Abstract

Strategic interaction parameters characterize the impact of actions of one economic agent on the payoff of another economic agent, and are of great interest in both theoretical and empirical work. In this paper, by considering econometric models involving simultaneous discrete systems of equations, we study how the information available to economic agents regarding other economic agents can influence the identifiability of these strategic information parameters from the observed actions. We consider two extreme cases: the complete information case where the information sets of participating economic agