplete information model with unobserved heterogeneity. Grieco (2010) focuses on a similar model where the distribution of unobserved heterogeneity and private shocks are assumed to be normal and the payoffs are parametric linear functions with the goal of testing for the presense of unobserved heterogeneity. In our paper we allow the distribution of unobserved heterogenity and the distribution of covariates to be nonparametric. As a result, we are considering a semiparametric strategic response model with unknown distribution of unobserved heterogeneity.

In the incomplete information setting, the private information of economic agents can be treated as an additional source of uncertainty in the model. Unlike the unobserved heterogeneity that is commonly observed by the economic agents but not observed by the econometrician, private information is not observed by the competing agents. As we show, the parameter of interest (i.e. the treatment effect or the strategic interaction parameter) has a positive Fisher information in the incomplete information model, whereas in the complete information model with zero variance of player types Fisher information is zero.

Recent econometrics work on inference in static game models (see, for instance, Bajari, Hong, Krainer, and Nekipelov (2010b), Aradillas-Lopez (2010) and de Paula and Tang (2011)) demonstrates that identification of strategic interaction parameters can be easier in games with incomplete information because one can find regions of the support of state variables where a unique equilibrium exists. However, in this paper we argue that positive Fisher information for strategic interaction parameters in the incomplete information models is not a result of an equilibrium refinement. This fundamental point is demonstrated in two ways.

First, we consider a triangular system of discrete response models, which is often used to model treatment effects in policy evaluation programs. These models, which are generally

---

[1]In the terminology of Fudenberg, Kreps, and Levine (1988), Tirole (1988), Kajii and Morris (1997).

coherent, in the sense that they do not suffer from multiple equilibria. Nonetheless, we find a stark contrast in cases of complete and incomplete information. For the complete information model, which has been studied in many papers, including Vytlacil and Yildiz (2007), Klein, Shan, and Vella (2011), Abrevaya, Hausman, and Khan (2011), and Jun, Pinkse, and Xu (2011), we establish zero information for the treatment effect parameter. This result implies that inference is nonstandard for the treatment effect. Therefore, for the triangular system, we can approach to the analysis of optimal estimators in terms of optimal rates of convergence (as opposed to efficiency). This approach to analysis of optimal estimators is often used for nonstandard models such as those in Stone (1982), Horowitz (1993), and more recently, Menzel and Morganti (2009) and Komarova, Linton, and Srisuma (2011). As we show, the optimal convergence rates will be directly related to the relative tail behavior of unobservable and observable variables.

We next consider a triangular binary response model and introduce incomplete information as an additive random shock in the second equation. This is a new model, and it represents the environment where the agent decides whether or not to comply with the treatment *before* the treatment is assigned. The treatment assignment remains uncertain to the agent until after he or she makes the decision to comply. As a result, the additional noise plays a loosely analogous role to what a "placebo" usually plays in the natural sciences in aiding with inference on treatment effects.[2] In the game-theoretic terms, this model can be described as a game between the two players, where player 1 is the treated individual and player 2 is the one assigning the treatment. The "type" of player 2 is not known to player 1. Then the decision problem of player 1 can be considered in the same strategic framework as in the standard static game of incomplete information. Our main finding is that the strategic interaction parameter has positive Fisher information in the incomplete information setting, and we derive the semiparametric efficiency bound for this parameter. Considering the incomplete information model with unobserved heterogeneity demonstrates that the complete information model can be viewed as a limiting case of the incomplete information model, where the variance of the random private shock converges to zero. We demonstrate this result by showing that the Fisher information of the incomplete information model, when expressed as a function of this variance, converges to zero as the variance of the random shock decreases.

Consideration of the triangular model naturally leads us to the strategic interaction models, which are represented by nontriangular systems, and argue that our result regarding Fisher information for strategic interaction parameters is not an artifact of equilibrium refinement. These models, which include the simultaneous move static game models as the

---

[2]With the difference that we do not need an unconfoundedness assumption that is frequently required in randomized studies with a placebo.

leading example, (e.g. Bjorn and Vuong (1985), Bresnahan and Reiss (1990), Bresnahan and Reiss (1991b), Bresnahan and Reiss (1991a), Tamer (2003), Andrews, Berry, and Jia (2004), Berry and Tamer (2006), Pakes, Ostrovsky, and Berry (2007), Ciliberto and Tamer (2009), Bajari, Hong, and Ryan (2010), Beresteanu, Molchanov, and Molinari (2011)), are known to suffer from problems of incoherency and multiple equilibria.

The behavior of the set of equilibria in games with vanishing uncertainty regarding the player types has been studied in the game theory literature, e.g. Fudenberg, Kreps, and Levine (1988), Dekel and Fudenberg (1990), Kajii and Morris (1997), and Chassang and Takahashi (2011). In the limiting case with no uncertainty regarding the player types, the model becomes a game of complete information. The literature has established that only particular equilibria of the complete information game are robust to adding noise to the players' payoffs.[3] As a result, the introduction of uncertain player types (by adding private noise to their payoffs) may result in equilibrium refinement.

We find that while strategic interaction parameters have zero Fisher information in the complete information model and positive Fisher information in the incomplete information model, it is not due to equilibrium refinement. We do so by assuming a simple *equilibrium selection rule*, so that the model is coherent and not indeterminate with point identified strategic interaction parameters in both the complete and incomplete information models. A striking fact is that the zero Fisher information result is driven by the relative tail behavior of unobserved heterogeneity and payoff-relevant covariates rather than the structure of the equilibrium set. It has been previously observed in Bajari, Hahn, Hong, and Ridder (2010) that including the equilibrium selection rule in the set of estimated parameters may lead to the zero Fisher information for the estimated payoffs due to the failure of the rank condition. Our result states that even if the equilibrium selection mechanism is known, the Fisher information for strategic interaction parameters in the players payoffs is still zero. Those parameters are, therefore, irregularly identified even in complete information oligopolistic competition models that assume simple equilbrium selection rules such as an incumbent firm's advantage or risk dominance. We emphasize that zero statistical information does not automatically imply the absence of point identification of the parameters of interest: in our problem it means that regular estimators for these parameters converging at the parametric rate may not exist. We derive the optimal convergence rate for estimation of strategic interaction parameters. This rate can be used to evaluate optimality of estimators for strategic interaction parameters in the complete information game.[4]

---

[3]Kajii and Morris (1997), for instance, find that the so-called **p**-dominant equilibria where mixed strategies select actions with a probability exceeding a certain threshold, are robust to adding noise to payoffs.

[4]Another alternative in this case is not even consider identifying the strategic interaction parameters and instead identify bounds for those parameters such as their signs. Such an alternative for incomplete information models has been proposed by de Paula and Tang (2011).

The incomplete information model contains both private payoff shocks and shocks that are commonly observed by the players but not observed by the econometrician (unobserved heterogeneity). We derive a semiparametric efficiency bound for the strategic interaction parameters when the joint distributon of the commonly observed shocks is unknown. This is a new result and distinct from existing efficiency calculations in Bajari, Hong, Krainer, and Nekipelov (2010b) and Aradillas-Lopez (2010) who do not allow for the presence of the unobserved heterogeneity. Furthermore, we show that the Fisher information in the complete information model can be viewed as the limiting case of the Fisher information in the incomplete information models. We recognize that this result *does not* imply the convergence of equilibria in the incomplete information game to those in the complete information games.

Our results indicate the importance of independent randomization in economic experiments where experimental subjects are strategic and unobserved heterogeneity plays an important role. Based on our results, if the agent is strategic about treatment compliance, the estimated treatment effect will not be robust to the unobserved heterogeneity. The convergence rate and the limiting distribution of the estimated parameter will both be strongly affected by the properties of the relative tail behavior of the distribution of unobserved heterogeneity and the observable individual-specific covariates. However, if the subject's payoff is independently randomized, the parameter of interest can be regularly estimated at the parametric rate. The experimental randomization in this has the effect of smoothing the strategic responses of experimental subjects.

The rest of the paper is organized as follows. In the following section we introduce a basic binary choice model with a binary endogenous variable determined by a reduced-form model. We find that the coefficient for the endogenous variable has zero Fisher information, which is a result similar to that in Khan and Tamer (2010) for the binary choice model with endogeneity in Lewbel (1998), and related to that in Chamberlain (1986) and Chen and Khan (1999) for heteroskedastic binary choice models (see also Graham and Powell (2009) for an example in panel data models). As this result implies the difficulties with inference for the parameter of interest, we further explore possible asymptotic properties for conducting inference in this model. We then consider the triangular system with incomplete information, which is the strategic behavior model with the agent playing against nature where nature has a type that is not observed by the agent. For the incomplete information framework we show in Section 3 that the Fisher information for the parameter of interest is positive, such that inference becomes standard. We then derive the semiparametric efficiency bound for the interaction parameter. In Section 4 we explore non-triangular systems using the example of a two-player simultaneous move game in which the incoherency is resolved by a choice randomization in cases of multiple equilibria or where pure strategy equilibria do not exist. We show the this "simplified" model has zero information for parameters

of interest. As with the triangular system, inference becomes complicated, even though the strategic interaction parameters are point identified, so we explore this issue further by finding their optimal convergence rates. In Section 5 we consider a game of incomplete information in which each player has a type, represented by a random shock to her payoff and players cannot observe the types of their opponents. This incomplete information game embeds the complete information game in Section 4, provided that types of players are represented by additive shocks to the payoffs of players in the incomplete information game. The presence of random payoff perturbations does not completely resolve the problem of multiple equilibria, but by introducing an equilibrium selection mechanism (as we did for the complete information game), we can now attain positive information for the strategic interaction parameters. The contrast illustrates that the positive (Fisher) information is not a result of equilibrium refinement, as both of the models are endowed with an equilibrium selection rule. Finally, Section 6 concludes the paper by summarizing and suggesting areas for future research. An appendix collects all the proofs of the theorems and additional results regarding the optimal rates and estimators attaining optimal rates.

2   Discrete response model

2.1   Information in discrete response model

Let Y_1 denote the dependent variable of interest, which is assumed to depend upon a vector of covariates Z_1 and a single endogenous variable Y_2.[5]

For the binary choice model with with a binary endogenous regressor in linear-index form with an additively separable endogenous variable, the specification is given by

$$Y_1 = \mathbf{1}\{Z_1'\gamma_0 + \alpha_0 Y_2 - U > 0\}. \tag{2.1}$$

Turning to the model for the endogenous regressor, the binary endogenous variable Y_2 is assumed to be determined by the following reduced-form model:

$$Y_2 = \mathbf{1}\{Z'\delta_0 - V > 0\}, \tag{2.2}$$

where $Z \equiv (Z_1, Z_2)$ is the vector of "instruments" and (U, V) is a pair of random shocks. The subcomponent Z_2 provides the exclusion restrictions in the model and is required to be nondegenerate conditional on $Z_1'\gamma_0$. We assume that the error terms U and V are jointly independent of Z. The endogeneity of Y_2 in (2.1) arises when U and V are not independent, while the estimation of the model in (2.2) is standard. When dealing with a binary

---

[5]We use capital letters to denote random variables and small letters to indicate regular arguments.

endogenous regressor, we will use the common terminology "treatment effect" rather then referring to the "causal effect of Y_2 on Y_1". Thus, for example, a positive treatment effect would correspond to the case of equation (2.1) where Y_2 can take on only two values.

This type of model fits into the class of models considered in Vytlacil and Yildiz (2007). In this paper we are interested in the parameter α_0, which is related to a treatment effect. To simplify exposition, we will assume the parameters δ_0 and γ_0 are known. What this part of the paper will focus on is the *information* for α_0 (see, e.g., Ibragimov and Has'minskii (1981), Chamberlain (1986), Newey (1990) for the relevant definitions). To simplify the notation, we introduce single indices $x_1 = z_1'\gamma_0$ and $x = z'\delta_0$.[6] The discrete response model can then be written as

$$Y_1 = \mathbf{1}\{X_1 + \alpha_0 Y_2 - U \geq 0\},$$
$$Y_2 = \mathbf{1}\{X - V \geq 0\}. \tag{2.3}$$

To give a full characterization of the class of distributions of errors and covariates that we consider, we introduce the following assumption:

Assumption 1 *(i) Single indices X_1 and X have a joint distribution with the full support on $\mathbb{R}^2$ which is not contained in any proper one-dimensional subspace;*

(ii) (U, V) are independent of X_1 and X and have an absolutely continuous density with full support on $\mathbb{R}^2$ and joint cdf $G(\cdot, \cdot)$. The partial derivative $\frac{\partial G(u,v)}{\partial u}$ exists and strictly positive on $\mathbb{R}^2$;

(iii) For each $t \in \mathbb{R}$ and fixed γ_0 and δ_0, there exists function $q(\cdot, \cdot)$ with $E[q(X_1, X)^2] < \infty$ which dominates $\frac{\partial G(x_1+t, x)}{\partial t}$

We recognize that this assumption is stronger than many assumptions that are used to identify semiparametric models with endogenous binary variables. We do so with an explicit intent to demonstrate that *our (negative) results regarding the quality of identification of parameter of interest α_0 occur even in this simple setup.*

We begin our analysis by noticing that we can construct examples of parametric distributions for the errors and covariates in the triangular model in which the variance of the score for parameter α_0 is infinite. The simplest way to construct such examples is to consider cases of high correlation between errors U and V. This can reflect the situation where both equations in the triangular system are driven by common shocks that are not observed by

---

[6] Although additional complications may arise if covariates Z are not truly exogenous, under our assumptions one can regularly identify parameters γ_0 and δ_0. The issues of inference for these parameters are discussed, for instance in Abrevaya, Hausman, and Khan (2011).

the econometrician. The infinite variance of the score in the parametric example which we give in Appendix D is driven by the large support of covariates and the thin tails of the normal distribution. We observe that one cannot construct a similar example for estimation of the constant in the single equation binary choice model. This result indicates that the information in the triangular model may have a different nature: while in the single equation model the information is determined by the smoothness of the joint distribution of errors and covariates, in the triangular model it is determined by the relative tail behavior of the distributions of errors and covariates. The zero information result can be "repaired" in parametric models by assuming that covariates have bounded support with density bounded away from zero on that support. This assumption may not be suitable in semiparametric models: when the distributions of covariates and the unobserved shocks are unknown, the restriction on the covariate support often leads to a loss of point identification of the parameter of interest.

As a result, when we allow the model to be semiparametric with unknown distributions of errors and covariates, we can find parametric submodels that have zero information. It turns out that these submodels can be constructed for each smooth distribution of errors U and V. The structure of the least favorable submodel is such that the shocks U and V are highly correlated eveywhere on the support. The single indices, on the other hand are highly correlated at the tails. As a result, the parameter of interest is nearly not identified. We formally state this result in the following theorem[7]:

Theorem 2.1 *Under Assumption 1, the Fisher information associated with parameter α_0 in model (2.3) is zero.*

We find that under our conditions the parameter α_0 cannot be estimated at the parametric rate. Even though this result has similar structure to the impossibility theorems in Chamberlain (1986) and Cosslett (1987), the issue of irregular identification of payoff parameters in strategic models was not well understood so far. We argue that the intercept in the semiparametric binary choice model and the coefficient of the endogenous dummy variable are different in terms of elements of the data distribution that identify them. While the intercept is identified from the location of the mean or the median of the marginal distribution of the errors, the coefficient of the endogenous dummy is identified from the mutual dependence between the error terms in two equations of the binary system. Hence, even if one imposes sufficient restrictions on the marginal distribution to regularly identify the intercept, the coefficient of the endogenous dummy may still be irregularly identified. In the subsequent sections we show that a similar result applies to the information of the strategic interaction parameter in static game models. The conditions of the theorem imply that for

---

[7]The proof of this and all subsequent theorems is provided in Appendix A

any distribution of errors we can find a parametric submodel for which the score will have an infinite variance. This does not mean that all parametric submodels will have the infinite variance of the score; for instance, if the class of densities of U and V covers all joint logistic densities, then normal distributions of covariates can deliver finite scores, and hence positive information. The assumption of the theorem rules out the cases when one only considers such distributions.

Remark 2.1 *This result, first shown in Khan and Nekipelov (2010), was alluded to in Abrevaya, Hausman, and Khan (2011), where they conducted inference on the sign of α_0, and indicated why the positive information found in Vytlacil and Yildiz (2007) was due to a relative support condition on unknown parameters. The delicacy of point identification was also made apparent in Shaikh and Vytlacil (2011), who partially identified this parameter. As we will see later in this paper, this zero information result can be overturned by introducing a little more uncertainty in the model (e.g. by reducing the information available to the treated agent (Player 1 in the game) regarding the treatment).*

2.2 Optimal rate for estimation of the interaction parameter

The fact that the information associated with the "interaction" parameter is zero does not imply that the parameter cannot be estimated consistently. We now describe the set of results regarding the convergence rates of the semiparametric estimator for α_0.

We take a constructive approach to establish the optimal convergence rate for the estimator for α_0. We begin with a definition of the optimal rate following Ibragimov and Has'minskii (1978). Let $\mathcal{G}$ characterize a class of joint densities of error terms (U, V) and single indices X_1 and X. First, we recall that for the class of distributions $\mathcal{G}$, we define the risk using a positive (rate) sequence r_n and a constant L as

$$R(\hat{\alpha}, r_n, L) = \sup_{f,g \in \mathcal{G}} P_{f,g} \left(r_n |\hat{\alpha} - \alpha_0| \geq L \right).$$

Using this notion of the risk, we introduce the definition of the convergence rates for the estimator of the parameter of interest.

Definition 2.1 *(i) We call the positive sequence r_n the lower rate of convergence for the class of densities $\mathcal{G}$ if there exists $L > 0$ such that*

$$\liminf_{n \to \infty} \inf_{\hat{\alpha}} R(\hat{\alpha}, r_n, L) \geq p_0 > 0.$$

(ii) We call the positive sequence r_n the upper rate of convergence if there exists an estimator $\hat{\alpha}_n$ such that

$$\lim_{L \to \infty} \limsup_{n \to \infty} R(\hat{\alpha}_n, r_n, L) = 0.$$

(iii) The positive sequence r_n is the minimax (or optimal) rate of convergence if it is both a lower and an upper rate.

We make a constructive argument to derive the upper convergence rate by providing an estimator that attains the upper rate of convergence in Definition 2.1(ii). The convergence rate of the resulting estimator relies on the tail behavior of the joint density of the error distribution. To be more specific about the class of considered error densities, we formulate assumptions that restrict the "thickness" of tails of the error distribution in addition to Assumption 1, which requires that the density of this distribution is smooth and the random shocks U and V are independent from the covariates X_1 and X. These assumptions are satisfied by many distributions that are conventional in applied research.[8]

Assumption 2 *Denote the joint cdf of unobserved payoff components U and V as $G(\cdot, \cdot)$, where $G_v(\cdot)$ is the marginal cdf of V. Let $\mathcal{G}$ be the class of distributions of errors $g(\cdot, \cdot)$ and covariates $f(\cdot, \cdot)$ which satisfy the assumptions of Theorem 2.1 and the following additional conditions:*

(i) There exists a non-decreasing function[9] $\nu(\cdot)$ such that for any $|t| < \infty$

$$\lim_{c \to +\infty} \frac{1}{\nu(c)} \sup_{f,g \in \mathcal{G}} E_{f,g} \left[\left(\frac{\partial G(X_1 + t, X)}{\partial t} \right)^2 G(X_1 + t, X)^{-1} \right.$$
$$\left. (G_v(X) - G(X_1 + t, X))^{-1} \, \middle| \, |X_1|, |X| < c \right] < \infty$$

(ii) There exists a non-increasing function $\beta(\cdot)$ such that for any $|t| < \infty$

$$\lim_{c \to +\infty} \beta(c) \sup_{f,g \in \mathcal{G}} E_{f,g} \left[\log \left(G(X_1 + t, X) \right. \right.$$
$$\left. \left. (G_v(X) - G(X_1 + t, X)) \right) \, \middle| \, |X_1|, |X| > c \right]^{-1} < \infty$$

This assumption allows the inverse joint cumulative distribution function to be non-integrable in the $\mathbb{R}^2$ plane (its improper integral diverges). It is, however, integrable on any square with finite edge and its integral can be expressed as a function of the length of the edge. A rough evaluation for such a function $\nu(\cdot)$, can come from evaluating the highest value attained by the inverse cumulative distribution of errors on $[-c, c] \times [-c, c]$. If the distribution of single indices decays sufficiently fast at the tails, this evaluation can be improved.

---

[8]We give concrete examples of such distributions including normal and logistic in Appendix C

[9]We use the same c to trim the support of covariates X and X_1 for notational and algebraic convenience only. Our analysis has a straightforward extension to the case where the relative tail behaviors of X_1 and X are different. In that case $\nu(\cdot)$ will be a function of two arguments.

The imposed assumptions play two roles. First, they require that for each finite c the expectation of inverse joint cdf and the logarithm of the joint cdf of the error terms are finite. Second, provided that those quantities are finite, they essentially allow the researcher to define functions $\nu(\cdot)$ and $\beta(\cdot)$ to offset the divergence of the corresponding conditional expectations to infinity. Assumption 2(ii) requires the population likelihood function of the model to be finite (provided that $\beta(\cdot)$ is a non-increasing function). In addition, if the support of the indices x_1 and x is restricted to a square with the edge of some large length c, the resulting restricted likelihood will be sufficiently close to the true population likelihood.

We maintain our Assumption 1 ensuring smoothness and measurability of the density of error terms U and V. We also need to add a technical assumption regarding the complexity of the class of considered densities of error terms to make sure that this density is estimable at a sufficiently fast rate. Following Kim and Pollard (1990), we refer to the class of densities satisfying our assumptions as *uniformly manageable*. We give the formal definition of this class in Appendix A.2.

The class of uniformly manageable densities of errors satisfying Assumptions 1 and 2 characterizes the class of error distributions that we will consider in our analysis. This is a large class of functions admitting many classes of distributions commonly used in applied research. We use a constructive approach to derive the optimal rates and first propose the estimator that attains the upper convergence rate.

We consider the following procedure to estimate α_0.[10] First, we look at the probability of the outcome $(0, 0)$ conditional on linear indices x_1 and x. This probability does not depend on the interaction parameter, and its derivative with respect to the linear indices will be equal to the joint error density. Therefore we estimate the joint probability of the outcome $(0, 0)$ and then differentiate it with respect to the arguments. The estimated density will be approximated by K terms in an orthogonal expansion.

Second, when the estimate of the error density is available, it can be substituted into the expression for the probabilities of outcomes $(1, 1)$ and $(0, 1)$ which depend on the interaction parameter. We then form the trimmed quasi-likelihood using the trimming sequence c_n. We need to trim the sample likelihood to avoid the divergence of its Hessian when covariates take large values. We define the estimator as the maximizer of the trimmed quasi-likelihood

$$\hat{\alpha}_{0,n}^* = \text{argmax}_\alpha \frac{1}{n} \sum_{i=1}^n \hat{l}_{K,c_n} \left(\alpha; y_{1i}, y_{2i}, x_{1i}, x_i \right). \tag{2.4}$$

where the subscript K, c_n denotes that the likelihood depends on the the tuning parameters K, c_n. The estimator will have a convergence rate that depends on the tail behavior of the error terms and the selected trimming sequence c_n. In Appendix A.2 we demonstrate that

---

[10] We provide a formal exposition of this estimator in Appendix B.1.

we can select the trimming sequence c_n such that the rate of the estimator $\hat{\alpha}_{0,n}^*$ attains the upper convergence rate for estimation of α_0.

The next question that must be addressed is whether the sequence c_n that makes the particular estimator $\hat{\alpha}_{0,n}^*$ attain the upper convergence rate is also the optimal rate of estimator $\hat{\alpha}_{0,n}$. The answer to this question involves combining our results with a fundamental result regarding the lower rate for semiparametric estimators provided in Koroselev and Tsybakov (1993) The following theorem outlies our main result regarding the optimal rate for the interaction parameter in the triangular model.

Theorem 2.2 *Suppose that Assumptions 1 and 2 hold. Suppose that $c_n \to \infty$ is a sequence such that $\frac{n\beta^2(c_n)}{\nu(c_n)} = O(1)$ with $n/\nu(c_n) \to \infty$. Then for this sequence $\sqrt{\frac{n}{\nu(c_n)}}$ is the optimal rate for the estimator for parameter α_0 in model (2.3).*

Remark 2.2 *We note that the stated conditions on c_n in the statement of the theorem resemble the usual bias variance tradeoff in nonparametric estimation. For the problem at hand, c_n converging to infinity will ensure the bias shrinks to 0, but unfortunately this can also cause the variance to explode. As in nonparametric estimation, there will be an optimal rate of c_n that balances this tradeoff to minimize mean squared error.*

This theorem shows that the majorant $\nu(\cdot)$ for the expectation of the inverse cumulative distribution of errors plays the role of the pivotizing sequence. Similar to the construction of the t-statistics where the de-meaned estimator is normalized by the standard deviation, we normalize the estimator by a function of the trimming sequence.

The above result reveals how widely the optimal rates vary, depending on the tail properties of the observed indices. Appendix C illustrates by considering widely used parametric distributions such as the normal and logistic distributions.

3   Triangular model with incomplete information

3.1   Identification and information of the model

In the previous section, we considered a classical triangular discrete response model and demonstrated that in general, that model has zero Fisher information for the interaction parameter α_0. Our results suggested that the optimal convergence rate for the estimator of the interaction parameter will be sub-parametric and will depend on the relative tail behavior of the error terms (U, V) and covariates. In this section, we set up a model which

can be arbitrarily "close" to the classical triangular model but have positive information.[11] We construct this model by adding a small noise to the second equation in the triangular system. Adding arbitrarily small but positive noise to this equation discontinuously changes the optimal rate to the standard parametric rate. One example of this approach is adding artificial noise to the treatment assignment in a controlled experiment, such that the experimental subjects do not know the specific realizations of the experimental noise but know its distribution. As a result, they will be responding to the *expected* treatment instead of the actual treatment. This model provides a strategic flavor to the conventional triangular model. As we emphasize it further, this model structure is interesting as it can be considered a special case of the static game of incomplete information.

Consider the model where the endogenous variable Y_2 defined as

$$Y_2 = \mathbf{1}\{X - V - \sigma\eta > 0\}.$$

We express our assumption regarding the additional noise component η formally:

Assumption 3 *Suppose that $\eta \perp (U, V)$ and $\eta \perp (X_1, X)$. The distribution of η has a differentiable density with the full support on $\mathbb{R}$ and a cdf $\Phi(\cdot)$ which is known by the economic agent and the econometrician.*[12]

Variable Y_1 reflects the response of agent who does not observe the realization of noise η but observes the error term V. As a result, the response in the first equation can be characterized as:

$$Y_1 = \mathbf{1}\{X_1 + \alpha_0 E_\eta[Y_2|X, V] - U > 0\}$$

where the parameter of interest is α_0 for which we wish to derive the information.[13] We can express the conditional expectation in the above term as $E_\eta[Y_2|x, v] = \Phi((x - v)/\sigma)$. The constructed discrete response model can the be written as

$$\begin{aligned} Y_1 &= \mathbf{1}\{X_1 - U + \alpha_0 E[Y_2|X, V] > 0\}, \\ Y_2 &= \mathbf{1}\{X - V + \sigma\eta > 0\}. \end{aligned} \tag{3.1}$$

---

[11]An alternative approach to address the zero information issue is to change the object of interest to some non-invertible function of the interaction parameter as in Abrevaya, Hausman, and Khan (2011), which focused on the sign of the treatment effect.

[12]We show further in the paper that this choice is a normalization

[13]We note that here α_0 may denote a different treatment parameter than before. Specifically, it now measures the response to probability of treatment, as opposed to treatment itself. We argue that it is still a useful parameter to conduct inference on, for two reasons: First, as the amount of noise becomes arbitrarily small, the probability of treatment becomes arbitrarily close to the standard treatment status indicator, and the new parameter approximates the standard parameter (the remainder of this section elaborates on this argument with more precision). Second, even if the amount of noise (quantified by σ) is not small, the new parameter will have the same sign as the old one.

Incorporating expectations as explanatory models is similar in spirit to work considered in Ahn and Manski (1993). In doing so, we are able to place the triangular binary model into the framework of modeling responses of economic agents to their expectations such as in Manski (1991), Manski (1993) and Manski (2000).

This model also has features of the continuous treatment model considered in Hirano and Imbens (2004), Florens, Heckman, Meghir, and Vytlacil (2008) and Imbens and Wooldridge (2009). While in the latter cases the economic agent responds to an intrinsically continuous quantity (such as dosage), in our case the continuity of treatment is associated with uncertainty of the agent regarding the treatment. Notably, even the triangular model in the previous section has a discrete response interpretation characterizing the optimal choice of an economic agent. This approach has been proven useful in the modern treatment effect literature, such as in Abadie, Angrist, and Imbens (2002), Heckman and Navarro (2004), Carneiro, Heckman, and Vytlacil (2010). Outside of the treatment effect setting, analysis of binary choice models with a continuous endogenous variable is also studied in Blundell and Powell (2004), who demonstrate the attainability of positive information for the coefficient on the endogenous variable.

The incomplete information triangular model presented here also places the standard triangular model considered in the previous section in the context of the models with strategic compliance of the treated subjects, as in Chassang and Snowberg (2010). The complete information triangular model will characterize the compliance behavior in the LATE model of Angrist and Imbens (1995), Abadie, Angrist, and Imbens (2002), and Imbens (2009) as a special case: the orthogonality assumption of LATE will be satisfied if the error terms U and V, in our terminology, are independent. Variable Y_2 corresponds to the "treatment assignment" (i.e. the binary instrument of the LATE model) and the variable y_1 corresponds to the compliance decision. The complete information model represents the case where the treated subject knows all of the inputs into the treatment decision. As a result, the compliance decision will be correlated with the treatment decision unless the unobservables in the two decisions are orthogonal. Once the treatment decision contains noise, which may come from the deliberate treatment randomization (e.g. through a placebo) or can suffer from the measurement error, the treated subject may only react to the expected treatment. This setting motivates the triangular model with treatment uncertainty.

We can illustrate the structure of the model using Figure 1. Panel (a) in Figure 1 corresponds to the classical binary triangular system and panels (b)-(d) correspond to the triangular system with incomplete information. The panels show the areas of joint support of U and V corresponding to the observable outcomes y_1 and y_2. When there is no noise in the second equation of the triangular system, the error terms U and V completely determine the outcome. On the other hand, when the noise with unbounded support is added to the

second equation, one can only determine the probability that the second indicator is equal to zero or one. Figures 1.b-1.d show the area where, for a given quantile q, the probability of Y_2 equal zero or one exceeds $1-q$. The noise in the second equation decreases from panel (d) to panel (b), which in the limit will approach to to the figure on panel (a).
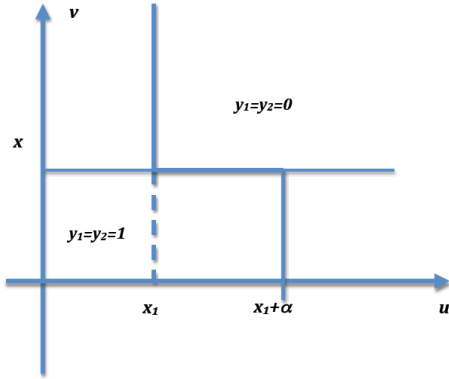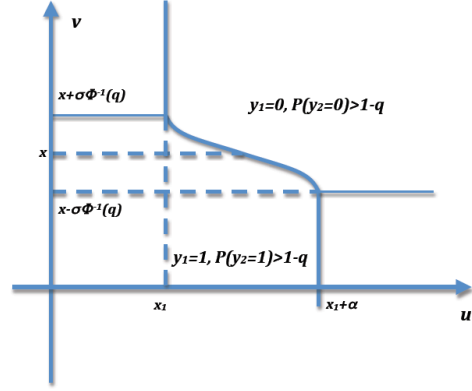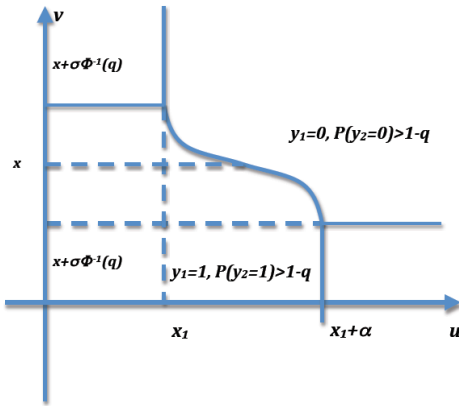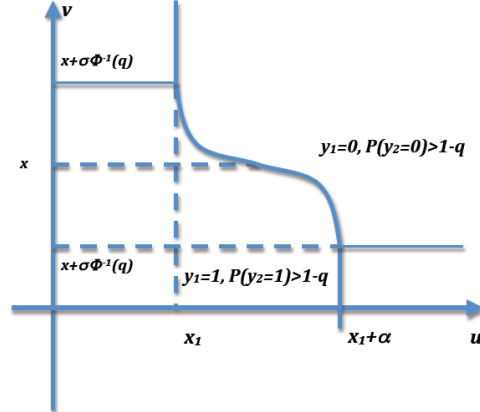

Figure 1.a

Figure 1.b

Figure 1.c

Figure 1.d

This discrete response model is related to game theory models with random payoff perturbations. If we associate discrete variable Y_1 with a discrete response, then the linear index in the first equation corresponds to the economic agent's payoff. As a result, this model is not a payoff perturbation model but rather a treatment perturbation model. The treatment perturbation can be considered in the experimental settings where the subjects are exposed to the placebo treatment with some fixed probability but they do not observe whether or not they get the placebo. In this case they will respond to the expected treatment. The error terms U and V in this setup can be interpreted as unobserved heterogeneity in the economic agent's payoff (determining Y_1) and in the treatment assignment rule (determining Y_2).

Given that this is a new model, we will need to establish first that the model is identified from the data. The following theorem considers the identification of the interaction parameter α_0.

Theorem 3.1 *Under Assumptions 1 and 3, the interaction parameter α_0 in model (3.1) is identified.*

We note that parameter α_0 is identified under conditions that are weaker than those that we needed to estiablish the optimal rate of convergence in the complete information triangular model (2.3). This becomes possible due to the presense of additional noise η which "smoothes out" the response of the economic agent. There is also a tradeoff between identification of the marginal distribution of the error term V and the distribution of noise η. The observable conditional probability of the indicator of the second equation equal to one can be written as

$$\mathbf{P}_1(x) = \int \Phi\left(\frac{x-v}{\sigma}\right) g_v(v)\, dv, \tag{3.2}$$

where $g_v(\cdot)$ is the marginal density of v. This expression represents a convolution of the marginal density and $\Phi(\cdot)$, the cdf of the noise distribution. Given that the Fourier transform of the convolution is equal to the product of Fourier transforms, the transform of the (observable) left-hand side is equal to the product of the Fourier transform of the cdf of the noise distribution and the marginal characteristic function of the distribution of V. If the distribution of V is known, then the cdf of the noise is identified via its Fourier transform. When the cdf of the noise is fixed, we can identify the marginal distribution of V.

The deconvolution argument provides a simple and convenient method to explicitly express the strategic interaction parameter α_0 using the observed conditional expectations of dummy variables Y_1 and Y_2, along with their interaction, conditional on covariates X_1 and X. In Appendix E we demonstrate how one can use the deconvolution argument along with the techniques for operating with generalized functions to identify the distribution of unobserved heterogeneity (U, V) along with the parameter of interest.

After establishing identification of the parameter of interest, we analyze its Fisher information. We find that for any finite variance σ^2 (which can be arbitrarily small) the information for α_0 in the incomplete information triangular model is strictly positive. Moreover, the Fisher information of the strategic interaction parameter α_0 approaches zero when the variance of noise shrinks to zero. In other words, *the smaller the informational asymmetry between agents, the smaller the Fisher information of the interaction parameter α_0.*

We state this in the following theorem:

Theorem 3.2 *Suppose that Assumptions 1 and 3 are satisfied.*

(i) *For any $\sigma > 0$ the information in the triangular model of incomplete information (3.1) is strictly positive.*

(ii) *As $\sigma \to 0$ the information in the triangular model of incomplete information (3.1) converges to zero.*

We note that this theorem also suggests an alternative estimator for the strategic interaction parameter in the complete information model. One can consider estimation of the complete information model assuming that it is "sufficiently" close to the incomplete information model and then for a fixed distribution of η choose a sequence of standard deviations $\sigma_n \to 0$ as $n \to \infty$. This approach is essentially a kernel smoothing-based estimator for parameter α_0.

3.2 Convergence rate for the interaction parameter

The previous subsection proved that the triangular model with incomplete information has positive Fisher information for any amount of noise added to the second equation. Our results, therefore, guarantee that the semiparametric efficiency bound is finite. We note that the analyzed model has two unknown nonparametric components: the distribution of covariates and the distribution of unobserved heterogeneity. Due to the independence of the unobserved heterogeneity and the observed covariates and the fact that the distribution of covariates does not depend on parameter α_0, this parameter is fully characterized by the expectations of and the covariance between the observed binary variables Y_1 and Y_2 conditional on covariates. In other words, the parameter of interest is characterized by a system of conditional moment equations. We explicitly compute the efficiency bound in Appendix F.1 and our results are based on the result for the semiparametric efficiency bound in conditional moment systems provided in Ai and Chen (2003). Our efficiency result provides the semiparametric efficiency bound for the new discrete response model.

Our final result expresses to the optimal convergence rate for the interaction parameter in the triangular model of incomplete information. Our result states that *the optimal rate of convergence is parametric and the minimum variance of the estimator converging at a parametric rate corresponds to the semiparametric efficiency bound.* Formally this result is formulated in the statement of the following theorem which combines the result of Theorem 3.2 and Theorem IV.1.1 in Ibragimov and Has'minskii (1981).

Theorem 3.3 *Under Assumptions 1 and 3 for any sub-convex loss function $w(\cdot)$ and standard normal random variable ξ:*

$$\liminf_{n\to\infty} \inf_{\hat{\alpha}_{0,n}} \sup_{(f,g)\in\mathcal{G}} E_{f,g}\left[w\left(\sqrt{n}(\hat{\alpha}_{0,n} - \alpha_0(f,g))\right)\right] \geq E[w\left(\Omega^{1/2}\xi\right)],$$

where $g(\cdot,\cdot)$ is the distribution of errors U and V, $f(\cdot,\cdot)$ is the distribution of covariates, $\hat{\alpha}_{0,n}$ is the estimator for the strategic interaction parameter and Ω is the semiparametric efficiency bound.[14]

4 Nontriangular Systems: A Static game of complete information

4.1 The Fisher information in the complete information game

In this section we consider the information of parameters of interest in a simultaneous discrete system of equations where we no longer impose the triangular structure of the previous sections. A leading example of this type of system is a 2-player discrete game with complete information (e.g. Bjorn and Vuong (1985) and Tamer (2003)). We will later extend this model to one with incomplete information in a manner analogous to our approach to the triangular system.

A simple binary game of complete information is characterized by the players' deterministic payoffs, strategic interaction coefficients, and random payoff components u and v. There are two players $i = 1, 2$ and the action space of each player consists of two points $A_i = \{0, 1\}$ with the actions denoted $y_i \in A_i$. The payoff of player 1 from choosing action $y_1 = 1$ can be characterized as a function of player 2's action:

$$y_1^* = z_1'\gamma_0 + \alpha_1 y_2 - u,$$

and the payoff of player 2 from choosing action $y_2 = 1$ is characterized as

$$y_2^* = z_2'\delta_0 + \alpha_2 y_1 - v.$$

For convenience of analysis we change notation to $x_1 = z_1'\gamma_0$ and $x_2 = z_2'\delta_0$. We normalize the payoff from action $y_i = 0$ to zero and we assume that realizations of covariates X_1 and X_2 are commonly observed by the players along with realizations of the errors U and V, which are not observed by the econometrician and thus characterize the unobserved heterogeneity in the players' payoffs. Under this information structure the pure strategy of each player

---

[14] We provide an explicit expression for Ω in terms of the primitives of the model in Appendix F.1.

is the mapping from the observable variables into actions: $(u, v, x_1, x_2) \mapsto 0, 1$. A pair of pure strategies constitute a Nash equilibrium if they reflect the best responses to the rival's equilibrium actions. The observed equilibrium actions are described by random variables (from the viewpoint of the econometrician) characterized by a pair of binary equations:

$$
\begin{aligned}
Y_1 &= \mathbf{1}\{X_1 + \alpha_1 Y_2 - U > 0\}, \\
Y_2 &= \mathbf{1}\{X_2 + \alpha_2 Y_1 - V > 0\},
\end{aligned}
\tag{4.1}
$$

where errors U and V are correlated with each other with an unknown distribution. In particular, we are interested in determining when the strategic interaction parameters α_1, α_2 can or cannot be estimated at the parametric rate. We formalize our restriction on the joint distribution of U and V in the following assumption, which is analogous to Assumption 1 in the triangular model.

Assumption 4 *Suppose that*

(i) *X_1 and X_2 have a continuous distribution with full support on $\mathbb{R}^2$ (which is not contained in any proper one-dimensional linear subspace);*

(ii) *(U, V) are independent of (X_1, X_2) and have a continuously differentiable density with the full support on $\mathbb{R}^2$ and joint cdf $G(\cdot, \cdot)$. Partial derivatives $\frac{\partial G(u,v)}{\partial u}$ $\frac{\partial G(u,v)}{\partial v}$ exist and strictly positive on $\mathbb{R}^2$;*

(iii) *For each $t_1, t_2 \in \mathbb{R}$, there exist functions $q_1(\cdot)$ and $q_2(\cdot)$ with $E[q_1(X_1, X_2)^2] < \infty$ $E[q_2(X_1, X_2)^2] < \infty$ which dominate $\frac{\partial G(x_1+t_1, x_2+t_2)}{\partial u}$ and $\frac{\partial G(x_1+t_1, x_2+t_2)}{\partial v}$, respectively.*

As noted in Tamer (2003), the system of simultaneous discrete response equations (4.1) has a fundamental problem of indeterminacy. To resolve this problem we impose the following additional assumption which is similar to the assumption of the existence of an equilibrium selection mechanism in game theory:

Assumption 5 *Denote $S_1 = [\alpha_1 + x_1, x_1] \times [\alpha_2 + x_2, x_2]$, $S_2 = [x_1, \alpha_1 + x_1] \times [x_2, \alpha_2 + x_2]$, $S_3 = [\alpha_1 + x_1, x_1] \times [x_2, x_2 + \alpha_2]$, and $S_4 = [x_1, x_1 + \alpha_1] \times [\alpha_2 + x_2, x_2]$. Note that $S_1 = \emptyset$ iff $\alpha_1 > 0, \alpha_2 > 0$ and $S_2 = \emptyset$ iff $\alpha_1 < 0, \alpha_2 < 0$.*

(i) *If $S_1 \neq \emptyset$ or $S_2 \neq \emptyset$ then $Pr(y_1 = y_2 = 1 | (u, v) \in S_k) \equiv \frac{1}{2}$ for $k = 1, 2$.*

(ii) *If $S_3 \neq \emptyset$ or $S_4 \neq \emptyset$ then $Pr(y_1 = (1 - y_2) = 1 | (u, v) \in S_k) \equiv \frac{1}{2}$ for $k = 3, 4$.*

Assumption 5 requires that when the system of binary responses has multiple solutions, then the realization of a particular solution is resolved over a symmetric coin flip. In regions where the system may have no solutions (corresponding to a unique mixed strategy equilibrium), we impose solutions via randomization. This assumption addresses the incoherency in the model. We select this simple setup *to emphasize that the complete information model has zero information even when there is no incoherency.* In principle, one can generalize this condition to cases where the distribution over multiple outcomes depends on additional covariates. However, given that the structure of results under this generalization remains the same, we do not consider it in this paper.

We now prove identification of strategic interaction parameters, arguing that the zero information result is not a consequence of poor identifiability. Our identification result, generally speaking, is new. We leave the distribution of unobserved payoff components to be fully non-parametric (and non-independent, unlike Bajari, Hong, and Ryan (2010), who assume independence and normality of unobserved components U and V, and Grieco (2010), who drops independence by keeps the assumption of normality) while imposing a linear index structure on the payoffs.[15]

Theorem 4.1 *Suppose that Assumptions 4 and 5 are satisfied. Then the interaction parameters α_1 and α_2 in model (4.1) are identified.*

Having established the identifiability of the parameters of interest, we now study the information associated with the strategic interaction parameters. The following result establishes that the information associated with the interaction parameters in the static game of complete information is zero. The important takeaway is that in the light of the identification result in Theorem 4.1, this result is not related to the incoherency of the static game and is a reflection of discontinuity of equilibrium strategies.

Theorem 4.2 *Suppose that Assumptions 4 and 5 are satisfied. Then the Fisher information associated with parameters α_1 and α_2 in model (4.1) is zero.*

Our result fully illustrates why the zero Fisher information of the interaction parameter is a problem that is not related to the multiplicity of equilibria. We have explicitly completed the model using randomization of oucomes so that it is coherent, yet we still cannot attain positive information. The estimation and inference of the interaction parameters are nonstandard even in a simplified model - a result analogous to that found for the triangular system in the previous sections. Here we aim to address the optimality of estimators of

---

[15]The proof of identification can be found in the companion paper "Information Bounds and Impossibility Theorems for Simultaneous Discrete Response Models".

the interaction parameters by deriving their optimal convergence rates. As we show in the next section, the convergence rate for the estimators of strategic interaction parameters will be affected by the considered class of distributions of unobserved heterogeneity. Provided that identification in this case relies on the full support of linear indices, the optimal rate of convergence for the estimator of the interaction parameters will be sub-parametric and reflect the relative tail behavior of the distribution of the unobserved payoff components. The parameter estimates will therefore be non-robust to the distribution of the unobserved heterogeneity.

4.2 Optimal rate for estimation of strategic interaction parameters

To analyze the optimal rates of convergence for the strategic interaction parameters we need to modify Assumption 2 to account for the presence of the interaction between both discrete response equations.

Assumption 6 *Denote the joint cdf of unobserved payoff components u and v as $G(\cdot, \cdot)$ and the joint density of single indices $f(\cdot, \cdot)$. Then assume that the following conditions are satisfied for these distributions.*

(i) There exists a non-decreasing function $\nu(\cdot)$ such that for any $|t| < \infty$ and $|s| < \infty$

$$\lim_{c \to \infty} \frac{1}{\nu(c)} \sup_{f,g \in \mathcal{G}} E_{f,g} \left[\max \left\{ \left(\frac{\partial G(X_1 + t, X_2 + s)}{\partial t} \right)^2, \left(\frac{\partial G(X_1 + t, X_2 + s)}{\partial s} \right)^2 \right\} \right.$$

$$\left. G(X_1 + t, X_2 + s)^{-1} \left(1 - G(X_1 + t, X_2 + s) \right)^{-1} \Big| \, |X_1|, |X_2| < c \right] < \infty$$

(ii) There exists a non-increasing function $\beta(\cdot)$ such that for any given $|t| < \infty$ and $|s| < \infty$

$$\lim_{c \to \infty} \beta(c) \sup_{f,g \in \mathcal{G}} \left| E_{f,g} \left[\log \left(G(X_1 + t, X_2 + s) \right. \right.$$

$$\left. \left. \left(1 - G(X_1 + t, X_2 + s) \right) \right) \Big| \, |X_1|, |X_2| > c \right] \right|^{-1} < \infty$$

In principle, we can consider a generalized version of Assumption 6 where we allow different behavior of the distribution tails in the strategic responses of different players. In that case we will need to select the trimming sequences differently for each equation. This will come at a cost of more tedious algebra. However, the conceptual result will be very similar.

We will use the assumption regarding the class of distributions of unobserved heterogeneity components U and V with minimal modifications and we will not reproduce it from

Section 1. We require that the density belongs to a"uniformly manageable" class of functions (as per the definition in Kim and Pollard (1990)). Assumptions 4, 6 and "uniform manageability" characterize the distributions of unobserved heterogeneity that we consider in our analysis. The error distributions commonly used in empirical analysis of games such as normal and logistic satisfy these assumptions.[16]

As in the case of the triangular model, we propose a constructive approach to analyzing the optimal rate for the estimators of the interaction parameters. The idea behind the estimation procedure in the case of triangular system was to use the case where both indicators are equal to zero, which allows one to directly observe the cumulative distribution of errors. This approach will not be immediately available for the complete information game. The outcome probability $\mathcal{P}_{00}(x_1, x_2; \alpha_1, \alpha_2)$ depends on the unknown parameters α_1 and α_2. We modify the estimator by replacing the two-step procedure with an iterative procedure where one can "profile out" the unknown density of errors at each iteration of likelihood maximization with respect to the strategic interaction parameters. Defining the sample log-likelihood

$$\hat{l}^p(\alpha_1, \alpha_2) = \sup_{a_{11}, \dots, a_{KK}} \frac{1}{n} \sum_{i=1}^{n} l\left(\alpha_1, \alpha_2; y_{1i}, y_{2i}, x_{1i}, x_{2i}\right),$$

where K is the number of terms in the orthogonal expansion of the density of U and V, we obtain the estimator as the maximizer of the profile log-likelihood:

$$(\hat{\alpha}_{1n}^*, \hat{\alpha}_{2n}^*) = \text{argmax}_{\alpha_1, \alpha_2} \hat{l}^p(\alpha_1, \alpha_2). \tag{4.2}$$

For an appropriately chosen sequence of cutoff points, this estimator attains the upper rate of convergence for the estimators of the strategic interaction parameters.[17] This property of the estimator based on the trimmed likelihood function is analogous to the result regarding the rate of the semiparametric two-stage estimator that we propose for the triangular system. Provided that we assume identical tail behavior for both error terms, the resulting rates for the interaction parameters are the same. As previously discussed, if the tails of the error distributions are different, the optimal rate result can established by choosing different trimming sequences for the support of covariates X_1 and X_2.

We can use the arguments employed in the case of the triangular system to show that that sequence c_n of trimming points that assures that estimator $(\hat{\alpha}_{1n}^*, \hat{\alpha}_{2n}^*)$ attains the upper rate also delivers the optimal rate in the class of all regular estimators for the strategic interaction parameters. The following theorem formally expresses the structure of the optimal convergence rate and echoes the argument of Theorem 2.2.

---

[16]We give concrete distribution examples in Appendix C.

[17] We provide the formal discussion of this estimator in Appendix B.1.1 and in Appendix A.7 we prove that this estimator attains the upper convergence rate.

Theorem 4.3 *Consider the model of the game of complete information in which the error distribution satisfies Assumptions 6 and 9. Suppose that $c_n \to \infty$ is a sequence such that $\frac{n\beta^2(c_n)}{\nu(c_n)} = O(1)$ with $n/\nu(c_n) \to \infty$. Then for this sequence $\sqrt{\frac{n}{\nu(c_n)}}$ is the optimal rate for the estimator for strategic interaction parameters α_1 and α_2.*

One of the important takeaways from this result is that the optimal rate for estimating strategic interaction parameters is, generally speaking, sub-parametric and depends on the tail behavior of the error terms even in cases with a fixed equilibrium selection mechanism.

5 Static game of incomplete information

5.1 Information in the game of incomplete information

Our triangular model with treatment uncertainty can be considered a special case of a static game of incomplete information. Theoretical results demonstrate that introduction of payoff perturbations leads to a reduction in the number of equilibria.[18] Here we attain regular identification for the interaction parameter as well, but our argument is not one of equilibrium refinement; as with the complete information game, we assume the simplest equilibrium selection rule, but in contrast, we now are able to attain positive information for the interaction parameter.

In this case we interpret the realizations of binary variables Y_1 and Y_2 as actions of player 1 and player 2. Each player is characterized by the deterministic payoff (corresponding to linear indices x_1 and x_2), interaction parameter, unobserved heterogeneity terms u and v, and the payoff perturbations η_1 and η_2. The payoff of player 1 from action $y_1 = 1$ can be represented as $y_1^* = x_1 + \alpha_1 y_2 - u - \sigma\eta_1$, while the payoff from action $y_1 = 0$ is normalized to 0. We impose the following informational assumptions.

Assumption 7 *Suppose that η_1 and η_2 are privately observed by the two players, where $\eta_1 \perp \eta_2$ and both satisfy Assumption 3.*

This model is a generalization of the incomplete information model usually considered in empirical applications because we allow for the presence of unobserved heterogeneity components u and v. This is an empirically relevant assumption if one considers the case where the same two players participate in repeated realizations of the static game. If initially

---

[18]Multiplicity of equilibria can still be an important issue in games of incomplete information as noted in Sweeting (2009) and de Paula and Tang (2011). Alternative approaches to estimation of games of incomplete information with multiple equilibria have been proposed in Lewbel and Tang (2011) and Sweeting (2009).

the unobserved utility components of players are correlated, then after sufficiently many replications of the game the players can learn about the structure of the component of the payoff shock that is correlated with their shock. The remaining elements that cannot be learned from replications of the game are the noise components η_1 and η_2, whose distributions are normalized. An alternative interpretation for this information structure is the that payoff components u and v are *a priori* known to the players but not to the econonometrician. The interaction of the players is considered in the experimental settings where the payoff noise (η_1, η_2) is introduced artificially by the experiment designer. For this reason its distribution is known both to the players and to the econometrician.

Assumption 7 lays the groundwork for the coherent characterization of the structure of equilibrium in this game of incomplete information. First, the strategy of player i is a mapping from the observable variables into actions: $(x_1, x_2, u, v, \eta_i) \mapsto \{0, 1\}$. Second, player i forms the beliefs regarding the action of the rival. Provided that η_1 and η_2 are independent, the beliefs will be functions only of u, v and linear indices. Thus, if $P_i(x_1, x_2, u, v)$ are players' beliefs regarding actions of opponent players, then the strategy, for instance, of player 1 can be characterized as a random variable

$$
\begin{aligned}
Y_1 &= \mathbf{1}\{E\left[Y_1^* \mid X_1, X_2, U, V, \eta_1\right] > 0\} \\
&= \mathbf{1}\{X_1 - U + \alpha_1 P_2(X_1, X_2, U, V) - \sigma\eta_1 > 0\}.
\end{aligned}
\tag{5.1}
$$

Similarly, the strategy of player 2 can be written as

$$
Y_2 = \mathbf{1}\{X_2 - V + \alpha_2 P_1(X_1, X_2, U, V) - \sigma\eta_2 > 0\}.
\tag{5.2}
$$

We note the resemblence of equations (5.1) and (5.2) with the first equation of the triangular system with treatment uncertainty.

To characterize the Bayes-Nash equilibrium in the simultaneous move game of incomplete information we consider a pair of strategies defined by (5.1) and (5.2). Moreover, the beliefs of players have to be consistent with their action probabilities conditional on the information set of the rival. Taking into consideration the independence of player types η and the fact that their cdf is known, we can characterize the pair of equilibrium beliefs as a solution to the system of nonlinear equations:

$$
\begin{aligned}
\sigma\Phi^{-1}(P_1) &= x_1 - u + \alpha_1 P_2 \\
\sigma\Phi^{-1}(P_2) &= x_2 - v + \alpha_2 P_1.
\end{aligned}
\tag{5.3}
$$

Our informational assumption regarding the independence of the unobserved heterogeneity components U and V from payoff perturbations η_1 and η_2 is to define the game with a coherent equilibrium structure. If we allow correlation between the payoff-relevant unobservable variables of two players, then their actions should reflect such correlation and the equilibrium

beliefs should also be functions of the noise components. This structure would not support an elegant form of the equilibrium correspondence (5.3). On the other hand, given that the unobserved heterogeneity components U and V are correlated, the econometrician will observe the individual actions to be correlated. In other words, we consider the structure of the game where actions of players are correlated without having to analyze a complicated equilibrium structure due to correlated unobserved player types.

The system of equations (5.3) can have multiple solutions.[19] To resolve the uncertainty over equilibria and maintain symmetry with our discussion of games of complete information, we assume that uncertainty over multiple possible equilibrium beliefs is resolved by independent coin flips. We formalize this idea in the following assumption.

Assumption 8 *If for some point $(x_1 - u, x_2 - v)$ the system of equations (5.3) has multiple solutions, then the uncertainty regarding the realization of an equilibrium is resolved via a uniform distribution over those solutions.*

We note that the incomplete information model that we constructed embeds the complete information model in the previous section. When σ approaches 0, the payoffs in the incomplete information model are identical to those in the complete information model and are observable by both players. We illustrate the transition from the complete to the incomplete information environment in Figure 2. When $\sigma = 0$, the actions of the players will be determined by U and V only. Figure 2.a. shows four regions, one for each possible pair of actions in the complete information model. There is a region in the middle where multiple pairs of actions are optimal, leading to multiple equilibria. With the introduction of uncertainty, we can only plot the probabilistic picture of players' actions (integrating over the payoff noise η_1 and η_2). We can then characterize the areas where specific action pairs are chosen with probability exceeding a given quantile $1 - q$. A decrease in the variance of payoff noise leads to the convergence of quantiles to the areas in the illustration of the complete information game in Figure 2.a.

---

[19]Sweeting (2009) considers a 2×2 game of incomplete information and gives examples of multiple equilibria in that game. Bajari, Hong, Krainer, and Nekipelov (2010a) develop a class of alorithms for efficient computation of all equilibria in incomplete information games with logistically distributed noise components.
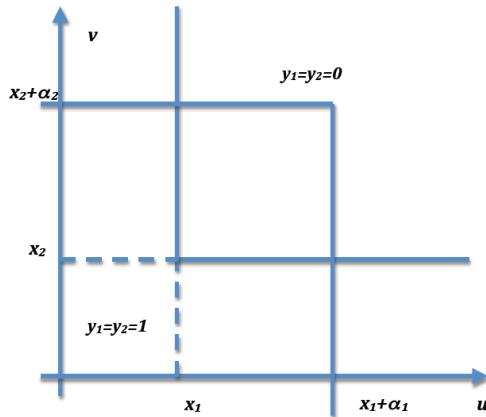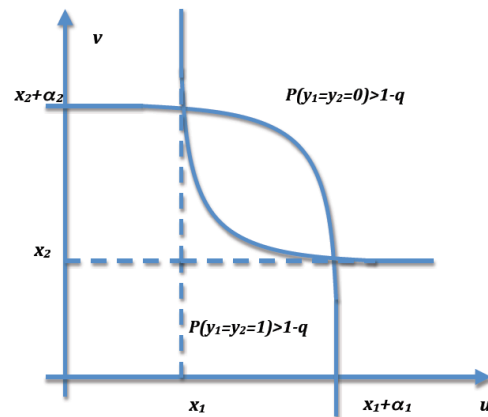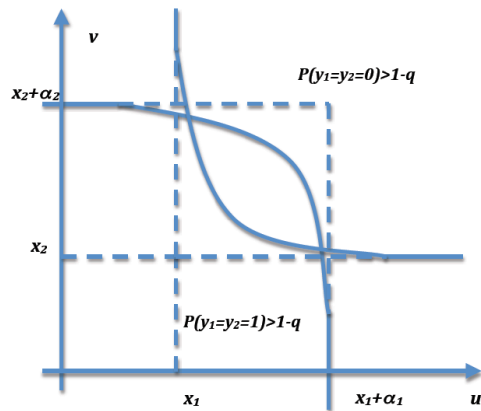
Figure 2.a

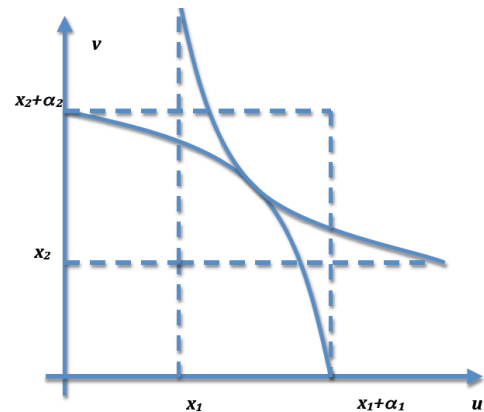**Figure 2.b**

Figure 2.c

Figure 2.d

First, we establish the fact that the strategic interaction parameters α_1 and α_2 and the distribution of errors (U, V) are identified in the considered model. Note that x_1, x_2, u and v enter the system of equations (5.1) and (5.2) in a way, such that the equilibrium beliefs are functions of $x_1 - u$ and $x_2 - v$. Conditional on the realizations x_1, x_2, u, and v, the choices of the two players are also independent. On the other hand, given that the realizations of u and v are not observable to the econometrician, conditional on x_1 and x_2, the choice are correlated. The observed actions are binary and the distribution of the covariates is directly observed in the data (due to independence of the errors (η_1, η_2) and the unobserved heterogeneity (U, V) from the covariates). Thus, the information that the data contains regarding the model is fully summarized by the conditional expectations $E[Y_1|x_1, x_2]$, $E[Y_2|x_1, x_2]$ and $E[Y_1 Y_2|x_1, x_2]$. The identification argument will then have two parts. First, one needs to solve system (5.3) to obtain mappings $P_1(x_1 - u, x_2 - v)$ and $P_2(x_1 - u, x_2 - v)$. Second, one can relate these mappings to the observable probabilities of

actions. Althought, with continuous distribution of the noise η_1 and η_2 the considered model has an equilibrium, the system of equilibrium choice probabilities can have multiple solutions. We approach cases of multiple equilibria by resolving the uncertainty via coin flips. Given our procedure for equilibrium selection, we can associate the observed equilibrium choice probability with the average value of the mappings P_1 and P_2 over the set of possible values for each given $x_1 - u$ and $x_2 - v$. Provided that the system of identifying equations is linear in the choice probabilities, in case of multiple equilibria the equilibrium choice probability has to be replaced by a mixture of possible equilibrium choice probabilities. We denote the "average" choice probabilities $\bar{P}_1$ and $\bar{P}_2$. Then, for instance, the conditional expectation $E[Y_1 Y_2 | x_1, x_2]$ can be expressed as

$$ E[Y_1 Y_2 | x_1, x_2] = \int \bar{P}_1(x_1 - u, x_2 - v)\bar{P}_2(x_1 - u, x_2 - v)g(u,v)\,du\,dv $$

Given strategic interaction parameters, the average probabilities $\bar{P}_1$ and $\bar{P}_2$ are known. We can use this expression to identify the distribution of unobserved heterogeneity for each value of the pair of srategic interaction parameters. Using the expectations $E[Y_1 | x_1, x_2]$ and $E[Y_2 | x_1, x_2]$, we can then identify the coefficients α_1 and α_2. In the following theorem we summarize our identification result.

Theorem 5.1 *Suppose that Assumptions 4, 7, and 8 are satisfied. Then the strategic interaction terms α_1 and α_2 in the model defined by (5.1) and (5.2) are identified.*

Given that parameters of interest are identified (along with the unobserved distribution of error terms), we can proceed with establishing the result regarding the information of the incomplete information game. We find that for any finite variance of noise σ^2 (which can be arbitrarily small) the information in the model of the incomplete information game is not zero. We also provide a result characterizing the Fisher information for the strategic interaction parameters as the variance of players' privately observed payoff shocks approaches zero. As in the incomplete information triangular model, the Fisher information of those parameters approaches zero.

Theorem 5.2 *Suppose that Assumptions 4, 7, and 8 are satisfied.*

(i) For any $\sigma > 0$ the information corresponding to parameters (α_1, α_2) in the incomplete information game defined by (5.1) and (5.2) is strictly positive.

(ii) As $\sigma \to 0$ the information corresponding to parameters (α_1, α_2) in the incomplete information game defined by (5.1) and (5.2) approaches zero.

As in the case of the triangular model, this result also suggests an alternative estimator for the strategic interaction parameters in the complete information game: we can use the estimates of the strategic interaction parameters from the incomplete information game with a small variance of the noise to approximate the strategic interaction parameters in the incomplete information game.

5.2   Convergence rate in the incomplete information game

We conclude the analysis with the following theorem which combines the result of Theorem 5.2 and Theorem IV.1.1 in Ibragimov and Has'minskii (1981). This theorem states that *the optimal convergence rate for the estimator for the strategic interaction parameters in the incomplete information game is parametric and the minimum variance of the estimator converging at the parametric rate corresponds to the semiparametric efficiency bound.*

Theorem 5.3 *Under Assumptions 4, 7, and 8 for any sub-convex loss function $w(\cdot)$ and standard normal random variable ξ:*

$$\liminf_{n\to\infty} \inf_{\hat{\alpha}_n=(\hat{\alpha}_1^n,\hat{\alpha}_2^n)'} \sup_{f,g\in\mathcal{G}} E_{f,g}\left[w\left(\sqrt{n}(\hat{\alpha}_n-\alpha(f,g))\right)\right] \geq E[w\left(\Omega^{1/2}\xi\right)],$$

where $g(\cdot)$ is the distribution of errors U and V, $f(\cdot)$ is the distribution of covariates, $\hat{\alpha}_n$ is the estimator for the strategic interaction parameters, $\alpha(f,g)$ is the true value of the strategic interaction parameters and Ω is the semiparametric efficiency bound.[20]

The result of this theorem is not surprising in light of our finding in Theorem 5.2: given that the information for the strategic interaction parameters is positive, the semiparametric efficiency bound which is equal to the inverse information matrix will be finite. An important additional result provided in Appendix F.2 is the explicit derivation of the semiparametric efficiency bound. This result demonstrates the structure of the variance of the efficient estimator for the strategic interaction parameter in the static game model with a non-parametric distribution of unobserved heterogeneity. The efficiency bound for a static two-payer game of incomplete information has been analyzed in Aradillas-Lopez (2010) without allowing for player-specific unobserved heterogeneity that is commonly observed by the players. Grieco (2010) allows for the individual-specific heterogeneity, but assumes a specific parametric form for both the payoff noise distribution and the distribution of unobserved heterogeneity. We provide the result that parametric inference remains feasible even when the distribution of unobserved heterogeneity remains fully nonparametric. Our efficiency result provides a

---

[20]We provide an explicit result for the semiparametric efficiency bound in Appendix F.2.

semiparametric efficiency bound for the generalized class of static games of incomplete information in Bajari, Hong, Krainer, and Nekipelov (2010b) as well as in Haile, Hortaçsu, and Kosenok (2008) for the games with quantal response equilibria considered in Palfrey (1985), provided that we allow for the presence of unobserved heterogeneity that is correlated across players and with unknown distribution.

6 Conclusions

This paper considers identification and inference in simultaneous equation models with discrete endogenous variables. We analyze triangular systems where the parameter of interest is the coefficient of a discrete endogenous variable, which is related to the treatment effect in certain settings. We also study nontriangular systems, focusing on simultaneous discrete games, where we are interested in the strategic interaction parameters. We then consider an incomplete information setting in which there is an additive random payoff disturbance which is only privately observed by the players. Our main findings are that the complete information models have zero Fisher information under our conditions, whereas the incomplete information models can have positive information. Our findings have important implications for both the triangular and nontriangular systems. In the triangular case, both the zero information and the optimal convergence rates we obtain indicate little advantage to estimating the parameter in this model relative to estimating the model proposed in Lewbel (1998). In the nontriangular case, zero Fisher information result implies that the difficulty in identification of the strategic interaction parameters is not due to incoherency (i.e. the presence of multiple equilibria or non-existence of the pure strategy equilibria), as we obtain this result even after introducing an equilibrium selection rule. In the incomplete information models (both triangular and nontriangular) the support of the endogenous variable is convexified by the additional payoff uncertainty which leads to the positive Fisher information.

The work here suggests many areas for future research. In the incomplete information models, with positive information, it would be useful to consider more general equilibrium selection rules and still attain positive information. Furthermore, we restricted our attention to static games, and it would be useful to explore information levels in both complete information and incomplete information in dynamic games. We leave these topics for future research.

References

ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings," *Econometrica*, 70(1), 91–117.

ABREVAYA, J., J. HAUSMAN, AND S. KHAN (2011): "Testing for Causal Effects in a Generalized Regression Model with Endogenous Regressors," *Econometrica*, 78, 2043–2061.

AHN, H., AND C. MANSKI (1993): "Distribution theory for the analysis of binary choice under uncertainty with nonparametric estimation of expectations," *Journal of Econometrics*, 56(3), 291–321.

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795–1844.

ANDREWS, D., S. BERRY, AND P. JIA (2004): "Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location," *manuscript, Yale University*.

ANGRIST, J., AND G. IMBENS (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity," *Journal of the American Statistical Association*, 90(430), 431–442.

ARADILLAS-LOPEZ, A. (2010): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information," *Journal of Econometrics*, 157, 409–431.

BAJARI, P., J. HAHN, H. HONG, AND G. RIDDER (2010): "A note on semiparametric estimation of finite mixtures of discrete choice models with application to game theoretic models," *International Economic Review*.

BAJARI, P., H. HONG, J. KRAINER, AND D. NEKIPELOV (2010a): "Computing equilibria in static games of incomplete information using the all-solution homotopy," *Working paper*.

BAJARI, P., H. HONG, M. KRAINER, AND D. NEKIPELOV (2010b): "Estimating Static Models of Strategic Interactions," *Journal of Business Economics and Statistics*, Forthcoming.

BAJARI, P., H. HONG, AND S. RYAN (2010): "Identification and Estimation of Discrete Games of Complete Information," *Econometrica*, 78, 1529–1578.

BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2011): "Sharp identification regions in games," *Econometrica, Forthcoming*.

BERRY, S., AND E. TAMER (2006): "Identification in Models of Oligopoly Entry," in *Advances in economics and econometrics: theory and applications, ninth World Congress*, vol. 2, p. 46. Cambridge University Press.

BJORN, P., AND Q. VUONG (1985): "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation," Caltech Working Paper 537.

BLUNDELL, R., AND J. POWELL (2004): "Endogeneity in Binary Response Models," *Review of Economic Studies*, 73.

BRESNAHAN, T., AND P. REISS (1990): "Entry in Monopoly Market," *The Review of Economic Studies*, 57(4), 531.

——— (1991a): "Empirical models of discrete games," *Journal of Econometrics*, 48(1-2), 57–81.

——— (1991b): "Entry and competition in concentrated markets," *Journal of Political Economy*, pp. 977–1009.

CARNEIRO, P., J. HECKMAN, AND E. VYTLACIL (2010): "Evaluating marginal policy changes and the average effect of treatment for individuals at the margin," *Econometrica*, 78(1), 377–394.

CHAMBERLAIN, G. (1986): "Asymptotic efficiency in semi-parametric models with censoring," *journal of Econometrics*, 32(2), 189–218.

CHASSANG, S., AND E. SNOWBERG (2010): "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments," Discussion paper, National Bureau of Economic Research.

CHASSANG, S., AND S. TAKAHASHI (2011): "Robustness to incomplete information in repeated games," *Theoretical Economics*, 6(1), 49–93.

CHEN, S., AND S. KHAN (1999): "Estimating Censored Regression Models in the Presence of Nonparametric Multiplicative Heteroscedasticity," forthcoming, Journal of Econometrics.

CHESHER, A. (2003): "Identification in nonseparable models," *Econometrica*, 71(5), 1405–1441.

——— (2007): "Endogeneity and discrete outcomes," *Working paper*.

——— (2010): "Instrumental variable models for discrete outcomes," *Econometrica*, 78(2), 575–601.

CILIBERTO, F., AND E. TAMER (2009): "Market structure and multiple equilibria in airline markets," *Econometrica*, 77(6), 1791–1828.

COSSLETT, S. (1987): "Efficiency Bounds for Distribution-Free Estimators of the Binary Choice and the Censored Regression Models," *Econometrics*, 55(3), 559–585.

DE PAULA, A., AND X. TANG (2011): "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information," *Econometrica*, Forthcoming.

DEKEL, E., AND D. FUDENBERG (1990): "Rational behavior with payoff uncertainty," *Journal of Economic Theory*, 52(2), 243–267.

DUNFORD, N., AND J. SCHWARTZ (1965): "Linear operators. Part II. Spectral theory," *Bull. Amer. Math. Soc.*, 2(9904), 11348–9.

FLORENS, J., J. HECKMAN, C. MEGHIR, AND E. VYTLACIL (2008): "Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects," *Econometrica*, 76(5), 1191–1206.

FUDENBERG, D., D. KREPS, AND D. LEVINE (1988): "On the robustness of equilibrium refinements," *Journal of Economic Theory*, 44(2), 354–380.

GRAHAM, B., AND J. POWELL (2009): "Identification and Estimation of'Irregular'Correlated Random Coefficient Models," Working Paper.

GRIECO, P. (2010): "Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets," Working Paper.

HAILE, P., A. HORTAÇSU, AND G. KOSENOK (2008): "On the empirical content of quantal response equilibrium," *The American Economic Review*, 98(1), 180–200.

HECKMAN, J., AND S. NAVARRO (2004): "Using matching, instrumental variables, and control functions to estimate economic choice models," *Review of Economics and statistics*, 86(1), 30–57.

HIRANO, K., AND G. IMBENS (2004): "The propensity score with continuous treatments," *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, pp. 73–84.

HOROWITZ, J. (1993): "Optimal Rates of Convergence of Parameter Estimators in the Binary Response Model with Weak Distributional Assumptions," *Econometric Theory*, 9, 1–18.

HU, Y., AND S. SCHENNACH (2008): "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 76(1), 195–216.

IBRAGIMOV, I., AND R. HAS'MINSKII (1978): "On the Capacity in Communication by Smooth Signals," *Soviet Math Dokl.*, 19, 1043–1047.

———— (1981): *Statistical Estimation Asymptotic Theory*. Springer-Verlag.

IMBENS, G. (2009): "Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009)," Discussion paper, National Bureau of Economic Research.

IMBENS, G., AND W. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77(5), 1481–1512.

IMBENS, G., AND J. WOOLDRIDGE (2009): "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47(1), 5–86.

JUN, S., J. PINKSE, AND H. XU (2011): "Tighter Bounds in Triangular Systems," *Journal of Econometrics*, forthcoming.

Kajii, A., and S. Morris (1997): "The robustness of equilibria to incomplete information," *Econometrica: Journal of the Econometric Society*, pp. 1283–1309.

Khan, S., and D. Nekipelov (2010): "Information Bounds and Impossibility Theorems for Simultaneous Discrete Response Models," *Working paper.*

Khan, S., and E. Tamer (2010): "Irregular Identification, Support Conditions and Inverse Weight Estimation," *Econometrica*, forthcoming.

Kim, J., and D. Pollard (1990): "Cube Root Asymptotics," *Annals of Statistics*, 18, 191–219.

Klein, R., C. Shan, and F. Vella (2011): "Semi-Parametric Estimation of Sample Selection Models with Binary Selection Rules and Binary Outcomes," Georgetown University working paper.

Komarova, T., O. Linton, and S. Srisuma (2011): "Estimation of Private Value Distributions in Second-Price Auctions and Generalized Competing Risks Models," *LSE Working paper.*

Koroselev, A., and A. Tsybakov (1993): *Minimax Theory of Image Reconstruction.* Springer-Verlag.

Lewbel, A. (1998): "Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors," *Econometrica*, 66(1), 105–122.

Lewbel, A., and X. Tang (2011): "Identification and Estimation of Discrete Bayesian Games with Multiple Equilibria Using Excluded Regressors," Working paper, Boston College.

Manski, C. (1991): "Nonparametric estimation of expectations in the analysis of discrete choice under uncertainty," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, p. 259. Cambridge University Press.

——— (1993): "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 60(3), 531.

——— (2000): "Identification problems and decisions under ambiguity: Empirical analysis of treatment response and normative analysis of treatment choice," *Journal of Econometrics*, 95(2), 415–442.

Menzel, K., and P. Morganti (2009): "Large Sample Properties for Estimators Based on the Order Statistics Approach in Auctions," NYUWorking Paper.

Newey, W. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5(2), 99–135.

Newey, W. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168.

PAKES, A., M. OSTROVSKY, AND S. BERRY (2007): "Simple Estimators for the Parameters of Dynamic Discrete Games (with Entry/Exit Examples)," *RAND Journal of Economics.*

PALFREY, T. (1985): "Uncertainty Resolution, Private Information Aggregation and the Counot Competitive Limit," *Review of Economic Studies*, 52, 69—83.

POLLARD, D. (1984): *Convergence of Stochastic Processes.* Springer Verlag.

SCHENNACH, S. (2004): "Estimation of nonlinear models with measurement error," *Econometrica*, 72(1), 33–75.

SHAIKH, A., AND E. VYTLACIL (2011): "Partial Identification in Triangular Systems of Equations With Binary Dependent Variables," *Econometrica*, 79, 949–955.

STONE, C. (1982): "Optimal Global Rates of Convergence for Nonparametric Regression," *Annals of Statistics*, 10, 1040–1053.

SWEETING, A. (2009): "The Strategic Timing of Radio Commercials: An Empirical Analysis Using Multiple Equilibria," *RAND Journal of Economics*, 40, 710–742.

TAMER, E. (2003): "Incomplete Bivariate Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147–167.

TIROLE, J. (1988): *The Theory of Industrial Organization.* MIT Press.

VLADIMIROV, V. (1971): *Equations of mathematical physics.* Marcel Dekker.

VYTLACIL, E. J., AND N. YILDIZ (2007): "Dummy Endogenous Variables in Weakly Separable Models," *Econometrica*, 75, 757–779.

ZINDE-WALSH, V. (2009): "Errors-in-variables models: a generalized functions approach," *Arxiv preprint arXiv:0909.5390.*

Appendix

A Proofs

A.1 Proof of Theorem 2.1

To simplify our argument, we assume that coefficients β_0 and δ_0 are known. We will thus refer to the indices in each equation as x_1 and x, respectively. To derive the information of the model we follow the approach in Chamberlain (1986) by demonstrating that for each triangular model generated by a distribution satisfying the conditions of Theorem 2.1 we can construct a parametric submodel passing through that model for which the information for parameter α is equal to zero. Suppose that Γ contains all distributions of errors that satisfy the conditions of Theorem 2.1 along with all distributions of indices $x_1 = \beta_0 \, z_1$ and $x = \delta_0 \, z$ for which $E[q(Z)^2] < \infty$ for $q(\cdot)$ defined in the statement of the theorem such that x_1 and x have a continuous joint distribution with a full support on $\mathbb{R}^2$. We first construct the likelihood function of the model and introduce the following notation:

$$P^{11}(t_1, t) = \Pr\left(U \leq t_1,\, V \leq t\right) = G(t_1, t),$$
$$P^{01}(t_1, t) = \Pr\left(U > t_1,\, V \leq t\right),$$
$$P^{10}(t_1, t) = \Pr\left(U \leq t_1,\, V > t\right),$$
$$P^{00}(t_1, t) = \Pr\left(U > t_1,\, V > t\right).$$

The likelihood function is determined by the density

$$r(y_1, y_2, x_1, x; \alpha, P) = P^{11}(x_1 + \alpha, x)^{y_1 y_2} P^{01}(x_1 + \alpha, x)^{(1-y_1)y_2}$$
$$\times P^{10}(x_1, x)^{y_1(1-y_2)} P^{00}(x_1, x)^{(1-y_1)(1-y_2)}$$

with respect to the measure μ defined on $\Omega = \{0,1\}^2 \times \mathbb{R}^2$ such that for any Borel set A in $\mathbb{R}^2$, $\mu(\{1,1\} \times A) = \mu(\{1,0\} \times A) = \mu(\{0,1\} \times A) = \mu(\{0,0\} \times A) = \nu(A)$, where $P((X_1, X) \in A) = \int_A d\nu$. Let $h : \mathbb{R}^2 \mapsto \mathbb{R}$ be a continuously differentiable function supported on the compact set with its derivative being continuous in the interior of that compact set such that $\frac{\partial h(u,v)}{\partial u} \geq B$ for some constant B on that compact set. We define $\tilde{\Lambda}$ as the collection of paths through the original model which we design as

$$\lambda^{11}(t_1, t; \delta) = P^{11}(t_1 + \delta(h(t_1, t) + 1), t),$$
$$\lambda^{01}(t_1, t; \delta) = P^{01}(t_1 + \delta(h(t_1, t) + 1), t),$$
$$\lambda^{10}(t_1, t; \delta) = P^{10}(t_1, t),$$
$$\lambda^{00}(t_1, t; \delta) = P^{00}(t_1, t),$$

where we note that these paths maintain the properties of the joint probability distribution (bounded between 0 and 1, sum up to 1) and, in a sufficiently small neighborhood about the origin containing

δ, they also maintain the monotonicity of the cdf (as the partial derivative of $h(\cdot, \cdot)$ is bounded from below).

We denote the likelihood function corresponding to the perturbed model $l_\lambda(y_1, y_2, x_1, x; \alpha, \delta)$. Provided the assumed dominance condition, it will be mean-square differentiable at $(\alpha_0, 0)$. In other words, we can find functions $\psi_\alpha(x_1, x)$ and $\psi_\delta(x_1, x)$ such that

$$l_\lambda^{1/2}(\cdot; \alpha, \delta) = \psi_\alpha(x_1, x)(\alpha - \alpha_0) + \psi_\lambda(x_1, x)\delta + R_{\alpha, \delta},$$

with

$$E\left[R_{\alpha, \delta}^2\right] / \left(|\alpha - \alpha_0| + |\delta|\right)^2 \to 0 \text{ as } \alpha \to \alpha_0, \delta \to 0.$$

We can explicitly derive the mean-square derivatives. In particular, the derivative with respect to the finite-dimensional parameter can be expressed as

$$\psi_\alpha(x_1, x) = \frac{1}{2}\{y_1 y_2 P^{11}(x_1 + \alpha_0, x)^{-1/2} - (1 - y_1)y_2 P^{01}(x_1 + \alpha_0, x)^{-1/2}\}\frac{\partial G(x_1 + \alpha_0, x)}{\partial u},$$

and the derivative with respect to λ can be expressed as

$$\psi_\lambda(x_1, x) = \frac{1}{2}\{y_1 y_2 P^{11}(x_1 + \alpha_0, x)^{-1/2} - (1 - y_1)y_2 P^{01}(x_1 + \alpha_0, x)^{-1/2}\}$$
$$\times \frac{\partial G(x_1 + \alpha_0, x)}{\partial u}(h(x_1 + \alpha_0, x) + 1).$$

We the can use the fact that the Fisher information can be bounded as

$$I_{\lambda, \delta} \leq 4 \int (\psi_\alpha - \psi_\lambda)^2 \, d\mu$$

$$= \int \frac{G_v(x)}{G(x_1 + \alpha_0, x)\left(G_v(x) - G(x_1 + \alpha_0, x)\right)} \left(\frac{\partial G(x_1 + \alpha_0, x)}{\partial u}\right)^2 h^2(x_1 + \alpha_0, x) \, d\nu(x_1, x)$$

We can define the measure on Borel sets in $\mathbb{R}^2$ as

$$\pi(A) = \int_A \frac{G_v(x)}{G(x_1, x)\left(G_v(x) - G(x_1, x)\right)} \left(\frac{\partial G(x_1, x)}{\partial u}\right)^2 d\nu(x_1 - \alpha_0, x),$$

allowing us to characterize

$$I_{\lambda, \delta} \leq 4\|h\|_{L_2(\pi)}^2$$

Chamberlain (1986) demonstrates that the space of differentiable functions with compact support is dense in $L_2(\pi)$. Moreover, we require the derivative of h to be continuous in the interior of its support. Let S be the support of h. We take $\epsilon^* > 0$ and construct the set S_{ϵ^*} to be a compact subset of S such that the Euclidean distance of the boundary of S from the boundary of S_{ϵ^*} is at least ϵ^*, where ϵ^* is selected such that $\pi(S \setminus S_{\epsilon^*}) < \sqrt{\epsilon}$. Since the set of differentiable functions is dense in $L_2(\pi)$, for any $\epsilon > 0$ we can find $a \in C_c^2(\mathbb{R}^2)$ such that $\|a\|_{L_2(\pi)} < \sqrt{\epsilon}$. The derivative $\frac{\partial a(u,v)}{\partial u}$ is continuous in the interior of S. Provided that $S_{\epsilon^*} \subset S$, this derivative is continuous on the entire set S_{ϵ^*} and, due to its compactness it is uniformly continuous there. As a result, there

exists $M = \sup\limits_{S_{\epsilon*}} \left| \frac{\partial a(u,v)}{\partial u} \right|$. There also exists $M' = \sup\limits_{S} |a|$. Then we pick the direction h^* as function with support on S such that $h^* = \frac{B}{2}(a/M)$ in $S_{\epsilon*}$. Then we note that

$$\|h^*\|_{L_2(\pi)} \leq \frac{B}{2M}\|a\|_{L_2(\pi)} + \frac{BM'}{2M}\|\mathbf{1}_{S\setminus S_{\epsilon*}}\|_{L_2(\pi)} < \frac{B(M'+1)}{2M}\sqrt{\epsilon}.$$

As a result, $I_{\lambda,\delta} \leq \frac{B^2(M'+1)^2}{M^2}\epsilon$. As the choice of ϵ was arbitrary, this proves that $\inf\limits_{\lambda\in\tilde{\Lambda}} I_{\lambda,\alpha} = 0$.
Q.E.D.

A.2   Convergence rate of the two-step estimator

We start with the formal definition of the *uniformly manageable* class of densities.

Assumption 9   *(i) For the class of densities $\mathcal{G}$ satisfying Assumptions 4 and 6, there exists a Hilbert space $\mathbf{H}$ with the basis of normalized Hermite polynomials $\{h_l\}_0^\infty$ such that*

(a) *For any sufficiently large $K \in \mathbb{N}$ and $\mathbf{H}_K = \{h_l\}_{l=0}^K \sup\limits_{g\in\mathcal{H}} \inf\limits_{\mu\in\mathbb{R}^K} \left\| g - \sum_{l=0}^K \mu^l h_l \right\|_{L_2} = O(K^{-r})$, for $r > 0$*

(b) $|h_l(\cdot)| \leq C$ *and* $\int^{(\cdot)} |h_l(z)|^2 \, dz \leq C$

(c) *For each l the class of functions $\mathcal{F}_{h,l} = \{h_l(\cdot + t),\, t \in \mathbb{R}\}$ is polynomial, i.e. the covering number $\sup\limits_{\mathcal{Q}} N(\epsilon, \mathcal{F}_{h,l}, L_2(\mathcal{Q})) < A\,\epsilon^{-\gamma}$, for some $\gamma > 0$ and probability measures $\mathcal{Q}$.*

(ii) *Consider functions $f(\cdot, t, s) = \int^{\cdot+t} \int^{\cdot+s} g(u,v)\, du\, dv$. For each $K \in \mathbb{N}$ the class of projections of these functions on each basis vector $\mathcal{F}_k = \{proj(f(\cdot, t, s), h_k),\, g \in \mathcal{G}, |t|, |s| < \infty\}$ has envelope F_k such that $E\left[F_k^2\right] < \infty$ and it has at most exponential covering number, i.e. there exist constants A' and γ' such that :*

$$\sup\limits_{\mathcal{Q}} \log N(\epsilon\|F\|, \mathcal{F}_k, L_2(\mathcal{Q})) < A'\,\epsilon^{-\gamma'}$$

Next, we prove the following lemma.

Lemma A.1 *Suppose that the choice probability functions are estimated via an orthogonal sequence $\mathcal{H}^{(K)}(\cdot, \cdot) = (\mathcal{H}_k(x_1, x))_{k=0}^K$ and $\inf\limits_{\mu\in\mathbb{R}^K} \|P(y_1, y_2|x_1, x) - \mu'\mathcal{H}^{(K)}(x_1, x)\| = O(K^{-r})$. We assume that $K \to \infty$ as $n \to \infty$ such that $n/(K\log n) \to \infty$. The estimator is then constructed by defining the likelihood with support restricted to the set $\{|x_1|, |x| \leq c_n\}$. Suppose that a sequence c_n is selected such that $\nu(c_n)/n \to 0$, $K^r/\nu(c_n) \to 0$, and $\nu(c_n)K^2/n \to \infty$. Then for any sequence $\hat{\alpha}_n$ with the function $\hat{l}(\alpha)$ corresponding to the maximand of (2.4) such that $\hat{l}_{K,c_n}(\hat{\alpha}_{0,n}) \geq \sup\limits_{\alpha} \hat{l}_{K,c_n}(\alpha) - o_p\left(\sqrt{\frac{\nu(c_n)}{n}}\right)$ we have*

$$\sqrt{\frac{n}{\nu(c_n)}}|\hat{\alpha}_{0,n}^* - \alpha_0| = O_p(1).$$

Proof: We introduce the "uncensored" objective function

$$l\left(\alpha; y_1, y_2, x_1, x\right) = y_1 y_2 \log \hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x) + (1 - y_1)y_2 \log \hat{\mathcal{P}}_n^{01}(x_1 + \alpha, x),$$

with $Q(\alpha) = E\left[l\left(\alpha; y_1, y_2, x_1, x\right)\right]$, and $\hat{\mathcal{P}}_n^{11}$ defined in Appendix B. Denote $\hat{l}\left(\alpha\right) = \frac{1}{n}\sum_{i=1}^{n} l\left(\alpha; y_{1i}, y_{2i}, x_{1i}, x_i\right)$. Also denote

$$\ell\left(\alpha; y_1, y_2, x_1, x\right) = y_1 y_2 \omega_n(x_1 + \alpha)\omega_n(x) \log \mathcal{P}^{11}(x_1 + \alpha, x)$$
$$+ (1 - y_1)y_2 \omega_n(x_1 + \alpha)\omega_n(x) \log \mathcal{P}^{01}(x_1 + \alpha, x),$$

and $\hat{\ell}\left(\alpha\right) = \frac{1}{n}\sum_{i=1}^{n} \ell\left(\alpha; y_{1i}, y_{2i}, x_{1i}, x_i\right)$. Now consider the following decomposition of the objective function:

$$\hat{l}(\alpha) - \hat{\ell}(\alpha_0) = R_1 + R_2 + R_3 + R_4 + R_5 + R_6,$$

where

$$R_1 = \hat{l}(\alpha) - \hat{\ell}(\alpha) - E\left[\hat{l}(\alpha)\right] + E\left[\hat{\ell}(\alpha)\right],$$
$$R_2 = \hat{\ell}(\alpha) - \hat{\ell}(\alpha_0) - E\left[\hat{\ell}(\alpha)\right] + E\left[\hat{\ell}(\alpha_0)\right],$$
$$R_3 = E\left[\hat{l}(\alpha)\right] - E\left[\hat{\ell}(\alpha)\right], \quad R_4 = E\left[\hat{\ell}(\alpha)\right] - Q(\alpha)$$
$$R_5 = -E\left[\hat{\ell}(\alpha_0)\right] + Q(\alpha_0), \quad R_6 = Q(\alpha) - Q(\alpha_0).$$

Term R_1

For convenience, we introduce new notation denoting

$$p^{Kk}(z) = \omega_n(x_1)\omega_n(x)\left[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x_1)\right]\left[\mathcal{H}_{l_2}(c_n) - \mathcal{H}_{l_2}(x)\right]$$

and introduce vectors $p^K(z) = \left(p^{K1}(z), \ldots, p^{KK}(z)\right)'$. Also let $d_i^{00} = (1 - y_{1i})(1 - y_{2i})$ and $d^{00} = \left(d_1^{00}, \ldots, d_n^{00}\right)'$. Let $\Delta(z) = E[d^{00}|z]$ and $\Delta = (\Delta(z_1), \ldots, \Delta(z_n))'$. We can project this function of z on K basis vectors of the sieve space. Let β be the vector of coefficients of this projection. As demonstrated in Newey (1997), for $P = \left(p^K(z_1), \ldots, p^K(z_n)\right)'$ and $\hat{Q} = P'P/n$

$$\|\hat{Q} - Q\| = O_p\left(\sqrt{\frac{K}{n}}\right), \quad \text{where his } \zeta_0\left(K\right) = C,$$

and Q is non-singular by assumption with the smallest eigenvalue bounded from below by some constant $\underline{\lambda} > 0$. Hence the smallest eigenvalue of $\hat{Q}$ will converge to $\underline{\lambda} > 0$. Following Newey (1997) we use the indicator 1_n to indicate the cases where the smallest eigenvalue of $\hat{Q}$ is above $\frac{\underline{\lambda}}{2}$ to avoid singularities. We also introduce

$$m^{Kk}(z) = \omega_n(x_1)\omega_n(x)\left[\mathcal{H}_{l_1}(x_1) - \mathcal{H}_{l_1}(-c_n)\right]\left[\mathcal{H}_{l_2}(x) - \mathcal{H}_{l_2}(-c_n)\right].$$

We then can write the estimator

$$\hat{\mathcal{P}}^{11}(x_1, x) = m^K(z)'\hat{Q}^{-1}P\,d^{00}/n$$

Note that

$$m^{K\prime}(z)\left(\hat{\beta}-\beta\right)=m^{K\prime}(z)\left(\hat{Q}^{-1}P'\left(d^{00}-\Delta\right)/n+\hat{Q}^{-1}P'\left(\Delta-P\beta\right)/n\right). \tag{A.1}$$

We can evaluate the component in the second term as

$$\|P\left(\Delta-P\beta\right)/n\|=\sqrt{\sum_{k=1}^{K}\left(\frac{1}{n}\sum_{i=1}^{n}p^{Kk}(z_i)\left(\Delta(z_i)-p^K(z_i)'\beta\right)\right)^2}$$

$$\leq\sqrt{K\,C\,K^{-2r}}=O(K^{\frac{1}{2}-r})$$

provided our assumption regarding the sieve space (Assumption 9 (iii) (a)). As we demonstrate, this result allows us to concentrate on the first term ignoring the second one. For the first term in (A.1), we can use the result that smallest eigenvalue of $\hat{Q}$ is converging to $\underline{\lambda}>0$. Then application of the Cauchy-Schwartz inequality leads to

$$\left|m^{K\prime}(z)\hat{Q}^{-1}P'\left(d^{00}-\Delta\right)\right|\leq\left\|\hat{Q}^{-1}m^K(z)\right\|\left\|P'\left(d^{00}-\Delta\right)\right\|.$$

Then $\left\|\hat{Q}^{-1}m^K(z)\right\|\leq\frac{C}{\underline{\lambda}}\sqrt{K}$, and

$$\left\|P'\left(d^{00}-\Delta\right)\right\|=\sqrt{\sum_{k=1}^{K}\left(\sum_{i=1}^{n}p^{Kk}(z_i)\left(d_i^{00}-\Delta(z_i)\right)\right)^2}$$

$$\leq\sqrt{K}\max_k\left|\sum_{i=1}^{n}p^{Kk}(z_i)\left(d_i^{00}-\Delta(z_i)\right)\right|$$

Thus,

$$\left|m^{K\prime}(z)\hat{Q}^{-1}P'\left(d^{00}-\Delta\right)/n\right|\leq\frac{CK}{\underline{\lambda}}\max_k\left|\frac{1}{n}\sum_{i=1}^{n}p^{Kk}(z_i)\left(d_i^{00}-\Delta(z_i)\right)\right|.$$

Denote $\mu_n=\mu\frac{n^{\delta/2}}{\sqrt{nK}}=\gamma_n/K$ for any $\delta\in(0,1]$. Next we adapt the arguments for proving Theorem 37 in Pollard (1984) to provide the bound for $P\left(\sup_z\frac{1}{n}\|m^{K\prime}(z)\hat{Q}^{-1}P'\left(d^{00}-\Delta\right)\|>K\mu_n\right)$. For K non-negative random variables Y_i we note that

$$P\left(\max_i Y_i>Kc\right)\leq\sum_{i=1}^{K}P\left(Y_i>c\right).$$

Using this observation, we can find that

$$P\left(\sup_z\frac{1}{n}\|m^{K\prime}(z)\hat{Q}^{-1}P'\left(d^{00}-\Delta\right)\|>K\mu_n\right)\leq\sum_{k=1}^{K}P\left(\left\|\frac{1}{n}\sum_{i=1}^{n}p^{Kk}(z_i)\left(d_i^{00}-\Delta(z_i)\right)\right\|>\gamma_n\right),$$

where we used our definition of $\gamma_n=K\mu_n$. This inequality allows us to substitute the tail bound for the class of functions $\mathcal{P}_n^{11}(\cdot,\cdot)$ by a tail bound for fixed functions

$$\mathcal{P}_{n,k}=\{p^{Kk}(\cdot)\left(d^{00}-\Delta(\cdot)\right)\}.$$

Then we can apply the inequality from Theorem 37 in (Pollard 1984) to obtain

$$P\left(\frac{1}{n}\left\|\sum_{i=1}^{n} p^{Kk}(z_i)\left(d_i^{00} - \Delta(z_i)\right)\right\| > \gamma_n\right) \leq 2\exp\left(-\frac{2n\gamma_n^2}{C^2} + A'\gamma_n^{-\gamma'}\right).$$

As a result, we find that

$$P\left(\sup_z \frac{1}{n}\|m^{K\prime}(z)\hat{Q}^{-1}P'\left(d^{00} - \Delta\right)\| > K\mu_n\right) \leq 2K\exp\left(-\frac{2n\gamma_n^2}{C^2} + A'\gamma_n^{-\gamma'}\right).$$

Then, provided that $n/\log K \to \infty$ and $\gamma' < 1$ we prove that the right-hand side of this inequality converges to zero. This means that

$$\sup_{(x_1,x)\in\mathcal{X}}\|\hat{\mathcal{P}}^{11}(x_1,x) - \mathrm{proj}\left(\mathcal{P}^{11}(x_1,x)\,|H_K\right)\| = o_p\left(n^{\frac{\gamma}{2}-\frac{1}{2}}\right).$$

From the second term we provide the evaluation

$$\sup_{\mathcal{P}^{11}\in H}\sup_{x_1,x}\|\mathrm{proj}\left(\mathcal{P}^{11}(x_1,x)\,|H_K\right) - \mathcal{P}^{11}(x_1,x)\| = O(K^{-r})$$

Therefore, if $K^r/n^{(1-\delta)/2} \to \infty$, then the "bias" term will be negligible. Next, we note that similar evaluations can be provided for $\mathcal{P}^{01}$. As the density of (U,V) is strictly positive on $\mathbb{R}^2$, the probabilities are bounded away from zero on any bounded subset of $\mathbb{R}^2$ and we can make the same evaluations for $\log \mathcal{P}^{11}(\cdot)$ and $\log \mathcal{P}^{01}(\cdot)$. As a result, we can deliver the rate

$$\sup_{\alpha}|\hat{l}(\alpha) - \hat{\ell}(\alpha) - E\left[\hat{l}(\alpha)\right] + E\left[\hat{\ell}(\alpha)\right]| = o_p\left(n^{-(1-\delta)/2}\right).$$

Term R_3

Consider the approximation bias term. Note that we can express

$$E\left[\hat{l}(\alpha)\right] = E\Big[\omega_n(x_1+\alpha)\omega_n(x)\Big(\mathcal{P}^{11}(x_1+\alpha,x)\log\hat{\mathcal{P}}_n^{11}(x_1+\alpha,x)$$
$$+ \mathcal{P}^{01}(x_1+\alpha,x)\log\hat{\mathcal{P}}_n^{01}(x_1+\alpha,x)\Big)\Big].$$

Similarly, we can express

$$E\left[\hat{\ell}(\alpha)\right] = E\Big[\omega_n(x_1+\alpha)\omega_n(x)\Big(\mathcal{P}^{11}(x_1+\alpha,x)\log\mathcal{P}^{11}(x_1+\alpha,x)$$
$$+ \mathcal{P}^{01}(x_1+\alpha,x)\log\mathcal{P}^{01}(x_1+\alpha,x)\Big)\Big].$$

One can attain a uniform rate

$$\sup_{x_1,x}\left\|\hat{\mathcal{P}}_n^{11}(x_1+\alpha,x) - \mathcal{P}^{11}(x_1+\alpha,x)\right\| = O_p\left(\sqrt{\frac{K}{n}} + K^{-r}\right),$$

given the quality of approximation by selected sieves. We can then evaluate the entire term

$$|R_3| = O\left(\sqrt{\frac{K}{n}} + K^{1-r}\right).$$

Terms R_4 and R_5

Consider term R_4. We can evaluate this term as

$$\left|E\left[\hat{\ell}(\alpha)\right] - Q(\alpha)\right| \leq 4\left|\int_{-\infty}^{c_n}\int_{-\infty}^{c_n} \mathcal{P}^{11}(x_1+\alpha,x)\log\mathcal{P}^{11}(x_1+\alpha,x)f(x_1,x)\,dx_1dx\right|.$$

We can then apply the Cauchy-Schwartz inequality and continue evaluation as

$$\left|E\left[\hat{\ell}(\alpha)\right] - Q(\alpha)\right| \leq 4E\left[y_1y_2\right]\left|\int_{-\infty}^{c_n}\int_{-\infty}^{c_n}\log\mathcal{P}^{11}(x_1+\alpha,x)f(x_1,x)\,dx_1\right|$$

$$\leq C\,\beta(c_n).$$

from Assumption 2.

Term R_2

We use the following assumption regarding the population likelihood function.

Assumption 10 *The population likelihood function $Q(\cdot)$ is twice continuously differentiable and uniquely maximized at α_0 with a negative definite Hessian.*

Consider the class of functions indexed by $\alpha \in A$ such that given

$$\ell(\alpha,y_1,y_2,x_1,x) = \left[y_1y_2\log\mathcal{P}^{11}(x_1+\alpha,x) + (1-y_1)y_2\log\mathcal{P}^{01}(x_1+\alpha,x)\right]\omega_n(x_1+\alpha)\omega_n(x)$$

$$\mathcal{F}_{n,\delta} = \{f = \ell(\alpha,\cdot) - \ell(\alpha_0,\cdot),\ \ |\alpha-\alpha_0| \leq \delta\}$$

Provided that the density of errors is twice differentiable in mean square with bounded mean square derivatives, there exist bounded functions $\dot{\mathcal{P}}^{11}$ and $\dot{\mathcal{P}}^{01}$ such that functions in class $\mathcal{F}_{n,\delta}$ have envelope

$$F_{n,\delta} = \mathbf{1}\{|x_1+\alpha_0| \leq c_n+\delta\}\omega_n(x)$$
$$\times\left[\frac{y_1y_2\dot{\mathcal{P}}^{11}}{\mathcal{P}^{11}} + \frac{(1-y_1)y_2\dot{\mathcal{P}}^{01}}{\mathcal{P}^{01}}\right]\delta.$$

Then, by Assumption 2, we can evaluate

$$\left(E\left[F_{n,\delta}^2\right]\right)^{1/2} = O\left(\nu(c_n)^{1/2}\delta\right).$$

Consider the re-parametrization of the model $\alpha = \alpha_0 + \frac{h}{r_n}$ for a sequence $r_n \to \infty$. Take $h \in [0,\eta r_n]$ for some large η and split the interval $[0,\eta r_n]$ into "shells" $S_{n,j} = \{h : 2^{j-1} < |h| < 2^j\}$. Suppose that $\hat{h}$ is the maximizer for $\hat{l}(\alpha_0 + \frac{h}{r_n})$. Then if $|\hat{h}| > 2^M$ for some M then $\hat{h}$ belongs to $S_{n,j}$ with $j \geq M$. As a result

$$P\left(|\hat{h}| > 2^M\right) \leq \sum_{j\geq M, 2^j < \eta r_n} P\left(\sup_{h\in S_{n,j}}\left(\hat{l}(\alpha_0 + \frac{h}{r_n}) - \hat{l}(\alpha_0)\right) \geq 0\right).$$

We now use the results from the evaluation of the terms R_1 and $R_3 to R_5$, taking into consideration that

$$Q(\alpha) - Q(\alpha_0) \leq -H|\alpha - \alpha_0|^2,$$

for some $H > 0$ due to the differentiability of $Q(\cdot)$ and the restriction on its Hessian at α_0 in Assumption 10. We can evaluate

$$P\left(\sup_{h \in S_{n,j}} \left(\hat{l}(\alpha_0 + \frac{h}{r_n}) - \hat{l}(\alpha_0) \right) \geq 0 \right)$$

$$\leq P\left(\sup_{h \in S_{n,j}} |R_2| \geq |R_1| + |R_3| + |R_4| + |R_5| + |R_6| \right)$$

$$= P\left(\sup_{h \in S_{n,j}} |R_2| \geq \frac{2^{2j-2}}{r_n^2} + O\left(\sqrt{\frac{K}{n}} + K^{1-r} + \beta(c_n)^{-1} \right) \right),$$

where we use that the difference of absolute values is smaller than the absolute value of the difference. Then we use the Markov inequality to obtain that

$$P\left(\sup_{h \in S_{n,j}} \left(\hat{l}(\alpha_0 + \frac{h}{r_n}) - \hat{l}(\alpha_0) \right) \geq 0 \right)$$

$$\leq \frac{E\left[\sup_{h \in S_{n,j}} \left| \hat{\ell}(\alpha_0 + \frac{h}{r_n}) - \hat{\ell}(\alpha_0) - E\left[\hat{\ell}(\alpha_0 + \frac{h}{r_n}) \right] + E\left[\hat{\ell}(\alpha_0) \right] \right| \right]}{\frac{2^{2j-2}}{r_n^2} + O\left(\sqrt{\frac{K}{n}} + K^{1-r} + \beta(c_n)^{-1} \right)}$$

Using the empirical process notation, we define the covering integral as

$$J(\delta, F) = \sup_Q \int_0^\delta \sqrt{1 + \log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q))}\, d\epsilon,$$

where Q is the probability measure, $\mathcal{F}$ is a class of functions with the envelope F, and $N(\cdot)$ is the covering number of the consider class. Provided the finiteness of the covering integral of the class $\mathcal{F}_{n,\delta}$, we can use the maximum inequality to evaluate

$$E\left[\sup_{h \in S_{n,j}} \sqrt{n} \left| \hat{\ell}(\alpha_0 + \frac{h}{r_n}) - \hat{\ell}(\alpha_0) - E\left[\hat{\ell}(\alpha_0 + \frac{h}{r_n}) \right] + E\left[\hat{\ell}(\alpha_0) \right] \right| \right]$$

$$\leq J(1, F_{n,h/r_n}) E\left[F_{n,h/r_n}^2 \right]^{1/2} = O\left(\nu(c_n)^{1/2} \frac{2^j}{r_n} \right).$$

Assuming that $r_n \beta(c_n)^{-1} = o(1)$, $r_n \sqrt{K/n} = o(1)$ and $r_n K^{-(d+1)/2} \to 0$, then

$$P\left(\sup_{h \in S_{n,j}} \left(\hat{l}(\alpha_0 + \frac{h}{r_n}) - \hat{l}(\alpha_0) \right) \geq 0 \right) \leq O\left(2^{-j+2} r_n \sqrt{\frac{\nu(c_n)}{n}} \right).$$

This implies that

$$P\left(|\hat{h}| > 2^M \right) \leq O\left(2^{-M+3} r_n \sqrt{\frac{\nu(c_n)}{n}} \right)$$

The right-hand side converges to zero for $M \to \infty$ if $r_n = \sqrt{\frac{n}{\nu(c_n)}}$.

Q.E.D.

A.3 Proof of Theorem 2.2

First, consider the following evaluation from the proof of Lemma A.1

$$P\left(\sup_{h\in S_{n,j}}\left(\hat{l}(\alpha_0+\frac{h}{r_n})-\hat{l}(\alpha_0)\right)\geq 0\right)$$

$$\leq \frac{E\left[\sup_{h\in S_{n,j}}\left|\hat{\ell}(\alpha_0+\frac{h}{r_n})-\hat{\ell}(\alpha_0)-E\left[\hat{\ell}(\alpha_0+\frac{h}{r_n})\right]+E\left[\hat{\ell}(\alpha_0)\right]\right|\right]}{\frac{2^{2j-2}}{r_n^2}+O\left(\sqrt{\frac{K}{n}}+K^{-(d+1)/2}+\beta(c_n)^{-1}\right)}$$

Using the maximum inequality as before we can conclude that the ratio can be evaluated as

$$P\left(\sup_{h\in S_{n,j}}\left(\hat{l}(\alpha_0+\frac{h}{r_n})-\hat{l}(\alpha_0)\right)\geq 0\right)\leq O\left(2^{-j+1}\frac{\nu(c_n)^{1/2}r_n}{\sqrt{n}}\right)$$

We note that evaluation here is different, because, unlike in Lemma A.1, here we allow $r_n\beta(c_n)=O(1)$. This allows us to obtain

$$P\left(\sqrt{\frac{n}{\nu(c_n)}}|\hat{h}|>2^M\right)\leq O\left(2^{-M+2}r_n\sqrt{\frac{\nu(c_n)}{n}}\right).$$

Thus, if $L=2^M$,

$$P\left(\sqrt{\frac{n}{\nu(c_n)}}|\hat{h}|>L\right)\leq O\left(\frac{4}{L}r_n\sqrt{\frac{\nu(c_n)}{n}}\right).$$

Provided that we choose $r_n\sqrt{\frac{\nu(c_n)}{n}}=1$, we assure that for the maximal risk

$$\lim_{L\to\infty}\limsup_{n\to\infty}R\left(\alpha_0+\frac{\hat{h}}{r_n},r_n,L\right)=0.$$

This means that r_n is the upper rate.

To derive the lower convergence rate we use the result from Koroselev and Tsybakov (1993). Denote the likelihood ratio $\Lambda(\mathcal{P}_1,\mathcal{P}_2)=\frac{dP_{\mathcal{P}_1}^n}{dP_{\mathcal{P}_2}^n}$. Then the following lemma is the result given in Koroselev and Tsybakov (1993).

Lemma A.2 *Suppose that $\alpha_0^1=\alpha(\mathcal{P}_1)$ and $\alpha_0^2=\alpha(\mathcal{P}_2)$, and let $\lambda>0$ be such that*

$$P_{\mathcal{P}_2}\left(\Lambda(\mathcal{P}_1,\mathcal{P}_2)>\exp\left(-\lambda\right)\right)\geq p>0,$$

and $|\alpha_0^1-\alpha_0^2|\geq 2s_n$. Then for any estimator $\hat{\alpha}_{0,n}$ we have $\max_{\mathcal{P}_1,\mathcal{P}_2}P\left(|\hat{\alpha}_{0,n}-\alpha_0|>s_n\right)\geq p\exp(-\lambda/2)$.

We can now use this lemma to derive the following result regarding the lower rate for the estimator of interest.

The log-likelihood function of the model is

$$n\hat{\mathcal{L}}(\alpha) = n\hat{\ell}(\alpha) + n\hat{e}(\alpha)$$

with

$$\hat{e}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\{y_{1i}\log\mathcal{P}^{11}(x_{1i}+\alpha, x_i) + (1-y_{1i})\log\mathcal{P}^{01}(x_{1i}+\alpha, x_i)\}$$

$$y_{2i}\mathbf{1}\{|x_{1i}| > c_n, |x_i| > c_n\}$$

Note that we use the same distribution of covariates x_1 and x. For $c_n \to \infty$, pick[21]

$$\mathcal{P}_2(\cdot, \cdot) = \mathcal{P}(\cdot, \cdot), \text{ and } \mathcal{P}_1(\cdot, \cdot) = \mathcal{P}(\cdot, \cdot)\omega_n(\cdot)\omega_n(\cdot).$$

Following from our previous analysis for such choices of $\mathcal{P}_1(\cdot)$ and $\mathcal{P}_2(\cdot)$, the corresponding likelihood maximizers satisfy

$$|\alpha_1 - \alpha_2| = O(\beta(c_n)).$$

We can then express

$$\begin{aligned}
\Lambda(\mathcal{P}_1, \mathcal{P}_2) &= \exp\left(n\hat{\mathcal{L}}_1(\alpha_1) - n\hat{\mathcal{L}}_2(\alpha_2)\right) \\
&= \exp\left(n\hat{\ell}(\alpha_1) - n\hat{\ell}(\alpha_2) - n\hat{e}(\alpha_2)\right) \\
&= \exp\left(n\left[\hat{\ell}(\alpha_1) - \hat{\ell}(\alpha_2) - \ell(\alpha_1) + \ell(\alpha_2)\right] - n\hat{e}(\alpha_2) - n\left(\ell(\alpha_2) - \ell(\alpha_1)\right)\right)
\end{aligned}$$

We note that $\hat{\ell}(\alpha_1) - \hat{\ell}(\alpha_2) - \ell(\alpha_1) + \ell(\alpha_2) = o_p(1)$ and $\hat{e}(\alpha_2) = o_p(1)$. As a result, the last term will dominate as $n \to \infty$. Then $\log\Lambda(\mathcal{P}_1, \mathcal{P}_2)$ is bounded from below as n approaches infinity if and only if $n(\ell(\alpha_2) - \ell(\alpha_1))$ is bounded. We note that α_1 maximizes $\ell(\alpha)$. This means that

$$\ell(\alpha_2) - \ell(\alpha_1) = -\frac{1}{2}H(c_n)(\alpha_2 - \alpha_1)^2 + o(|\alpha_2 - \alpha_1|).$$

Invoking the Cauchy-Schwartz inequality, we can evaluate $H(c_n) = O(\nu(c_n)^{-1})$. As a result, we find that

$$n[\ell(\alpha_2) - \ell(\alpha_1)] = O\left(\frac{n\beta(c_n)^2}{\nu(c_n)}\right).$$

This means that $\frac{n\beta(c_n)^2}{\nu(c_n)} = O(1)$, suggesting that for large n there exists a lower bound on the likelihood ratio. By invoking Lemma A.2, we obtain the desired result.
Q.E.D.

---

[21] The selected $\mathcal{P}_1$ may not be a probability measure; however it bears the properties of the measure, such that characteristics of the measure such as a Radon-Nykodim derivative are still well-defined.

A.4 Proof of Theorem 3.1

Our model is generated by two binary variables, Y_1 and Y_2. As a result, its parametric components will be fully characterized by conditional probabilities $E[Y_1|x_1, x]$, $E[Y_2|x_1, x]$ and $E[Y_1 Y_2|x_1, x]$. In Appendix E we derive the Fourier transformation of each of these probabilities. Consider (E.6) and (E.8). First, using the limit result (E.4), we can transform (E.6) to

$$\mathcal{Q}(0, t_2) = \left[\alpha \frac{\chi_\Phi(\sigma t_2)}{it_2} + 2\pi^2 \delta(t_2) \right] \chi_v(t_2),$$

where $\mathcal{Q}(\cdot, \cdot)$ is the Fourier transform of $E[Y_1|x_1, x]$ and we use the result that $\chi_{uv}(0, t_2) = \chi_v(t_2)$. Multiplying both sides by $it_2/\chi_\Phi(\sigma t_2)$ and performing deconvolution provides us with

$$\alpha g_v(v) = \frac{1}{2\pi} \int e^{it_2 v} \frac{it_2 \mathcal{Q}(0, t_2)}{\chi_\Phi(\sigma t_2)} \, dt_2.$$

Similarly, we note that the Fourier transform of $E[Y_2|x_1, x]$ can be expressed as

$$\mathcal{F}(t_2) = \left[\frac{\chi_\Phi(\sigma t_2)}{it_2} + 2\pi^2 \delta(t_2) \right] \chi_v(t_2).$$

Multiplying both sides by $it_2/\chi_\Phi(\sigma t_2)$ and performing deconvolution leads to

$$g_v(v) = \frac{1}{2\pi} \int e^{it_2 v} \frac{it_2 \mathcal{F}(t_2)}{\chi_\Phi(\sigma t_2)} \, dt_2.$$

By our assumption the density of the distribution of unobserved heterogeneity is strictly above zero on $\mathbb{R}^2$. Evaluating the ratio of the above expressions, we can find

$$\alpha = \left(\int e^{itv} \frac{it\mathcal{F}(t)}{\chi_\Phi(\sigma t)} \, dt \right)^{-1} \int e^{itv} \frac{it\mathcal{Q}(0, t)}{\chi_\Phi(\sigma t)} \, dt. \tag{A.2}$$

Therefore, parameter α is identified.
Q.E.D.

A.5 Proof of Theorem 3.2

A.5.1 Proof for part (i)

In the proof of Theorem 3.1 we presented an explicit expression for the parameter of interest as expressed by (A.2). To compute the information corresponding to the parmeter of interest, we construct the log-likelihood of the model by explicitly expressing the probabilities:

$$P_{11}(x_1, x; \alpha, g) = \int \mathbf{1}\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\} \Phi\left(\frac{x-v}{\sigma}\right) g(u, v) \, du \, dv,$$

$$P(x_1, x; \alpha, g) = \int \Phi\left(\frac{x-v}{\sigma}\right) g_v(v) \, dv,$$

$$Q(x_1, x; \alpha, g) = \int \mathbf{1}\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\} g(u, v) \, du \, dv.$$

We can then express all probabilities of interest as

$$P_{01}(x_1, x; \alpha, g) = P(x_1, x; \alpha, g) - P_{11}(x_1, x; \alpha, g),$$

$$P_{10}(x_1, x; \alpha, g) = Q(x_1, x; \alpha, g) - P_{11}(x_1, x; \alpha, g),$$

$$P_{00}(x_1, x; \alpha, g) = 1 - Q(x_1, x; \alpha, g) - P(x_1, x; \alpha, g) + P_{11}(x_1, x; \alpha, g),$$

and the derivatives of the probabilities of interest as

$$\frac{\partial P_{11}(x_1, x; \alpha, g)}{\partial \alpha} = \int \Phi\left(\frac{x-v}{\sigma}\right) g\left(x_1 + \alpha \Phi\left(\frac{x-v}{\sigma}\right), v\right) dv \equiv D_1(x_1, x; \alpha, g),$$

$$\frac{\partial Q(x_1, x; \alpha, g)}{\partial \alpha} = \int g\left(x_1 + \alpha \Phi\left(\frac{x-v}{\sigma}\right), v\right) dv \equiv D_2(x_1, x; \alpha, g).$$

We adopt the notation of the proof of zero information in the complete information model. We consider the square root of the density generating the model:

$$r(y_1, y_2, x_1, x; \alpha, g)^{1/2} = y_1 y_2 P_{11}(x_1, x; \alpha, g)^{1/2} + y_1(1 - y_1)P_{10}(x_1, x; \alpha, g)^{1/2}$$

$$+ (1 - y_1)y_2 P_{01}(x_1, x; \alpha, g)^{1/2} + (1 - y_1)(1 - y_1)P_{00}(x_1, x; \alpha, g)^{1/2}.$$

We can express the mean square derivative with respect to α as

$$\psi_\alpha(y_1, y_2, x_1, x) = \frac{1}{2}\left[y_1 y_2 P_{11}(x_1, x; \alpha, g)^{-1/2} - (1 - y_1)y_2 P_{01}(x_1, x; \alpha, g)^{-1/2}\right] D_1(x_1, x; \alpha, g)$$

$$+ \frac{1}{2}\left[(1 - y_1)(1 - y_1)P_{00}(x_1, x; \alpha, g)^{1/2} - y_1(1 - y_1)P_{10}(x_1, x; \alpha, g)^{1/2}\right]$$

$$\times (D_1(x_1, x; \alpha, g) - D_2(x_1, x; \alpha, g)).$$

Thus, we can express the information for parameter α as

$$I_\alpha = 4 \int (\psi_{\alpha_0})^2 \, d\mu.$$

If ν is the measure on $\mathbb{R}^2$ corresponding to the distribution of x_1 and x, following the approach in the derivation of information of the complete information model, we define the measures on Borel subsets of $\mathbb{R}^2$

$$\pi_1(A) = \int_A \frac{P_1(x_1, x; \alpha_0, g)}{P_{11}(x_1, x; \alpha_0, g)\left(P_1(x_1, x; \alpha_0, g) - P_{11}(x_1, x; \alpha_0, g)\right)} d\nu(x_1, x)$$

and

$$\pi_2(A) = \int_A \frac{1 - P_1(x_1, x; \alpha_0, g)}{P_{00}(x_1, x; \alpha_0, g)\left(1 - P_1(x_1, x; \alpha_0, g) - P_{00}(x_1, x; \alpha_0, g)\right)} d\nu(x_1, x).$$

We can then express the information of the model as

$$I_\alpha = \|D_1(x_1, x; \alpha_0, g)\|^2_{L_2(\pi_1)} + \|D_1(x_1, x; \alpha_0, g) - D_2(x_1, x; \alpha_0, g)\|^2_{L_2(\pi_2)} \tag{A.3}$$

We construct the measure π^* which "picks out" for each Borel subset A one of the measures π_1 or π_2 which gives this set less weight: $\pi^*(A) = \min\{\pi_1(A), \pi_2(A)\}$. Based on this structure of the measure, we can write:

$$I_\alpha \geq \|D_1(x_1, x; \alpha_0, g)\|^2_{L_2(\pi^*)} + \|D_2(x_1, x; \alpha_0, g) - D_1(x_1, x; \alpha_0, g)\|^2_{L_2(\pi^*)}$$

Denoting $w(t) = \Phi(t/\sigma)$ and $t = x - v$, we express

$$D_1(x_1, x; \alpha_0, g) = \int w(t) g(x_1 + \alpha_0 w(t), x - t) \, dt$$

and

$$D_2(x_1, x; \alpha_0, g) - D_1(x_1, x; \alpha_0, g) = \int (1 - w(t)) g(x_1 + \alpha_0 w(t), x - t) \, dt.$$

Suppose that $S \subset \mathbb{R}^2$ is a compact set such that $\pi^*(S) > C$. Then given that $g(\cdot)$ is continuous and strictly positive, there exists $M(t) = \inf_{(x_1, x) \in S} |g(x_1 + \alpha w(t), x - t)|$ which is not equal to zero at least for some $t \in \mathbb{R}$. We take $\sqrt{\epsilon} = \sup_{t \in [-B, B]} |M(t)|$, where B is selected such that $[-B, B]$ contains at least one point where $M(t) \neq 0$. Suppose that the supremum is attained at point t^*. By continuity, there exists some neighborhood of t^* where $M(t) > \sqrt{\epsilon/2}$. Denote the size of this neighborhood R. Invoking triangle inequality and bounds provided above results in

$$I_\alpha \geq \|D_2(x_1, x; \alpha_0, g)\|^2_{L_2(\pi^*)} \geq \|D_2(x_1, x; \alpha_0, g) \mathbf{1}_S\|^2_{L_2(\pi^*)}$$
$$\geq C \left\| \int_{\mathbb{R}} M(t) \, dt \right\|^2 \geq C \left\| \int_{-B}^{B} M(t) \, dt \right\|^2 \geq \frac{1}{2} C R^2 \epsilon > 0.$$

Therefore, the information corresponding to parameter α is strictly positive.
Q.E.D.

A.5.2   Proof for part (ii)

Consider the expression for the information in the incomplete information triangular model expressed in (A.3):

$$I_\alpha = \|D_1(x_1, x; \alpha_0, g)\|^2_{L_2(\pi_1)} + \|D_1(x_1, x; \alpha_0, g) - D_2(x_1, x; \alpha_0, g)\|^2_{L_2(\pi_2)}$$

We construct the measure π^{**} (it may be not a probability measure) that is constructed as an integral over: $\frac{d\pi^{**}}{d\nu} = \max\{\frac{d\pi_1}{d\nu}, \frac{d\pi_1}{d\nu}\}$, where the maximum is considered in the pointwise sense over all regular points of measures π^1 and π^2 and where $\frac{d\pi_1}{d\nu}$ is the Radon-Nykodim density with respect to the σ-finite measure ν. Then we note that $\pi^{**}(\mathbb{R}^2) < \Pi < \infty$, assuming that both measures are defined on the entire $\mathbb{R}^2$. We denote $w(t) = \Phi(t/\sigma)$ and $t = x - v$ and express

$$D_1(x_1, x; \alpha_0, g) = \int w(t) g(x_1 + \alpha_0 w(t), x - t) \, dt \leq \int \max_{w \in [0,1]} g(x_1 + \alpha_0 w, x - t) \, dt$$

and

$$D_2(x_1, x; \alpha_0, g) - D_1(x_1, x; \alpha_0, g) = \int (1 - w(t)) g(x_1 + \alpha_0 w(t), x - t) \, dt.$$
$$\leq \int \max_{w \in [0,1]} g(x_1 + \alpha_0 w, x - t) \, dt$$

As a result, we find that

$$I_\alpha \leq \|D_1(x_1, x; \alpha_0, g)\|^2_{L_2(\pi^{**})} + \|D_1(x_1, x; \alpha_0, g) - D_2(x_1, x; \alpha_0, g)\|^2_{L_2(\pi^{**})} \to 2 \left\| \max_{w \in [0,1]} g_u(x_1 + \alpha_0 w) \right\|^2_{L_2(\pi^{**})}$$

Provided that the class of twice continuously differentiable functions is dense in $L_2(\pi^{**})$, for each $\epsilon > 0$ we can find an element of this class such that its norm in $L_2(\pi^{**})$ is less than ϵ. This means that the limiting information is smaller than 2ϵ. Since, the choice of ϵ was arbitrary, we conclude that the limiting information is zero.

Q.E.D.

A.6   Proof of Theorem 4.2

To derive the information of the model, we follow the approach in Chamberlain (1986) by demonstrating that for each complete information static game model generated by a distribution satisfying the conditions of Theorem 4.2 we can construct a parametric submodel passing through that model for which the information for parameters α_1 and α_2 is equal to zero.

Suppose that Γ contains all distributions of errors that satisfy the conditions of Theorem 4.2 along with distributions of indices x_1 and x_2. First we construct the likelihood function of the model and introduce the following notation:

$$P^{11}(t_1, t) = \Pr(U \leq t_1, V \leq t) = G(t_1, t),$$
$$P^{01}(t_1, t) = \Pr(U > t_1, V \leq t),$$
$$P^{10}(t_1, t) = \Pr(U \leq t_1, V > t),$$
$$P^{00}(t_1, t) = \Pr(U > t_1, V > t).$$

Without loss of generality, we focus on the case where the signs of coefficients α_1 and α_2 coincide. We construct the probability mass corresponding to the region with multiple equilbria as

$$\Delta(t_1, t_2; \alpha_1, \alpha_2) = \Pr(t_1 < U \leq t_1 + \alpha_1, \, t_2 < V \leq t_2 + \alpha_2)$$

We write the density of the data as

$$r(y_1, y_2, x_1, x_2; \alpha, P) = \left(P^{11}(x_1 + \alpha_1, x_2 + \alpha_2) - \frac{1}{2}\Delta(x_1, x_2; \alpha_1, \alpha_2) \right)^{y_1 y_2}$$
$$\times P^{01}(x_1 + \alpha_1, x_2)^{(1-y_1)y_2} P^{10}(x_1, x_2 + \alpha_2)^{y_1(1-y_2)}$$
$$\times \left(P^{00}(x_1, x) - \frac{1}{2}\Delta(x_1, x_2; \alpha_1, \alpha_2) \right)^{(1-y_1)(1-y_2)}$$

with respect to the measure μ defined on $\Omega = \{0,1\}^2 \times \mathbb{R}^2$ such that for any Borel set A in $\mathbb{R}^2$, $\mu(\{1,1\} \times A) = \mu(\{1,0\} \times A) = \mu(\{0,1\} \times A) = \mu(\{0,0\} \times A) = \nu(A)$, where $P((X_1, X_2) \in A) = \int_A d\nu$.

Let $h_1 : \mathbb{R}^2 \mapsto \mathbb{R}$ and $h_2 : \mathbb{R}^2 \mapsto \mathbb{R}$ be continuously differentiable functions supported on the compact set with continuous derivatives in the interior of that compact set such that $\frac{\partial h_i(u,v)}{\partial u} \geq B$ and $\frac{\partial h_i(u,v)}{\partial v} \geq B$ for some constant B on that compact set and $i = 1, 2$. Define $\tilde{\Lambda}$ as the collection of paths through the original model which we design as

$$\lambda^{11}(t_1, t_2; \delta_1, \delta_2) = P^{11}(t_1 + \delta_1(h_1(t_1, t_2) + 1), t_2 + \delta_2(h_2(t_1, t_2) + 1)),$$

$$\lambda^{01}(t_1, t_2; \delta_1, \delta_2) = P^{01}(t_1 + \delta_1(h_1(t_1, t_2 + \alpha_2) + 1), t_2),$$

$$\lambda^{10}(t_1, t_2; \delta_1, \delta_2) = P^{11}(t_1, t_2 + \delta_2(h_2(t_1 + \alpha_1, t_2) + 1)),$$

$$\lambda^{00}(t_1, t_2; \delta_1, \delta_2) = P^{11}(t_1, t),$$

$$\gamma(t_1, t_2; \alpha_1, \alpha_2, \delta_1, \delta_2) = \Pr\Big(t_1 < U \leq t_1 + \alpha_1 + \delta_1(h_1(t_1 + \alpha_1, t_2 + \alpha_2) + 1),$$

$$t_2 < V \leq t_2 + \alpha_2 + \delta_2(h_2(t_1 + \alpha_1, t_2 + \alpha_2) + 1)\Big)$$

where we note that these paths maintain the properties of the joint probability distribution (bounded between 0 and 1, sum up to 1) and, in a sufficiently small neighborhood about the origin containing δ, they also maintain the monotonicity of the cdf (as the partial derivatives of $h_1(\cdot, \cdot)$ and $h_2(\cdot, \cdot)$ are bounded from below).

Denote the likelihood function corresponding to the perturbed model $l_\lambda(y_1, y_2, x_1, x_2; \alpha, \delta)$. Provided the assumed dominance condition, it will be mean-square differentiable at $(\alpha_0, 0)$. In other words, we can find vector functions $\psi_\alpha(x_1, x_2)$ and $\psi_\delta(x_1, x_2)$ such that

$$l_\lambda^{1/2}(\cdot; \alpha, \delta) = \psi_\alpha(x_1, x_2)'(\alpha - \alpha_0) + \psi_\lambda(x_1, x_2)'\delta + R_{\alpha,\delta},$$

with

$$E\left[R_{\alpha,\delta}^2\right] / (|\alpha - \alpha_0| + |\delta|)^2 \to 0 \text{ as } \alpha \to \alpha_0, \, \delta \to 0.$$

We can explicitly derive the mean-square derivatives. For convenience, we introduce notation

$$P^{++}(x_1, x_2; \alpha) = P^{11}(x_1 + \alpha_1, x_2 + \alpha_2) - \frac{1}{2}\Delta(x_1, x_2, \alpha_1, \alpha_2)$$

$$P^{-+}(x_1, x_2; \alpha) = P^{01}(x_1 + \alpha_1, x_2)$$

$$P^{+-}(x_1, x_2; \alpha) = P^{10}(x_1, x_2 + \alpha_2)$$

$$P^{--}(x_1, x_2; \alpha) = P^{00}(x_1, x_2) - \frac{1}{2}\Delta(x_1, x_2, \alpha_1, \alpha_2)$$

In particular, the components of the derivative with respect to the finite-dimensional parameter can be expressed as

$$\psi_{\alpha_1}(x_1, x_2) = \frac{1}{4}\left\{ y_1 y_2 P^{++}(x_1, x_2; \alpha)^{-1/2} - (1 - y_1)(1 - y_2)P^{--}(x_1, x_2; \alpha)^{-1/2} \right\}$$

$$\times \frac{\partial G(x_1 + \alpha_1, x_2 + \alpha_2)}{\partial u}$$

$$- \frac{1}{2}(1 - y_1)y_2 P^{-+}(x_1, x_2; \alpha)^{-1/2}\frac{\partial G(x_1 + \alpha_1, x_2)}{\partial u},$$

and

$$\psi_{\alpha_2}(x_1, x_2) = -\frac{1}{4}\left\{y_1 y_2 P^{++}(x_1, x_2; \alpha)^{-1/2} - (1 - y_1)(1 - y_2)P^{--}(x_1, x_2; \alpha)^{-1/2}\right\}$$
$$\times \frac{\partial G(x_1 + \alpha_1, x_2 + \alpha_2)}{\partial v}$$
$$-\frac{1}{2}y_1(1 - y_2)P^{+-}(x_1, x_2; \alpha)^{-1/2}\frac{\partial G(x_1, x_2 + \alpha_2)}{\partial v}.$$

The derivative with respect to λ can be expressed as

$$\psi_{\lambda,1}(x_1, x_2) = \frac{1}{4}\left\{y_1 y_2 P^{++}(x_1, x_2; \alpha)^{-1/2} - (1 - y_1)(1 - y_2)P^{--}(x_1, x_2; \alpha)^{-1/2}\right\}$$
$$\times \frac{\partial G(x_1 + \alpha_1, x_2 + \alpha_2)}{\partial u}(h_1(x_1 + \alpha_1, x_2 + \alpha_2) + 1)$$
$$-\frac{1}{2}(1 - y_1)y_2 P^{-+}(x_1, x_2; \alpha)^{-1/2}\frac{\partial G(x_1 + \alpha_1, x_2)}{\partial u}(h_1(x_1 + \alpha_1, x_2 + \alpha_2) + 1),$$

and

$$\psi_{\lambda,2}(x_1, x_2) = -\frac{1}{4}\left\{y_1 y_2 P^{++}(x_1, x_2; \alpha)^{-1/2} - (1 - y_1)(1 - y_2)P^{--}(x_1, x_2; \alpha)^{-1/2}\right\}$$
$$\times \frac{\partial G(x_1 + \alpha_1, x_2 + \alpha_2)}{\partial v}(h_2(x_1 + \alpha_1, x_2 + \alpha_2) + 1)$$
$$-\frac{1}{2}y_1(1 - y_2)P^{+-}(x_1, x_2; \alpha)^{-1/2}\frac{\partial G(x_1, x_2 + \alpha_2)}{\partial v}(h_2(x_1 + \alpha_1, x_2 + \alpha_2) + 1).$$

We note that the corresponding score has mean zero.

We cuse the fact that the Fisher information can be bounded as

$$I_{\lambda,\delta_1} \leq 4 \int (\psi_{\alpha_1} - \psi_{\lambda,1})^2 \, d\mu$$
$$= \int \frac{1}{4}\left(\left[P^{++}(x_1, x_2; \alpha_0)^{-1} + P^{--}(x_1, x_2; \alpha_0)^{-1}\right] \left(\frac{\partial G(x_1 + \alpha_1, x_2 + \alpha_2)}{\partial u}\right)^2 \right.$$
$$\left. + P^{-+}(x_1, x_2; \alpha_0)^{-1} \left(\frac{\partial G(x_1 + \alpha_1, x_2)}{\partial u}\right)^2 \right) h_1^2(x_1 + \alpha_1, x_2 + \alpha_2) \, d\nu(x_1, x_2)$$

Define the measure on Borel sets in $\mathbb{R}^2$ as

$$\pi_1(A) = \int_A \frac{1}{4}\left(\left[P^{++}(x_1, x_2; \alpha_0)^{-1} + P^{--}(x_1, x_2; \alpha_0)^{-1}\right] \left(\frac{\partial G(x_1 + \alpha_1, x_2 + \alpha_2)}{\partial u}\right)^2 \right.$$
$$\left. + P^{-+}(x_1, x_2; \alpha_0)^{-1} \left(\frac{\partial G(x_1 + \alpha_1, x_2)}{\partial u}\right)^2 \right) d\nu(x_1 - \alpha_1, x_2 - \alpha_2)$$

allowing us to characterize

$$I_{\lambda,\alpha_1} \leq \|h_1\|_{L_2(\pi_1)}^2$$

Chamberlain (1986) demonstrates that the space of differentiable functions with compact support is dense in $L_2(\pi)$. Replicating the argument in the proof of Theorem 2.1, we can demonstrate that $\inf_{\lambda \in \tilde{\Lambda}} I_{\lambda,\alpha_1} = 0$. Similarly, we can also show that $\inf_{\lambda \in \tilde{\Lambda}} I_{\lambda,\alpha_2} = 0$.
Q.E.D.

A.7 Convergence rate for the iterative estimator

We use our previous assumption regarding the uniformly manageable class of functions and establish the result regarding the convergence rate of the constructed estimator.

Lemma A.3 *Suppose that a sequence c_n is selected such that $\nu(c_n)/n \to 0$, $K^r/\nu(c_n) \to 0$, $\nu(c_n)K^2/n \to \infty$. Then for any sequence $\hat{\alpha}_n$ with the function $\hat{l}(\alpha)$ corresponding to the maximand of (4.2) such that $\hat{l}^p(\hat{\alpha}_n) \geq \sup_\alpha \hat{l}^p(\alpha) - o_p\left(\sqrt{\frac{\nu(c_n)}{n}}\right)$ we have*

$$\sqrt{\frac{n}{\nu(c_n)}}|\hat{\alpha}_{1n}^* - \alpha_{1,0}| = O_p(1), \quad and \quad \sqrt{\frac{n}{\nu(c_n)}}|\hat{\alpha}_{2n}^* - \alpha_{2,0}| = O_p(1).$$

Proof:

For simplicity of notation, denote $y = (y_1, y_2)$ and $x = (x_1, x_2)$. Let $\hat{\mathcal{P}}_K^{ij}(\cdot; \alpha)$ be the K-term approximation of the probability of the outcome $y_1 = i$ and $y_2 = j$. The conditional likelihood function that uses approximate probabilities can be written as

$$l(\alpha; y, x) = \sum_{i,j=0}^{1} \mathbf{1}\{y_1 = i\}\mathbf{1}\{y_2 = j\} \log \hat{\mathcal{P}}_K^{ij}(x; \alpha),$$

with $Q(\alpha) = E[l(\alpha; y, x)]$, and $\hat{\mathcal{P}}_n^{11}$ defined in Appendix B.1.1. Denote

$$\hat{l}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} l(\alpha; y_i, x_i).$$

Also denote

$$\ell(\alpha; y, x) = \omega_n(|x_1| + |\alpha_1|)\omega_n(|x_2| + |\alpha_2|) \sum_{i,j=0}^{1} \mathbf{1}\{y_1 = i\}\mathbf{1}\{y_2 = j\} \log \mathcal{P}^{ij}(x; \alpha),$$

and

$$\hat{\ell}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \ell(\alpha; y_i, x_i).$$

The iterative estimator suggests computing the coefficients of the orthogonal expansion for each parameter value. For each fixed number of terms K this step is equivalent to running a regression of the dummy variables $y_1 y_2$, $(1 - y_1)y_2$, $y_1(1 - y_2)$, and $(1 - y_1)(1 - y_2)$ on the orthogonal terms $p^{kK}(\cdot)$. We also note that the next steps will replicate the derivation of the rate for the two-step estimator. In fact, we can perform the following decomposition

$$\hat{l}(\alpha) - \hat{\ell}(\alpha_0) = R_1 + R_2 + R_3 + R_4 + R_5 + R_6,$$

where

$$R_1 = \hat{l}(\alpha) - \hat{\ell}(\alpha) - E\left[\hat{l}(\alpha)\right] + E\left[\hat{\ell}(\alpha)\right],$$

$$R_2 = \hat{\ell}(\alpha) - \hat{\ell}(\alpha_0) - E\left[\hat{\ell}(\alpha)\right] + E\left[\hat{\ell}(\alpha_0)\right],$$

$$R_3 = E\left[\hat{l}(\alpha)\right] - E\left[\hat{\ell}(\alpha)\right],$$

$$R_4 = E\left[\hat{\ell}(\alpha)\right] - Q(\alpha),$$

$$R_5 = -E\left[\hat{\ell}(\alpha_0)\right] + Q(\alpha_0),$$

$$R_6 = Q(\alpha) - Q(\alpha_0).$$

Then the first term uniformly in probability converges to zero with stochastic order $o_p\left(n^{-(1-\delta)/2}\right)$, which for sufficiently small δ approaches the parametric rate. The second term is the main "variance" term. Its stochastic order is determined by the imposed bound $\nu(c_n)/n$ and the size of the neighborhood containing parmeters (α_1, α_2). Terms R_3 to R_5 are bias terms that converge to zero under assumptions of the theorem. The last term provides the second-order expansion for the true objective function at the true parameter, which will maintain the quadratic term in $(\alpha_1 - \alpha_{10}, \alpha_1 - \alpha_{10})$. Equating these two terms delivers the stochastic order for the estimated parameters and thus produces the rate of convergence.

Q.E.D.

A.8   Proof of Theorem 4.3

In Appendix A.7 we drew a direct parallel between our proof for the rate of the parameter in the triangular model and the rate that we obtained for the static game of complete information. Following the same strategy, we can note that the results from the proof of Theorem 2.2 apply with the same arguments for the rate optimality for the sequence $n\beta(c_n)^2/\nu(c_n) = O(1)$.

Q.E.D.

A.9   Proof of Theorem 5.1

Step 1 First, we prove that equilibrium belief functions are uniquely determined by parameters α_1 and α_2. In other words, if $(P_1(\cdot, \cdot), P_2(\cdot, \cdot))$ correspond to (α_1, α_2) and $(P_1'(\cdot, \cdot), P_2'(\cdot, \cdot))$ correspond to (α_1', α_2') then $(P_1(\cdot, \cdot), P_2(\cdot, \cdot)) = (P_1'(\cdot, \cdot), P_2'(\cdot, \cdot))$ almost everywhere if and only if $(\alpha_1, \alpha_2) = (\alpha_1', \alpha_2')$.

Provided the structure of the model, we can characterize the system of equations defining the equilibrium choice probabilities as

$$\sigma\Phi^{-1}(P_1) = q_1 + \alpha_1 P_2,$$

$$\sigma\Phi^{-1}(P_2) = q_2 + \alpha_2 P_1,$$

where $q_1 = x_1 - u$ and $q_2 = x_2 - u$. In other words, both probabilities are functions of two composite arguments and parameters α_1 and α_2. We note that provided the differentiability of the distribution of perturbations η_1 and η_2, the solution of this system will also be (locally) differentiable. We can express the Jacobi matrix as

$$J^q = \frac{1}{1 + a_1 a_2 \alpha_1 \alpha_2} \begin{pmatrix} a_1 & a_1 a_2 \alpha_1 \alpha_2 \\ a_1 a_2 \alpha_1 \alpha_2 & a_2 \end{pmatrix},$$

where $a_1 = \phi\left(\Phi^{-1}(P_1)\right)/\sigma$ and $a_2 = \phi\left(\Phi^{-1}(P_2)\right)/\sigma$. We note that this matrix is non-singular if and only if $a_1 a_2 \neq 0$ and $a_1 a_1 \alpha_1 \alpha_2 \neq 1$. As demonstrated in Bajari, Hong, Krainer, and Nekipelov (2010a), the set of argument values (q_1, q_2) where the system of equilibrium beliefs has multiple solutions is compact. Denote this set S^m. Whenever $(q_1, q_2) \in S^m$, matrix J^q can be constructed at each solution pair. Provided that we impose a trivial equilibrium selection rule, the effective Jacobi matrix $\bar{J}^q$ is a simple average of Jacobi matrices in all solutions. We note that monotonicity and differentiability properties apply to matrices $\bar{J}^q$ and J^q in the same way. Thus the "effective" Jacobi matrix can be defined as $J^q \mathbf{1}\{(q_1, q_2) \notin S^m\} + \bar{J}^q \mathbf{1}\{(q_1, q_2) \in S^m\}$. The matrix defined this way is *globally non-singular* since we replace it with a non-singular matrix at each point where there are multiple equilibria and thus J^q is rank-defficient.

We note that functions $P_1(\cdot)$ and $P_2(\cdot)$ are differentiable and strictly monotone if the cdf $\Phi(\cdot)$ is strictly positive on $\mathbb{R}$. Moreover, as the density $\phi(\cdot)$ is unimodal, then equation $a_1 a_1 \alpha_1 \alpha_2 = 1$ has at most two solutions. If P_1^* and P_2^* is a solution of this equation, then the set of q_1 and q_2 that lead to equilibrium beliefs correspond to a cut of the graphs of $P_1(\cdot)$ and $P_2(\cdot)$: $\{(q_1, q_2) : P_1(q_1, q_2) = P_1^*, P_2(q_1, q_2) = P_2^*\}$. Provided that the graph of a differentiable and strictly monotone function has Lebesque measure zero, its cut also has Lebesque measure zero (due to Caratheodori's theorem).

We then analyze the Jacobi matrix of equilibrium beliefs with respect to strategic interaction parameters. Directly evaluating the Jacobi matrix

$$J^\alpha = \begin{pmatrix} \frac{\partial P_1}{\partial \alpha_1} & \frac{\partial P_1}{\partial \alpha_2} \\ \frac{\partial P_2}{\partial \alpha_1} & \frac{\partial P_2}{\partial \alpha_2} \end{pmatrix} = \frac{a_1 a_2}{1 + \alpha_1 \alpha_2 a_1 a_2} \begin{pmatrix} P_1/a_2 & \alpha_1 P_1 \\ \alpha_2 P_2 & P_2/a_1 \end{pmatrix}.$$

This matrix is non-singular whenever $a_1 a_2 \neq 0$ and $a_1 a_2 \alpha_1 \alpha_2 \neq 1$. Thus J^α is non-singular if and only if J^q is non-singular. However, we determined that J^q is non-singular almost everywhere and at the points of its singularity it is replaced with matrix $\bar{J}^q$. Therefore, J^α is globally non-singular. Thus taking an arbitrary point in the support of equilibrium beliefs, we can uniquely solve for the pair of the strategic interaction parameters and for each pair of strategic interaction parameters there exists a unique pair of equilibrium beliefs.

Step 2 Second, we show that from observed conditional expectations $E[Y_1|x_1, x_2]$, $E[Y_1|x_1, x_2]$, and $E[Y_1 Y_2|x_1, x_2]$ we can uniquely recover the corresponding equilbrium choice probabilities. In Appendix E we demonstrate how one can recover the density of the distribution of unobserved heterogeneity when strategic interaction parameters are given. The core of the argument was in defining probability measures that assign to all subsets of $\mathbb{R}^2$ of the form $S = (-\infty, x_1] \times$

$(-\infty, x_2]$ values corresponding to belief probabilities $P_1(\cdot)$, $P_1(\cdot)$, and $P_1(\cdot)P_2(\cdot)$ evaluated at point (x_1, x_2). By the uniqueness of extension of the measure (see Dunford and Schwartz (1965)), each measure will be uniquely defined on Borel subsets of $\mathbb{R}^2$. Without loss of generality, we assume that $\alpha_1, \alpha_2 > 0$ (otherwise, we "flip" the signs of the derivatives of the belief functions with their absolute values). Provided that the distribution $\Phi(\cdot)$ is twice continuously differentiable, we can express the characteristic function defined in Appendix E as

$$\chi_{P_1}(t_1, t_2) = \int e^{-it_1 q_1 - it_2 q_2} \frac{\partial^2 P_1(q_1, q_2)}{\partial q_1 \partial q_2} dq_1 dq_2$$

Then

$$\lim_{t_1 \to 0} \frac{\chi_{P_1}(t_1, t_2)}{it_1} = -\int e^{-it_2 q_2} q_1 \frac{\partial^2 P_1(q_1, q_2)}{\partial q_1 \partial q_2} dq_1 dq_2 = \int e^{-it_2 q_2} \frac{\partial P_1(q_1, q_2)}{\partial q_2} dq_1 dq_2.$$

Previously we denoted $a_1 = \frac{\phi(\Phi^{-1}(P_1))}{\sigma}$ and $a_2 = \frac{\phi(\Phi^{-1}(P_2))}{\sigma}$. Then

$$\frac{\partial P_1}{\partial q_2} = \frac{a_1 a_2}{1 - \alpha_1 \alpha_2 a_1 a_2} \alpha_2.$$

Provided our assumption regarding the signs of α_1 and α_2, we conclude that

$$\frac{a_1 a_2}{1 - \alpha_1 \alpha_2 a_1 a_2} \leq \frac{\phi\left(\frac{q_1 + \alpha_1 q_2}{\sigma}\right) \phi\left(\frac{q_2 + \alpha_2 q_1}{\sigma}\right)}{1 - \frac{\alpha_1 \alpha_2}{2\pi}}$$

The function

$$\int_{-\infty}^{+\infty} \phi\left(\frac{q_1 + \alpha_1 q_2}{\sigma}\right) \phi\left(\frac{q_2 + \alpha_2 q_1}{\sigma}\right) dq_1$$

is bounded and decreasing at infinity. Moreover, provided that $\int t^2 \phi(t) \, dt < \infty$, its Fourier transform exists as a regular complex-valued function. Denote this function

$$A_1(t_2) = \int e^{-it_2 q_2} \frac{a_1 a_2}{1 - \alpha_1 \alpha_2 a_1 a_2} \, dq_1 \, dq_2.$$

Similarly, we can conclude that

$$A_2(t_1) = \int e^{-it_1 q_1} \frac{a_1 a_2}{1 - \alpha_1 \alpha_2 a_1 a_2} \, dq_1 \, dq_2,$$

is a regular complex-valued function. Thus if $\mathcal{Q}(t_1, t_2)$ is the Fourier transform of $E[Y_1 | x_1, x_2]$ and $\mathcal{F}(t_1, t_2)$ is the Fourier transform of $E[Y_2 | x_1, x_2]$ then

$$\mathcal{Q}(0, t_2) = \frac{\alpha_2}{it_2} A_1(t_2) \chi_v(t_2) \left(1 + \pi it_2 \delta(t_2)\right),$$

$$\mathcal{F}(t_1, 0) = \frac{\alpha_1}{it_2} A_2(t_1) \chi_u(t_1) \left(1 + \pi it_1 \delta(t_1)\right).$$

Performing an inverse Fourier transform, we find that

$$\alpha_2 g_v(v) = \frac{1}{2\pi} \int e^{it_2 v} \frac{it_2 \mathcal{Q}(0, t_2)}{A_1(t_2)} \, dt_2,$$

$$\alpha_1 g_u(u) = \frac{1}{2\pi} \int e^{it_1 u} \frac{it_1 \mathcal{F}(t_1, 0)}{A_2(t_1)} \, dt_1,$$

(A.4)

In Appendix E we demonstrate how one recovers the density $g_{uv}(\cdot, \cdot)$. This allows us to also recover the marginal distributions from this density. We thus recover the strategic interaction parameters from expressions (A.4).

Q.E.D.

A.10 Proof of Theorem 5.2

A.10.1 Proof of result (i)

To compute the information of the static game model with incomplete information, we note that this game can have multiple equilibria corresponding to multiple solutions of the system of equilibrium beliefs. Provided the continuous differentiability of the distribution of random perturbations, we can characterize the boundary of the set of multiple equilibria as the set of points on $\mathbb{R}^2$ where the curves corresponding to the best responses of the players to their beliefs regarding their opponents touch for the first time. This corresponds to the set of points on $\mathbb{R}^2$ where:

$$\sigma \Phi^{-1}(P_1) = q_1 + \alpha_1 P_2,$$
$$\sigma \Phi^{-1}(P_2) = q_2 + \alpha_2 P_1,$$
$$\alpha_1 \phi \left(\frac{1}{\sigma} \left(q_1 + \alpha_1 P_2 \right) \right) = \left(\alpha_2 \phi \left(\Phi^{-1}(P_2) \right) \right)^{-1}.$$

For given parameters α_1, α_2, this defines a mapping from the set of covariates q_1, q_2 to the beliefs. This mapping reduces the dimensionality of the overall mapping by 2, as it incorporates the original system of equations for the beliefs and the restriction on the derivatives of the belief functions. It will be a 1-dimensional closed curve $e(q_1, q_2) = 0$. This curve will be differentiable in the strategic interaction parameters due to continuous differentiability of the density of the payoff noise. This curve represents the boundary of the set of multiple equilibria, which we denote $S^m(\alpha_1, \alpha_2)$.

The likelihood of the model can then be characterized by four objects:

$$E\left[Y_1 Y_2 \,|\, x_1, x_2\right] = P_{11}(x_1, x_2; \alpha) = \int \Phi \left(\frac{x_1 - u + \alpha_1 P_2(x_1 - u, x_2 - v)}{\sigma} \right)$$
$$\times \Phi \left(\frac{x_2 - v + \alpha_2 P_2(x_1 - u, x_2 - v)}{\sigma} \right) g(u, v) \, du \, dv,$$
$$E\left[Y_1 \,|\, x_1, x_2\right] = Q_1(x_1, x_2; \alpha) = \int \Phi \left(\frac{x_1 - u + \alpha_1 P_2(x_1 - u, x_2 - v)}{\sigma} \right) g(u, v) \, du \, dv,$$
$$E\left[Y_2 \,|\, x_1, x_2\right] = P_1(x_1, x_2; \alpha) = \int \Phi \left(\frac{x_2 - v + \alpha_2 P_1(x_1 - u, x_2 - v)}{\sigma} \right) g(u, v) \, du \, dv,$$
$$\Pr\left((X_1 - U, X_2 - V) \in S^m(\alpha_1, \alpha_2) \,|\, x_1, x_2 \right) = \Delta(x_1, x_2; \alpha)$$
$$= \int \mathbf{1}\{ (x_1 - u, x_2 - v) \in S^m(\alpha_1, \alpha_2) \} g(u, v) \, du \, dv.$$

We assume that $\alpha_1\alpha_2 > 0$ without loss of generality. We construct the probabilities corresponding to observed equilibrium outcomes as

$$P^{++}(x_1, x_2; \alpha) = P_{11}(x_1, x_2; \alpha) - \frac{1}{2}\Delta(x_1, x_2; \alpha),$$

$$P^{-+}(x_1, x_2; \alpha) = P_1(x_1, x_2; \alpha) - P_{11}(x_1, x_2; \alpha) + \frac{1}{2}\Delta(x_1, x_2; \alpha),$$

$$P^{+-}(x_1, x_2; \alpha) = Q_1(x_1, x_2; \alpha) - P_{11}(x_1, x_2; \alpha) + \frac{1}{2}\Delta(x_1, x_2; \alpha),$$

$$P^{--}(x_1, x_2; \alpha) = 1 - P_1(x_1, x_2; \alpha) - Q_1(x_1, x_2; \alpha) + P_{11}(x_1, x_2; \alpha) - \frac{1}{2}\Delta(x_1, x_2; \alpha).$$

Denote the gradients $D_1(x_1, x_2; \alpha) = \frac{\partial}{\partial \alpha'}\left(P_{11}(x_1, x_2; \alpha) - \frac{1}{2}\Delta(x_1, x_2; \alpha)\right)$, $D_2(x_1, x_2; \alpha) = \frac{\partial}{\partial \alpha'}P^{-+}(x_1, x_2; \alpha)$, and $D_3(x_1, x_2; \alpha) = \frac{\partial}{\partial \alpha'}P^{+-}(x_1, x_2; \alpha)$.

We focus on the square root of the density corresponding to the likelihood of the model:

$$r(y_1, y_2|x_1, x_2; \alpha)^{1/2} = y_1 y_2 P^{++}(x_1, x_2; \alpha)^{1/2} + (1 - y_1)y_2 P^{-+}(x_1, x_2; \alpha)^{1/2}$$
$$+ y_1(1 - y_2)P^{+-}(x_1, x_2; \alpha)^{1/2} + (1 - y_1)(1 - y_2)P^{--}(x_1, x_2; \alpha)^{1/2}$$

Then we can express the mean-square gradient of this density as

$$\psi_\alpha(x_1, x_2) = \frac{1}{2}\left\{y_1 y_2 P^{++}(x_1, x_2; \alpha)^{-1/2} - (1 - y_1)(1 - y_2)P^{--}(x_1, x_2; \alpha)^{-1/2}\right\}$$
$$\times D_1(x_1, x_2; \alpha)$$
$$+ \frac{1}{2}\left\{(1 - y_1)y_2 P^{-+}(x_1, x_2; \alpha)^{-1/2} - (1 - y_1)(1 - y_2)P^{--}(x_1, x_2; \alpha)^{-1/2}\right\}$$
$$\times D_2(x_1, x_2; \alpha)$$
$$+ \frac{1}{2}\left\{y_1(1 - y_2)P^{+-}(x_1, x_2; \alpha)^{-1/2} - (1 - y_1)(1 - y_2)P^{--}(x_1, x_2; \alpha)^{-1/2}\right\}$$
$$\times D_3(x_1, x_2; \alpha).$$

We note that the corresponding score has mean zero and that conditional on the covariates, the terms in this expression are positively correlated. Then by definition,

$$I_\alpha = 4 \int \psi_\alpha(x_1, x_2)\psi_\alpha(x_1, x_2)' \, d\mu$$

Thus, if ν is the measure on $\mathbb{R}^2$ corresponding to the distribution of x_1 and x, following the approach in the derivation of information of the complete information model, we define the measures on Borel subsets of $\mathbb{R}^2$

$$\pi_1(A) = \int_A \frac{1 - P^{-+}(x_1, x_2; \alpha_0) - P^{+-}(x_1, x_2; \alpha_0)}{P^{++}(x_1, x; \alpha_0)P^{--}(x_1, x; \alpha_0)} d\nu(x_1, x)$$

and

$$\pi_2(A) = \int_A \frac{1 - Q_1(x_1, x_2; \alpha_0)}{P^{-+}(x_1, x; \alpha_0)P^{--}(x_1, x; \alpha_0)} d\nu(x_1, x),$$

and

$$\pi_3(A) = \int_A \frac{1 - P_1(x_1, x_2; \alpha_0)}{P^{-+}(x_1, x; \alpha_0) P^{--}(x_1, x; \alpha_0)} d\nu(x_1, x).$$

Due to discovered positive correlation between the components of the mean-square gradient, we can evaluate the information as

$$I_\alpha \geq \|D_1(x_1, x_2; \alpha_0)\|^2_{L_2(\pi_1)} + \|D_2(x_1, x_2; \alpha_0)\|^2_{L_2(\pi_2)} + \|D_3(x_1, x_2; \alpha_0)\|^2_{L_2(\pi_3)}.$$

Then we can construct the measure π^* which "picks out" for each Borel subset A one of the measures π_1 or π_2 which gives this set less weight: $\pi^*(A) = \min\{\pi_1(A), \pi_2(A)\}$. Based on this structure of the measure, we can write:

$$I_\alpha \geq \|D_1(x_1, x_2; \alpha_0)\|^2_{L_2(\pi^*)} + \|D_2(x_1, x_2; \alpha_0)\|^2_{L_2(\pi^*)} + \|D_3(x_1, x_2; \alpha_0)\|^2_{L_2(\pi_3)}.$$

By combining the triangle inequality and taking into account the non-negativity of the square, we can evaluate

$$I_\alpha \geq \|D_1(x_1, x_2; \alpha_0) + D_2(x_1, x_2; \alpha_0)\|^2_{L_2(\pi^*)}.$$

Then we note that

$$D_1(x_1, x_2; \alpha_0) + D_2(x_1, x_2; \alpha_0) = \int \phi\left(\frac{1}{\sigma}\Phi^{-1}(P_1)\right)\left(\alpha_1 \frac{\partial P_2}{\partial \alpha} + (P_2, 0)'\right) g(u, v) \, du \, dv$$

We denote $t_1 = x_1 - u$ and $t_2 = x - v$. Then

$$D_1(x_1, x_2; \alpha_0) + D_2(x_1, x_2; \alpha_0) = \int \phi\left(\frac{1}{\sigma}\Phi^{-1}(P_1(t_1, t_2))\right)$$
$$\times \left(\alpha_1 \frac{\partial P_2(t_1, t_2)}{\partial \alpha} + (P_2(t_1, t_2), 0)'\right) g(t_1 + x_1, t_2 + x_2) \, dt_1 \, dt_2.$$

Denote $w(t_1, t_2) = \phi\left(\frac{1}{\sigma}\Phi^{-1}(P_1(t_1, t_2))\right)\left(\alpha_1 \frac{\partial P_2(t_1, t_2)}{\partial \alpha} + (P_2(t_1, t_2), 0)'\right)$. Then we can express

$$D_1(x_1, x_2; \alpha_0) + D_2(x_1, x_2; \alpha_0) = \int w(t_1, t_2) g(x_1 + t_1, x_2 + t_2) \, dt_1 \, dt_2.$$

Suppose that $S \subset \mathbb{R}^2$ is a compact set such that $\pi^*(S) > C$. Then given that $g(\cdot, \cdot)$ is continuous and strictly positive, there exists $M(t_1, t_2) = \inf_{(x_1, x_2) \in S} |g(x_1 + t_1, x_2 + t_2)|$ which is not equal to zero for at least some $(t_1, t_2) \in \mathbb{R}^2$. We take $\sqrt[4]{\epsilon} = \sup_{t \in [-B, B] \times [-B, B]} |M(t)|$, where B is selected such that $[-B, B] \times [-B, B]$ contains at least one point where $M(t) \neq 0$. Suppose that the supremum is attained at point (t_1^*, t_2^*). By continuity, there exists some neighborhood of (t_1^*, t_2^*) where $M(t) > \sqrt{\epsilon/2}$. Denote the size of this neighborhood R. By construction $w(t_1, t_2)$ is a continuous function which is not equal to zero (given that we assumed that $\alpha_1 \alpha_2 > 0$, we have $\alpha_1 \frac{\partial P_2}{\partial \alpha_1} > 0$). Thus this function attains its lower bound in every compact set and that lower bound is above zero $\inf_{(t_1, t_2) \in B_R(t_1^*, t_2^*)} \|w(t_1, t_2)\| = A\sqrt[4]{\epsilon} > 0$.

We substitute our evaluations into the bound for the information:

$$I_\alpha \geq \|D_1(x_1, x_2; \alpha_0) + D_2(x_1, x_2; \alpha_0)\|^2_{L_2(\pi^*)} \geq \|(D_1(x_1, x_2; \alpha_0) + D_2(x_1, x_2; \alpha_0))\, \mathbf{1}_S\|^2_{L_2(\pi^*)}$$

$$\geq CA^2 \sqrt{\epsilon} \left\|\int_{\mathbb{R}} M(t_1, t_2)\, dt\right\|^2 I_{2\times 2} \geq \frac{1}{2} CA^2 R^2 \epsilon\, I_{2\times 2} > 0,$$

where $I_{2\times 2}$ is the identity matrix. Therefore the information corresponding to parameters α_1 and α_2 is strictly positive.

A.10.2   Proof of result (ii)

Consider the measures introduced in the proof of Theorem 5.2. Suppose that measure $\pi^{**}(A)$ picks out the measure assigning the highest weight to the set A. Then we can see that

$$I_\alpha \leq \|D_1(x_1, x_2; \alpha_0)\|^2_{L_2(\pi^{**})} + \|D_2(x_1, x_2; \alpha_0)\|^2_{L_2(\pi^{**})} + \|D_3(x_1, x_2; \alpha_0)\|^2_{L_2(\pi^{**})}$$

$$+ 2\|D_1 D_2\|^2_{L_2(\pi^{**})} + 2\|D_1 D_3\|^2_{L_2(\pi^{**})} + 2\|D_2 D_3\|^2_{L_2(\pi^{**})}.$$

Provided that the complete information game is constructed to be a limit of the incomplete information game and provided that we use the same equilibrium selection rule, we can conclude that

$$P_1(u - x_1, v - x_2) \to \mathbf{1}\{u \leq x_1 + \alpha_1\} - \frac{1}{2}\mathbf{1}\{x_1 \leq u \leq x_1 + \alpha_1\}$$

as $\sigma \to 0$. Therefore, by dominated convergence theorem, we conclude that

$$D_1(x_1, x_2; \alpha) \to \left(\int_{-\infty}^{x_2+\alpha_2} g(x_1 + \alpha_1, v)\, dv, \ \int_{-\infty}^{x_1+\alpha_1} g(u, x_2 + \alpha_2)\, du\right)'$$

Therefore, component-wise

$$\int_{-\infty}^{x_2+\alpha_2} g(x_1 + \alpha_1, v)\, dv \leq g_u(x_1 + \alpha_1).$$

Using a similar evaluation, we can provide the limits for all components leading to

$$D_2(x_1, x_2; \alpha) \to \bar{D}_2(x_1, x_2; \alpha), \ \ \text{and} \ \ D_3(x_1, x_2; \alpha) \to \bar{D}_3(x_1, x_2; \alpha)$$

such that

$$\bar{D}_i(x_1, x_2; \alpha) \leq (g_u(x_1 + \alpha_1),\, g_v(x_2 + \alpha_2))',$$

with $i = 1, 2, 3$. We evaluate the information as

$$I_\alpha \to \bar{I} \leq 3I_{2\times 2}$$

$$\times \left(\|g_u(x_1 + \alpha_1)\|^2_{L_2(\pi^{**})} + \|g_v(x_2 + \alpha_2)\|^2_{L_2(\pi^{**})} + 2\|g_u(x_1 + \alpha_1)\|_{L_2(\pi^{**})}\|g_u(x_1 + \alpha_1)\|_{L_2(\pi^{**})}\right),$$

where $I_{2\times 2}$ is a 2×2 identity matrix. In this evaluation the marginal densities of the unobserved heterogeneity can be treated as separate functions. Provided that the space of twice continuously

differentiable functions is dense in $L_2(\pi^{**})$, for any $\epsilon > 0$ we can find an element of this space such that $\|f\|_{L_2(\pi^{**})} < \sqrt{\epsilon}$. As a result, when we select $g_u = g_v = f$, the limiting information matrix is

$$\bar{I} \le 9\epsilon I_{2\times 2}.$$

As the choice of ϵ was arbitrary, the resulting information converges to zero.
Q.E.D.

B   Estimators with optimal rate

B.1   Triangular model: Two-step estimator

Step 1. Consider the family of normalized Hermite polynomials and denote $h_l(x) = (\sqrt{2\pi} l!)^{-1/2} e^{-\frac{x^2}{4}} H_l(x)$, where $H_l(\cdot)$ is the l-th degree Hermite polynomial. Also denote $\mathcal{H}_l(x) = \int_{-\infty}^{x} h_l(z)\,dz$. We note that this sequence is orthonormal for the inner product defined as $\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g(x)\,dx$. We take the sequence $c_n \to \infty$, and define the function $\omega_n(x) = \mathbf{1}\{|x| \le c_n\}$ and estimate the probability of both indicators are equal to zero ($y_1 = y_2 = 0$) as

$$\hat{\mathcal{P}}_n^{00}(x_1, x) = \sum_{l_1, l_2 = 1}^{K(n)} a_{l_1 l_2} \omega_n(x_1) \left[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x_1)\right] \omega_n(x) \left[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x)\right]$$

The estimates can be obtained via a regression of $\omega_n(x_1) \left[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x_1)\right] \omega_n(x) \left[\mathcal{H}_{l_2}(c_n) - \mathcal{H}_{l_2}(x)\right]$ on the indicators $(1 - y_1)(1 - y_2)$. Then the estimator for the joint density of errors can be obtained from the regression coefficients as

$$\hat{g}_n(x_1, x) = \sum_{l_1, l_2 = 1}^{K(n)} \hat{a}_{l_1 l_2} \omega_n(x_1) h_{l_1}(x_1) \omega_n(x) h_{l_2}(x).$$

Step 2. Using the estimator for the density, we compute the fitted values for conditional probabilities of $y_1 = y_2 = 1$ and $y_1 = 0$, $y_2 = 1$ as

$$\hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x) = \sum_{l_1, l_2 = 1}^{K(n)} \hat{a}_{l_1 l_2} \omega_n(x_1 + \alpha) \left[\mathcal{H}_{l_1}(x_1 + \alpha) - \mathcal{H}_{l_1}(-c_n)\right] \omega_n(x) \left[\mathcal{H}_{l_1}(x) - \mathcal{H}_{l_1}(-c_n)\right],$$

and

$$\hat{\mathcal{P}}_n^{01}(x_1 + \alpha, x) = \sum_{l_1, l_2 = 1}^{K(n)} \hat{a}_{l_1 l_2} \omega_n(x_1 + \alpha) \left[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x_1 + \alpha)\right] \omega_n(x) \left[\mathcal{H}_{l_1}(x) - \mathcal{H}_{l_1}(-c_n)\right].$$

Using these fitted probabilities we can form the conditional log-likelihood function

$$\begin{aligned} l\left(\alpha; y_1, y_2, x_1, x\right) =\, & y_1 y_2 \omega_n(x_1 + \alpha) \omega_n(x) \log \hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x) \\ & + (1 - y_1) y_2 \omega_n(x_1 + \alpha) \omega_n(x) \log \hat{\mathcal{P}}_n^{01}(x_1 + \alpha, x). \end{aligned}$$

Then we can express the empirical score as

$$s\left(\alpha; y_1, y_2, x_1, x\right) = \left[\frac{y_1 y_2}{\hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x)}\right.$$
$$\left. - \frac{(1 - y_1)y_2}{\hat{\mathcal{P}}_n^{01}(x_1 + \alpha, x)}\right]\frac{\partial \hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x)}{\partial \alpha}\omega_n(x_1 + \alpha)\omega_n(x)$$

This expression can be rewritten as

$$s\left(\alpha; y_1, y_2, x_1, x\right) = \frac{\omega_n(x_1 + \alpha)\omega_n(x)y_2}{\hat{\mathcal{P}}_n^{11}(c_n, x)} \frac{y_1 - \frac{\hat{\mathcal{P}}_n^{11}(x_1+\alpha,x)}{\hat{\mathcal{P}}_n^{11}(c_n,x)}}{\left(1 - \frac{\hat{\mathcal{P}}_n^{11}(x_1+\alpha,x)}{\hat{\mathcal{P}}_n^{11}(c_n,x)}\right)\frac{\hat{\mathcal{P}}_n^{11}(x_1+\alpha,x)}{\hat{\mathcal{P}}_n^{11}(c_n,x)}} \frac{\partial \hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x)}{\partial \alpha}.$$

Setting the empirical score equal to zero, we obtain the estimator for α_0 as

$$\hat{\alpha}_n^* = \operatorname{argmax}_\alpha \frac{1}{n}\sum_{i=1}^n l\left(\alpha; y_{1i}, y_{2i}, x_{1i}, x_i\right). \tag{B.1}$$

B.1.1   Iterative estimator

As in the case of the binary triangular system we approximate the error density using normalized Hermite polynomials. We take a sequence $c_n \to \infty$ and define the function $\omega_n(x) = \mathbf{1}\{|x| \le c_n\}$. We introduce the function

$$\Delta(x_1, x_2; \alpha_1, \alpha_2) = \sum_{l_1, l_2=1}^{K(n)} a_{l_1 l_2}\omega_n(x_1)\left[\mathcal{H}_{l_1}(x_1 + \alpha_1) - \mathcal{H}_{l_1}(x_1)\right]\omega_n(x_2)\left[\mathcal{H}_{l_2}(x_2 + \alpha_2) - \mathcal{H}_{l_2}(x_2)\right].$$

Then we approximate the probabilities of the indicators taking values $y_1 = y_2 = 0$ as

$$\hat{\mathcal{P}}_n^{00}(x_1, x_2) = \sum_{l_1, l_2=1}^{K(n)} a_{l_1 l_2}\omega_n(x_1)\left[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x_1)\right]\omega_n(x_2)\left[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x_2)\right]$$
$$- \frac{1}{2}\Delta(x_1, x_2; \alpha_1, \alpha_2).$$

Similarly, we approximate the remaining probabilities

$$\hat{\mathcal{P}}_n^{11}(x_1, x_2) = \sum_{l_1, l_2=1}^{K(n)} \hat{a}_{l_1 l_2}\omega_n(x_1 + \alpha_1)\left[\mathcal{H}_{l_1}(x_1 + \alpha_1) - \mathcal{H}_{l_1}(-c_n)\right]\omega_n(x_2)\left[\mathcal{H}_{l_1}(x_2 + \alpha_2) - \mathcal{H}_{l_1}(-c_n)\right]$$
$$- \frac{1}{2}\Delta(x_1, x_2; \alpha_1, \alpha_2)$$

and

$$\hat{\mathcal{P}}_n^{01}(x_1, x_2) = \sum_{l_1, l_2=1}^{K(n)} \hat{a}_{l_1 l_2}\omega_n(x_1 + \alpha_1)\left[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x_1 + \alpha_1)\right]\omega_n(x_2)\left[\mathcal{H}_{l_1}(x_2) - \mathcal{H}_{l_1}(-c_n)\right].$$

Using these approximations to the joint probabilities for the binary indicators we can form the conditional log-likelihood function

$$l\left(\alpha_1, \alpha_2; y_1, y_2, x_1, x_2\right) = y_1 y_2 \omega_n(x_1)\omega_n(x_2) \log \hat{\mathcal{P}}_n^{11}(x_1, x_2)$$
$$+ (1 - y_1)y_2\omega_n(x_1)\omega_n(x_2) \log \hat{\mathcal{P}}_n^{01}(x_1, x_2)$$
$$+ y_1(1 - y_2)\omega_n(x_1)\omega_n(x_2) \log \hat{\mathcal{P}}_n^{10}(x_1, x_2)$$
$$+ (1 - y_1)(1 - y_2)\omega_n(x_1)\omega_n(x_2) \log \hat{\mathcal{P}}_n^{00}(x_1, x_2).$$

We can consider thw sample profile log-likelihood

$$\hat{l}^p(\alpha_1, \alpha_2) = \sup_{a_{11},\dots,a_{KK}} \frac{1}{n} \sum_{i=1}^n l\left(\alpha_1, \alpha_2; y_{1i}, y_{2i}, x_{1i}, x_{2i}\right).$$

The parameter estimates can be obtained as maximizers of the profile log-likelihood:

$$(\hat{\alpha}_{1n}^*, \hat{\alpha}_{2n}^*) = \mathrm{argmax}_{\alpha_1,\alpha_2} \hat{l}^p(\alpha_1, \alpha_2). \tag{B.2}$$

C   Examples of convergence rates for common classes of distributions

Logistic errors with logistic covariates

To evaluate function $\nu(\cdot)$ we consider the one dimensional case. Let $F(\cdot)$ be the cdf of interest and $\phi(\cdot)$ be the pdf of the covariates. We evaluate the term of interest as

$$\int_0^c \frac{\phi(x)}{1 - F(x)}\,dx = \int_0^c \frac{e^x}{1 + e^x}\,dx$$

A change of variables $z = e^x$ allows us to re-write this expression as

$$\int_1^{e^c} \frac{dz}{1 + z} = O(c)$$

Given that we have a two-dimensional distribution, we can select $\nu(c) = c^2$. Next, we evaluate function $\beta(\cdot)$, whose leading term can be represented as

$$\int_c^\infty \log((1 + e^x)^{-1})\frac{e^x}{(1 + e^x)^2}\,dx = O(e^{-c}).$$

Therefore, we can select $\beta(c) = e^{-c}$ and the optimal rate will be $\sqrt{n/c_n^2}$ with $c_n e^{c_n}/n = O(1)$. For instance, we can select $c_n = \delta\sqrt{\log n}$ for some $0 < \delta < 1$, delivering convergence rate $\sqrt{n/\log n}$.

Logistic errors with normal covariates

Using the same notation as before, we evaluate the leading term for $\nu(\cdot)$ as

$$\int_0^c \frac{\phi(x)}{1 - F(x)}\, dx = \frac{1}{\sqrt{2\pi}} \int_0^c (1 + e^x) e^{-x^2/2}\, dx = O(1)$$

Then we can express the order of the bias term as

$$\beta(c) = \frac{1}{\sqrt{2\pi}} \int_c^\infty \log(1 + e^x) e^{-x^2/2}\, dx = O(e^{-c^2/2}).$$

As a result, we can use $\nu(c) \equiv 1$ and the bias will vanish. This choice gives the parametric optimal rate $\sqrt{n}$.

Normal errors with logistic covariates

We will use the same approach as before and try to evaluate the function $\nu(\cdot)$ using the leading term of the representation of the integral

$$\int_0^c \frac{\phi(x)}{1 - F(x)}\, dx$$

First note that one can arrive at the asymptotic evaluation for the normal cdf via a change of variable $t = 1/z$ and subsequent Taylor expansion

$$1 - \Phi(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = \frac{1}{\sqrt{2\pi}} \int_0^{1/x} \frac{e^{-1/(2t^2)}}{t^2}\, dt = O\left(\frac{e^{-x^2/2}}{x}\right)$$

Then we obtain that

$$\frac{\phi(x)}{1 - F(x)} = O(xe^{x^2/2 - x}),$$

for sufficiently large x. This means that the leading term for the integral is $O(e^{c^2/2})$. As a result, we find that $\nu(c) = e^{c^2}$. We then evaluate the leading component of the bias term as

$$\int_c^\infty \log\left(\frac{e^{-x^2/2}}{x}\right) \frac{e^x}{(1 + e^x)^2}\, dx = O\left(c^2 e^{-c}\right).$$

Therefore, we can select $\beta(c) = c^2 e^{-c}$, and we can determine the optimal trimming sequence by solving

$$nc_n^4 e^{-c_n^2} = O(1).$$

The convergence rate will correspond to $\sqrt{n} e^{-c_n^2}$. This means that, for instance, selection of $c_n = \log n^{1/2}$ delivers the convergence rate $n^{1/4}$.

Normal errors with normal covariates

Using our previous evaluation of the normal cdf, we can provide the representation for the lead term of the ratio

$$\frac{\phi(x)}{1 - F(x)} = O(x).$$

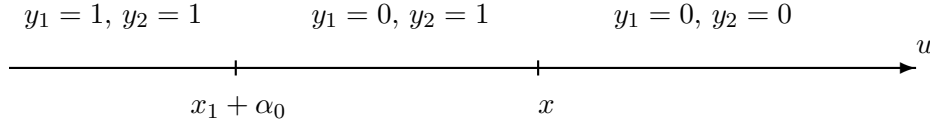Therefore, we can evaluate $\nu(c) = c^4$. Then we evaluate the bias term as

$$\int_c^\infty \log\left(\frac{e^{-x^2/2}}{x}\right) e^{-x^2/2}\, dx = O(ce^{-c^2/2})$$

The optimal rate corresponds to $\sqrt{n}/c_n^2$ with c_n solving $c_n^2 e^{-c_n^2}/n = O(1)$.
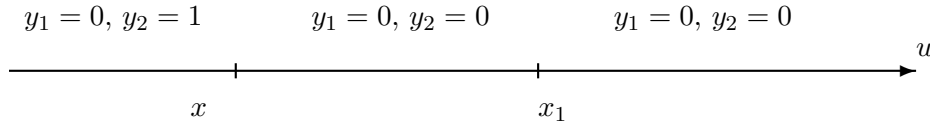
D Zero information in a smooth parametric model

Consider the case where $u = v$ with the standard normal distribution, and x_1 and x are scalar covariates such that $x_1 = x + \sqrt{u}\phi(x_1 + \alpha_0)\phi(x_1)/\Phi(x_1 + \alpha_0) - \alpha_0$ where u is a uniformly distributed variable on $[0, 1]$ and x_1 is a standard normal random variable. This setting is a threshold crossing model. The model is determined by two cases:

Case 1: $x_1 + \alpha_0 < x$

$y_1 = 1, y_2 = 1$ $y_1 = 0, y_2 = 1$ $y_1 = 0, y_2 = 0$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad u$

$\qquad\qquad x_1 + \alpha_0 \qquad\qquad\qquad\qquad\qquad x$

Case 2: $x_1 \geq x$

$y_1 = 0, y_2 = 1$ $y_1 = 0, y_2 = 0$ $y_1 = 0, y_2 = 0$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad u$

$\qquad\quad x \qquad\qquad\qquad\qquad\qquad\qquad x_1$

The score will be above zero only in Case 1:

$$s(\alpha_0; y_1, y_2, x_1, x) = \frac{y_1 - \Phi(x_1 + \alpha_0)}{\Phi(x_1 + \alpha_0)\left(\Phi(x_1 + \alpha_0) - \Phi(x)\right)} y_2 \phi(x_1 + \alpha_0).$$

The variance of the score can be expressed as $E\left[\left(\Phi(x_1 + \alpha_0)\left(\Phi(x_1 + \alpha_0) - \Phi(x)\right)\right)^{-1}\phi(x_1 + \alpha_0)^2\right]$. This expectation does not exist. In fact, as $x_1 \to -\infty$, $\Phi(x_1 + \alpha_0) - \Phi(x) = -\sqrt{u}\phi(x_1 + \alpha_0)^2\phi(x_1)/\Phi(x_1 + \alpha_0) + o_p(1)$, and

$$E\left[\frac{\phi(x_1 + \alpha_0)^2}{\Phi(x_1 + \alpha_0)\left(\Phi(x_1 + \alpha_0) - \Phi(x)\right)}\right] \geq 2\int_0^1 \frac{du}{\sqrt{u}}\int_{-K}^0 dx_1 \to +\infty, \text{ as } K \to +\infty,$$

provided that the distribution of x approaches the distribution of $x_1 - \alpha_0$. Therefore, the score in this model has infinite variance, and the Fisher information for parameter α_0 in this parametric model is equal to zero.

E   Recovering the distribution of the unobserved heterogeneity from the observed binary choice probabilities.

We used the Fourier transformation technique to recover the distribution of unobserved heterogeneity from the observed choice probabilities. However, it needs to be emphasized that this is a non-trivial problem and it extends beyond simple functional inversions. The importance of correct treatment of functional inverse problems is clearly understood in the nonlinear measurement errors literature, e.g. Schennach (2004), Hu and Schennach (2008) among others. As we demonstrate in this appendix, the problems of functional inversion from the observed choice probabilities are much more severe, rather like those noted in Zinde-Walsh (2009), and they require advanced generalized function techniques, outlined for instance, in Vladimirov (1971). We omit the technical definitions and introductory results for operations with generalized functions in the context of convolution problems, as these are expressed in detail in Zinde-Walsh (2009).

We start with a very simple example where one observes a single choice probability $P(x)$ as a function of the observed covariate x which describes the expected response of the individual to the realization of the unobserved heterogeneity u. The actual binary decision of the individual can be expressed as $Y = \mathbf{1}\{X - U \geq 0\}$. The realizations of the unobserved heterogeneity are not observed by the researcher and probability $P(\cdot)$ corresponds to the expectation of the binary decision with respect to the distribution of the unobserved heterogeneity:

$$P(x) = \int_{-\infty}^{+\infty} \mathbf{1}\{x - u \geq 0\} g(u) \, du.$$

The density of this distribution $g(\cdot)$ is not known to the researcher. However, it is known that it has full support on $\mathbb{R}$, as does the distribution of the observed covariate x. We note that the right-hand side of the expression for the probability is a convolution between the indicator function $\mathbf{1}\{\cdot \geq 0\}$ and the density function $g(\cdot)$. To perform the inversion, one may therefore use an attractive feature of the convolution with respect to the Fourier transform: the Fourier transform of the convolution is equal to the product of Fourier transforms of the convolved functions. Note that the Fouier transform of the density $g(\cdot)$ is a characteristic function of the distribution of U, $\chi_u(\cdot)$. However, we note that the Fourier image of $\mathbf{1}\{\cdot \geq 0\}$ exists only as a generalized function:

$$\int e^{-itx} \mathbf{1}\{x \geq 0\} \, dx = \frac{1}{it} + \pi\delta(t),$$

where $\delta(\cdot)$ is a Dirac δ-function. An important observation is that δ - function is a singular generalized function and it can only be defined in the way it operates on the test function $\varphi(\cdot)$:

$\int \varphi(x)\delta(x)\,dx = \varphi(0)$. Thus one cannot perform non-linear algebraic manipulations on it. If we perform the Fourier transform of the choice probability (and denote $\mathcal{F}(t) = \int e^{-itx}P(x)\,dx$, then we can write

$$\mathcal{F}(t) = \left(\frac{1}{it} + \pi\delta(t)\right)\chi(t),$$

where the characteristic function of the distribution of unobserved heterogeneity (which is the object of interest) is multiplied by the generalized function. The irregularity of the Fourier transform is generated by the fact that the transformed function does not decay at infinity. This problem can be solved in this particular case using a "trick": we multiply both sides by it and then make an inverse Fourier transform. We note that $\int e^{itx}t\delta(t)\chi(t)\,dt = 0$, by the property of delta-function and that

$$g(u) = \frac{1}{2\pi}\int e^{itx}it\mathcal{F}(t)\,dt.$$

This manipulation essentially allows us to mitigate the singularity at zero. One of the objects of interest in our paper is the behavior of the smoothed choice probabilities. We describe them by introducing a smooth symmetric distribution with a cdf $\Phi(\cdot)$ and a pdf $\phi(\cdot)$. Then the smoothed choice probability can be expressed as

$$P_\sigma(x) = \int \Phi(\frac{x-u}{\sigma})g(u)\,du.$$

We observe that the Fourier transform of the cdf resembles the Fourier transform of the indicator function (which is a special case of the cdf). First, we represent

$$\Phi(x) = \frac{1}{2} + \left(\Phi(x) - \frac{1}{2}\right).$$

Due to the symmetry of the distribution about zero, function $\Phi(x) - \frac{1}{2}$ is odd. Next we note that

$$\int e^{-itx}\left(\Phi(x) - \frac{1}{2}\right)\,dx = \int \left(\mathbf{1}\{z - x \le 0\} - \frac{1}{2}\right)e^{-itx}\phi(z)\,dz\,dx$$
$$= \int e^{-itz}\phi(z)\,dz\int e^{-itu}\left(\mathbf{1}\{u \ge 0\} - \frac{1}{2}\right)\,du,$$

where we used a change of variable $u = x - z$. Then we note that the first integral is equal to the characteristic function of the distribition $\Phi(\cdot)$ and the second integral is evaluated through its Cauchy principal value and is equal to $1/(it)$.

The Fourier transform of $\frac{1}{2}$ leads to $\pi\delta(t)$. As a result, we can express the Fourier transform of the constructed cdf as the generalized function

$$\int_{-\infty}^{+\infty} \Phi(x)e^{-itx}\,dx = \frac{\chi_\Phi(t)}{it} + \pi\delta(t).$$

We then denote the Fourier transform of $P_\sigma(\cdot)$ by $\mathcal{F}_\sigma(\cdot)$, and we express the Fourier transform of the convolution as

$$\mathcal{F}_\sigma(t) = \left(\frac{\chi_\Phi(\sigma t)}{it} + \pi\delta(t)\right)\chi_u(t).$$

Then we can obtain the density of the distribution of unobserved heterogeneity via a deconvolution:

$$g(u) = \frac{1}{2\pi} \int e^{itu} \frac{it\mathcal{F}_\sigma(t)}{\chi_\Phi(\sigma t)}\, dt.$$

The analysis of two-dimensional cases with partial smoothing poses more challenges. We have noticed that symmetry is essential to obtaining closed-form expressions for the Fourier transforms of the observable choice probabilities. We start our analysis with the case of the triangular model with incomplete information. The triangular incomplete information model is characterized by three conditional expectations: $E[Y_1|x_1, x]$, $E[Y_2|x_1, x]$, and $E[Y_1Y_2|x_1, x]$. We use the expectation of the cross-product to characterize the distribution of unobserved heterogeneity for each parameter value α. We note that

$$E[Y_1Y_2 \,|\, x_1, x] = P_{11}(x_1, x) = \int \mathbf{1}\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) \geq 0\}\Phi\left(\frac{x-v}{\sigma}\right) g(u, v)\, du\, dv.$$

As this expression makes clear, an essential component for recovering the distribution of unobserved heterogeneity is to find the Fourier transform of the indicator. We consider the Fourier transform $\int e^{-it_1y_1 - it_2y_2} \mathbf{1}\{y_1 + \alpha\Phi(y_2) \geq 0\}\, dy_1\, dy_2$.

We notice that function $\mathbf{1}\{y_1 + \alpha\Phi(y_2) \geq 0\} - \frac{1}{2}$ is centrally anti-symmetric about the point $(-\frac{\alpha}{2}, 0)$. In fact, if $z_1 = y_1 + \frac{\alpha}{2}$ and $z_2 = y_2$, then

$$\mathbf{1}\{z_1 + \alpha\left(\Phi(z_2) - \frac{1}{2}\right) \geq 0\} - \frac{1}{2} = -\mathbf{1}\{-z_1 + \alpha\left(\Phi(-z_2) - \frac{1}{2}\right) \geq 0\} + \frac{1}{2}.$$

Then we can consider the Fourier transform

$$\int e^{-it_1y_1 - it_2y_2}\left(\mathbf{1}\{y_1 + \alpha\Phi(y_2) \geq 0\} - \frac{1}{2}\right) dy_1\, dy_2$$
$$= e^{-\frac{1}{2}it_1\alpha} \int e^{-it_1z_1 - it_2z_2}\left(\mathbf{1}\{z_1 + \alpha\left(\Phi(z_2) - \frac{1}{2}\right) \geq 0\} - \frac{1}{2}\right) dz_1\, dz_2.$$

Provided that the transformed function is odd, integration over z_1 leads to

$$\int e^{-it_1z_1 - it_2z_2}\left(\mathbf{1}\{z_1 + \alpha\left(\Phi(z_2) - \frac{1}{2}\right) \geq 0\} - \frac{1}{2}\right) dz_1\, dz_2$$
$$= \frac{1}{it_1} \int e^{it_1\alpha\left(\Phi(z_2) - \frac{1}{2}\right) - it_2z_2}\, dz_2. \tag{E.3}$$

The generalized function $\Gamma(t_1, t_2; \alpha) = \int_{-\infty}^{+\infty} e^{it_1\alpha\left(\Phi(z) - \frac{1}{2}\right) - it_2z}\, dz$ belongs to the class of tempered distributions (see Vladimirov (1971)) and the integral does not exist in the regular sense. This generalized function is not separable into singular and regular components. In fact, we note that if $\Phi(\cdot)$ approaches to a uniform distribution on $\left[-\frac{1}{2}, \frac{1}{2}\right]$, then $\Gamma(t_1, t_2; \alpha)$ behaves like $\delta(t_2 - \alpha t_1)$. When $\Phi(\cdot)$ approaches to a degenerate distribution with point mass at the origin, then $\Gamma(t_1, t_2; \alpha)$ approaches $\frac{2\sin\left(\frac{\alpha}{2}t_1\right)}{t_2} + 2\cos\left(\frac{\alpha}{2}t_1\right)\delta(t_2)$. We note that

$$\lim_{t_1 \to 0} \frac{1}{it_1}\Gamma(t_1, t_2; \alpha) = \alpha \int \left(\Phi(z) - \frac{1}{2}\right) e^{-it_2z}\, dz = \alpha\frac{\chi_\Phi(t_2)}{it_2}. \tag{E.4}$$

The last useful property of this generalized function is that

$$\int t_2 \Gamma(t_1, t_2; \alpha) \, dt_2 = 2\pi i \int e^{it_1 \alpha \left(\Phi(z) - \frac{1}{2}\right)} \delta'(z) \, dz = -2\pi t_1 \alpha \phi(0), \tag{E.5}$$

where $\delta'(\cdot)$ is the appropriately defined derivative of the δ-function. As a result, if we denote $Q(x_1, x) = E\left[Y_1 \,|x_1, x\right]$ then its Fourier transformation can be expressed as

$$\mathcal{Q}(t_1, t_2) = \left(\frac{\sigma e^{-\frac{1}{2}it_1\alpha}}{it_1}\Gamma(t_1, \sigma t_2) + 2\pi^2 \delta(t_1)\delta(t_2)\right) \chi_{uv}(t_1, t_2). \tag{E.6}$$

Next, we consider the Fourier transform of the product $\mathbf{1}\{y_1 + \alpha\Phi(y_2) \geq 0\}\Phi(y_2)$. We note that the product $\mathbf{1}\{y_1 + \alpha\Phi(y_2) \geq 0\}\Phi(y_2)$ is a probability distribution and thus we can define the measure $R(\cdot, \cdot)$ on $\mathbb{R}^2$ that for each subset $S = \{(-\infty, y_1] \times (-\infty, y_2]\}$ with $|y_1|, |y_2| < \infty$ can be defined as

$$R(S) = \mathbf{1}\{y_1 + \alpha\Phi(y_2) \geq 0\}\Phi(y_2).$$

Denote as $\chi_R(\cdot)$ the characteristic function of the random variable whose distribution is defined by $R(\cdot)$. If $\Phi(\cdot)$ is defined by a continuous probability distribution with the full support on $\mathbb{R}$, the corresponding characteristic function exists and does not vanish. If $dR(\cdot, \cdot)$ is the Radon-Nykodim density associated with the measure $R(\cdot)$, we can express the Fourier transform

$$\int e^{-it_1 y_1 - it_2 y_2} \mathbf{1}\{y_1 + \alpha\Phi(y_2) \geq 0\}\Phi(y_2) dy_1 \, dy_2$$

$$= \int e^{-it_1 y_1 - it_2 y_2} \mathbf{1}\{z_1 \leq y_1\}\mathbf{1}\{z_2 \leq y_2\} dR(z_1, z_2) = -\frac{\chi_R(t_1, t_2)}{t_1 t_2}(1 + \pi i t_1 \delta(t_1))(1 + \pi i t_2 \delta(t_2)).$$

Considering the expectation $P_{11}(x_1, x) = E\left[Y_1 Y_2 \,|x_1, x\right]$ with the corresponding Fourier transform $\mathcal{F}_{11}(t_1, t_2)$, we can express

$$\mathcal{F}_{11}(t_1, t_2) = -\frac{\chi_R(t_1, \sigma t_2)}{t_1 t_2}(1 + \pi i t_1 \delta(t_1))(1 + \pi i \sigma t_2 \delta(t_2))\chi_{uv}(t_1, t_2). \tag{E.7}$$

Finally, we consider expectation $P(x) = E\left[Y_2 | x_1, x\right]$, and using the results derived above, we can express its Fourier transform $\mathcal{F}(t_2)$ as

$$\mathcal{F}(t_2) = \left(\frac{\chi_\Phi(\sigma t_2)}{it_2} + \pi\delta(t_2)\right)\chi_v(t_2). \tag{E.8}$$

This concludes our description of the Fourier transformations of the system of identifying equations for the triangular model with incomplete information.

We also consider a more complicated case of the game of incomplete information. In this case, we need to use equilibrium belief functions $P_1(\cdot)$ and $P_2(\cdot)$ to smooth the distribution of unobserved heterogeneity. As we have seen in the previous discussion, symmetry of the smoothing functions is an important feature for deriving the closed-form expression for the Fourier transform. We note that due to the symmetry of the distribution of the experimental noise, $\Phi(0) = \frac{1}{2}$. Then we can

make a transformation of the equilibrium beliefs $P_i(q_1, q_2) = \frac{1}{2} - \Delta_i(q_1, q_2)$, for $i = 1, 2$ where we define functions $\Delta_i(q_1, q_2)$ as solutions of the system of equations

$$\Delta_1(q_1, q_2) = \Phi\left(\frac{1}{\sigma}\left[(q_1 + \frac{\alpha_1}{2}) + \alpha_1\Delta_2(q_1, q_2)\right]\right) - \frac{1}{2},$$

$$\Delta_2(q_1, q_2) = \Phi\left(\frac{1}{\sigma}\left[(q_2 + \frac{\alpha_2}{2}) + \alpha_2\Delta_1(q_1, q_2)\right]\right) - \frac{1}{2},$$

We note that functions $\Delta_i(\cdot)$ are centrally symmetric around the point $(-\frac{1}{2}\alpha_1, -\frac{1}{2}\alpha_2)$. In fact, if $t_i = q_i + \frac{\alpha_i}{2}$, then the system above can be re-written as

$$\Delta_1(t_1, t_2) = \Phi\left(\frac{1}{\sigma}\left[t_1 + \alpha_1\Delta_2(t_1, t_2)\right]\right) - \frac{1}{2},$$

$$\Delta_2(t_1, t_2) = \Phi\left(\frac{1}{\sigma}\left[t_2 + \alpha_2\Delta_1(t_1, t_2)\right]\right) - \frac{1}{2},$$

and if for some (t_1, t_2) the pair (Δ_1, Δ_2) solves this system, then given that $\Phi(\cdot) - \frac{1}{2}$ is an odd function, $(-\Delta_1, -\Delta_2)$ will be a solution for the pair $(-t_1, -t_2)$. We consider recovering the density of the distribution of unobserved heterogeneity from the expectation $E[Y_1Y_2 \,|\, x_1, x_2]$. In particular, we can write

$$P_{11}(x_1, x_2) = E[Y_1Y_2 \,|\, x_1, x_2] = \int P_1(x_1 - u, x_2 - v)P_2(x_1 - u, x_2 - v)g(u, v)\, du\, dv.$$

We will use the deconvolution technique to recover the density $g(\cdot)$ for which we need to find an expression for the Fourier transform of the product of equilibrium beliefs:

$$\mathcal{M}_{12}(t_1, t_2) = \int e^{-it_1q_1 - it_2q_2} P_1(q_1, q_2)P_2(q_1, q_2)\, dq_1\, dq_2$$

Denoting the Fourier transforms of individual beliefs by

$$\mathcal{M}_1(t_1, t_2) = \int e^{-it_1q_1 - it_2q_2} P_1(q_1, q_2)\, dq_1\, dq_2,$$

$$\mathcal{M}_2(t_1, t_2) = \int e^{-it_1q_1 - it_2q_2} P_2(q_1, q_2)\, dq_1\, dq_2,$$

we can construct the closed-form expression for the Fourier transform of the product in the following way. First, denote $p_i(\cdot, \cdot)$ a bivariate density associated with the measure $P_i(\cdot, \cdot)$ defined by the equilibrium beliefs. Observe that P_i is a probability measure with $P_i(\mathbb{R}^2) = 1$ and has an absolutely continuous density (by differentiability of $\Phi(\cdot)$). Then

$$\int e^{-it_1q_1 - it_2q_2} P_1(q_1, q_2), dq_1\, dq_2$$

$$= \int e^{-it_1q_1 - it_2q_2} \mathbf{1}\{z_1 - q_1 \leq 0\}\mathbf{1}\{z_2 - q_2 \leq 0\}p_1(z_1, z_2)dz_1\, dz_2\, dq_1\, dq_2$$

$$= -\frac{\chi_{P_1}(t_1, t_2)}{t_1 t_2}\left(1 + \pi it_1\delta(t_1)\right)\left(1 + \pi it_2\delta(t_2)\right).$$

Similarly, we can define the bivariate density associated with the measure $P_1(\cdot,\cdot)P_2(\cdot,\cdot)$ and express

$$\int e^{-it_1q_1-it_2q_2}P_1(q_1,q_2)P_2(q_1,q_2), dq_1\,dq_2$$
$$= -\frac{\chi_{P_1P_2}(t_1,t_2)}{t_1t_2}\left(1+\pi it_1\delta(t_1)\right)\left(1+\pi it_2\delta(t_2)\right).$$

We then consider the Fourier transform of the expectation $E\left[Y_1Y_2\,|\,x_1,x\right]$, which we denote $\mathcal{F}_{11}(t_1,t_2)$, leading to the following expression for the density of the distribution of unobserved heterogeneity:

$$g(u,v) = -\frac{1}{(2\pi)^2}\int e^{it_1u+it_2v}\frac{t_1t_2\mathcal{F}_{11}(t_1,t_2)}{\chi_{P_1P_2}(t_1,t_2)}\,dt_1\,dt_2.$$

F Semiparametric efficiency bounds in incomplete information models

F.1 Semiparametric efficiency bound in the triangular model with incomplete information

The semiparametric efficiency bound provides the minimum variance for the finite-dimensional parameters over admissible sets of non-parametric components of the model. In our case, it will reflect the minimum variance of the strategic interaction parameter. To find the semiparametric efficiency bound, we use the result in Ai and Chen (2003). We note that the model is represented by a system of semiparametric conditional moment equations:

$$
\begin{aligned}
P_{11}(x_1,x) &= E\left[y_1y_2\,|\,x_1,x\right] \\
&= \int \mathbf{1}\{x_1-u+\alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\}\Phi\left(\frac{x-v}{\sigma}\right)g(u,v)\,du\,dv, \\
P(x_1,x) &= E\left[y_2\,|\,x_1,x\right] = \int \Phi\left(\frac{x-v}{\sigma}\right)g_v(v)\,dv, \\
Q(x_1,x) &= E\left[y_1\,|\,x_1,x\right] = \int \mathbf{1}\{x_1-u+\alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\}g(u,v)\,du\,dv.
\end{aligned}
$$

(F.9)

These equations fully characterize the conditional distribution of the outcome variables, provided that the outcome variables are binary. Due to independence of the distribution of errors (U,V) and covariates (X_1,X), the distribution of covariates does not provide any information regarding the

strategic interaction parameter. We can re-write this system of equations in an equivalent form as

$$m_1(x_1, x; \alpha, g) = E\left[y_1 y_2 - \int \mathbf{1}\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\}\right.$$

$$\left.\Phi\left(\frac{x-v}{\sigma}\right) g(u,v)\, du\, dv \middle| x_1, x\right] = E[\rho_1(y, x; \alpha, g)|x_1, x],$$

$$m_2(x_1, x; \alpha, g) = E\left[y_1 - \int \mathbf{1}\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\}g(u,v)\, du\, dv\right] \tag{F.10}$$

$$= E[\rho_2(y, x; \alpha, g)|x_1, x],$$

$$m_3(x_1, x; \alpha, g) = E\left[y_2 - \int \Phi\left(\frac{x-v}{\sigma}\right) g_v(v)\, dv \middle| x_1, x\right] = E[\rho_3(y, x; \alpha, g)|x_1, x],$$

Consider the derivatives of these moment equations with respect to parameter α:

$$\frac{dm_1}{d\alpha} = -\int \Phi\left(\frac{x-v}{\sigma}\right)^2 \frac{\partial}{\partial v} G\left(x_1 + \alpha\Phi\left(\frac{x-v}{\sigma}\right), v\right) du\, dv,$$

$$\frac{dm_2}{d\alpha} = -\int \Phi\left(\frac{x-v}{\sigma}\right) \frac{\partial}{\partial v} G\left(x_1 + \alpha\Phi\left(\frac{x-v}{\sigma}\right), v\right) du\, dv,$$

$$\frac{dm_3}{d\alpha} = 0.$$

Then considering the space of densities that are *uniformly manageable* and satisfy Assumption 1, we take a direction in this space h and

$$\frac{dm_1}{dg}[h] = -\int \mathbf{1}\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\}\Phi\left(\frac{x-v}{\sigma}\right) h(u,v)\, du\, dv,$$

$$\frac{dm_2}{dg}[h] = -\int \mathbf{1}\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\}h(u,v)\, du\, dv,$$

$$\frac{dm_3}{dg}[h] = -\int \Phi\left(\frac{x-v}{\sigma}\right) h_v(v)\, dv.$$

We introduce the vector with elements

$$\psi_1(x_1, x, u, v) = \mathbf{1}\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\}\left(\Phi\left(\frac{x-v}{\sigma}\right) g(u,v) - h(u,v)\right),$$

$$\psi_2(x_1, x, u, v) = \mathbf{1}\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\}\left(g(u,v) - h(u,v)\right),$$

$$\psi_3(x_1, x, u, v) = -h(u,v),$$

and denote

$$\zeta_1(x_1, x, u, v) = \mathbf{1}\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\},$$

$$\zeta_2(x_1, x, u, v) = \mathbf{1}\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\},$$

$$\zeta_3(x_1, x, u, v) = 1,$$

and

$$\xi_1(x_1, x, u, v) = \mathbf{1}\{x_1 - u + \alpha\Phi\left(\frac{x - v}{\sigma}\right) > 0\}\Phi\left(\frac{x - v}{\sigma}\right),$$

$$\xi_2(x_1, x, u, v) = \mathbf{1}\{x_1 - u + \alpha\Phi\left(\frac{x - v}{\sigma}\right) > 0\},$$

$$\xi_3(x_1, x, u, v) = 1.$$

We express

$$D_h(x_1, x) = \frac{dm}{d\alpha} - \frac{dm}{dg}[h] = \int \Phi\left(\frac{x - v}{\sigma}\right)\psi(x_1, x, u, v)\,du\,dv,$$

which is a linear functional of $h(\cdot, \cdot)$, in fact

$$D_h(x_1, x) = \int \Phi\left(\frac{x - v}{\sigma}\right)\xi(x_1, x, u, v)g(u, v)\,du\,dv - \int \Phi\left(\frac{x - v}{\sigma}\right)\zeta(x_1, x, u, v)h(u, v)\,du\,dv.$$

Next, we find the conditional covariance matrix

$$\Sigma(x_1, x) = P_{11}\left(I - \begin{pmatrix} P_{11} & Q & P \\ Q & 1 - \frac{Q(1-Q)}{P_{11}} & \frac{PQ}{P_{11}} \\ P & \frac{PQ}{P_{11}} & 1 - \frac{P(1-P)}{P_{11}} \end{pmatrix}\right) \tag{F.11}$$

The semiparametric efficiency bound will be associated with the "least favorable" direction h. To find this direction one needs to solve the minimization problem

$$\min_{h \in \mathcal{G} - g_0} E\left[D_h(X_1, X)'\,\Sigma^{-1}(X_1, X)\,D_h(X_1, X)\right].$$

It is convenient to define the least favorable direction as $h = q^2$ to ensure that the solution is positive and also require that $\int q^2(u, v)\,du\,dv = 1$. Then the minimization problem becomes a constrained optimization problem. We have previously noted that $D_h(X_1, X)$ is a linear functional. We can thus find the minimum using the standard calculus of variation for a constrained isoperimetric problem. The considered minimized functional is quadratic and we can express the necessary condition for its minimum as

$$E\left[\Phi\left(\frac{X - v}{\sigma}\right)\zeta(X_1, X, u, v)'\,\Sigma^{-1}(X_1, X)\,D_{h^*}(X_1, X)\right] + \lambda = 0,$$

where λ is the Lagrange multiplier and $h^* = q^{*2}$ corresponds to the optimal solution. Finally, we can transform this equation by isolating the terms for h^* and g and introducing notations

$$K(u, v, u', v') = E\left[\Phi\left(\frac{X - v}{\sigma}\right)\Phi\left(\frac{X - v'}{\sigma}\right)\zeta(X_1, X, u, v)'\,\Sigma^{-1}(X_1, X)\,\zeta(X_1, X, u', v')\right]$$

and

$$R(u, v, u', v') = E\left[\Phi\left(\frac{X - v}{\sigma}\right)\Phi\left(\frac{X - v'}{\sigma}\right)\zeta(X_1, X, u, v)'\,\Sigma^{-1}(X_1, X)\,\xi(X_1, X, u', v')\right].$$

Thus,

$$\int K(u,v,u',v')h^*(u',v')\,du'dv' = \lambda + \int R(u,v,u',v')g(u',v')\,du'dv'$$

Given that $K(u,v,u',v')$ is a non-separable symmetric kernel. Thus it has an infinite coutable set of eigenfunctions with real eigenvalues. Moreover, provided that $K(u,v,u',v')$ is strictly positive and decays with $|u|,|v| \to \infty$, it satisfies the Picard criterion. Therefore, the Fredholm integral equation above has a solution. Ths solution to this equation that is strictly positive and normalizes to 1 yields the semiparametric efficiency bound

$$\Omega = E\left[D_{h^*}(X_1,X)'\,\Sigma^{-1}(X_1,X)\,D_{h^*}(X_1,X)\right]$$

F.2 Semiparametric efficiency bound in the static game model with incomplete information

We note that the observed equilibrium responses are characterized by two binary variables, Y_1 and Y_2. Given the independence of the unobserved heterogeneity (U,V) and the covariates and the fact that the distribution of covariates does not depend on the parameters of interest, the conditional distribution of observed actions is fully characterized by three expectations: $E[Y_1|x_1,x_2]$, $E[Y_2|x_1,x_2]$, and $E[Y_1Y_2|x_1,x_2]$. These expectations characterize the conditional moments that identify the strategic interaction parameters:

$$P_{11}(x_1,x_2) = E[Y_1Y_2|x_1,x_2] = \int P_1(x_1-u,x_2-v)P_2(x_1-u,x_2-v)g(u,v)\,du\,dv,$$

$$Q(x_1,x_2) = E[Y_1|x_1,x_2] = \int P_1(x_1-u,x_2-v)g(u,v)\,du\,dv,$$

$$P(x_1,x_2) = E[Y_2|x_1,x_2] = \int P_2(x_1-u,x_2-v)g(u,v)\,du\,dv,$$

We can re-write this system of equations in an equivalent form as

$$m_1(x_1,x_2;\alpha,g) = E\left[Y_1Y_2 - \int P_1(X_1-u,X_2-v)P_2(X_1-u,X_2-v)\right.$$
$$\left. \times g(u,v)\,du\,dv \,\Big|\, x_1,x_2\right] = E[\rho_1(Y,X;\alpha,g)|x_1,x],$$

$$m_2(x_1,x_2;\alpha,g) = E\left[Y_1 - \int P_1(X_1-u,X_2-v)g(u,v)\,du\,dv\right] \tag{F.12}$$
$$= E[\rho_2(Y,X;\alpha,g)|x_1,x_2],$$

$$m_3(x_1,x_2;\alpha,g) = E\left[Y_2 - \int P_2(X_1-u,X_2-v)g(u,v)\,du\,dv \,\Big|\, x_1,x_2\right]$$
$$= E[\rho_3(Y,X;\alpha,g)|x_1,x],$$

Under our assumption regarding the distribution of errors η_1 and η_2, equilibrium beliefs are monotone functions of the parameters. Previously, we derived the Jacobi matrix corresponding to the

derivatives of the equilibrium beliefs with respect to the parameters as:

$$J^\alpha = \begin{pmatrix} \frac{\partial P_1}{\partial \alpha_1} & \frac{\partial P_1}{\partial \alpha_2} \\ \frac{\partial P_2}{\partial \alpha_1} & \frac{\partial P_2}{\partial \alpha_2} \end{pmatrix} = \frac{a_1 a_2}{1 + \alpha_1 \alpha_2 a_1 a_2} \begin{pmatrix} P_1/a_2 & \alpha_1 P_1 \\ \alpha_2 P_2 & P_2/a_1 \end{pmatrix},$$

where $a_i = \sigma^{-1} \phi \left(\Phi^{-1}(P_i) \right)$.

We can express the Jacobi matrix of the moment vector $m(\cdot)$ with respect to the finite-dimensional parameters α_1 and α_2 as:

$$\frac{dm(x_1, x_2; \alpha, g)}{d\alpha'} = \int M(x_1 - u, \, x_2 - v) \, J^\alpha(x_1 - u, \, x_2 - v) g(u, v) \, du \, dv,$$

where

$$M = \begin{pmatrix} P_2 & P_1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = (\mu_1, \mu_2).$$

Then considering the space of densities that are *uniformly manageable* and satisfy Assumption 4, we take a direction in this space h and obtain

$$\frac{dm(x_1, x_2; \alpha, g)}{dg}[h] = \int \psi(x_1 - u, \, x_2 - v) h(u, v) \, du \, dv,$$

where $\psi(q_1, q_2) = (P_1(q_1, q_2) P_1(q_1, q_2), \, P_1(q_1, q_2), \, P_2(q_1, q_2))'$. The semiparametric efficiency bound will be associated with a vector of two least favorable directions h_1^* and h_2^* such that h_i^* minimizes

$$E \left[D_{h_i}(X_1, X_2) \, \Sigma(X_1, X_2)^{-1} D_{h_i}(X_1, X_2) \right],$$

where $D_{h_i}(x_1, x_2) = \frac{dm(x_1, x_2; \alpha, g)}{d\alpha_i} - \frac{dm(x_1, x_2; \alpha, g)}{dg}[h_i]$ and $\Sigma(\cdot, \cdot)$ is determined by (F.11). We note that $D_{h_i}(x_1, x_2)$ is linear in h_i. We can minimize the considered objective function under the constraint that the solution has to be a density function. This optimization leads us to the expression

$$E \left[\psi(X_1 - u, \, X_2 - v)' \, \Sigma(X_1, X_2)^{-1} D_{h_i}(X_1, X_2) \right] + \lambda = 0,$$

where λ is the Lagrange multiplier. We introduce notation

$$K(u, v, u', v') = E \left[\psi(X_1 - u, \, X_2 - v)' \, \Sigma(X_1, X_2)^{-1} \psi(X_1 - u', \, X_2 - v') \right]$$

and

$$R_i(u, v, u', v') = E \Bigg[\psi(X_1 - u, \, X_2 - v)' \, \Sigma(X_1, X_2)^{-1}$$
$$\mu_i(X_1 - u', \, X_2 - v') J^\alpha(X_1 - u', \, X_2 - v') \Bigg].$$

Then we can find the least favorable direction for $i = 1, 2$ as a solution to

$$\int K(u, v, u', v') \, h_i^*(u', v') \, du' \, dv' = \lambda + \int R(u, v, u', v') \, g(u', v') \, du' \, dv'.$$

The kernel function $K(u, v, u', v')$ is positive, symmetric, non-separable, and square-integrable. Thus, the Hilbert space $\mathcal{G}$ has an orthonormal basis consisting of the eigenvectors of the integral operator with the kernel $K(u, v, u', v')$, and the solution for h_i^* will be in this basis.

The semiparametric efficiency bound will then be constructed from

$$D_{h^*}(x_1, x_2) = \left(D_{h_1^*}(x_1, x_2),\ D_{h_2^*}(x_1, x_2) \right)'.$$

We can express the bound as

$$\Omega = E \left[D_{h^*}(X_1, X_2)\, \Sigma(X_1, X_2)^{-1} D_{h^*}(X_1, X_2) \right]^{-1}.$$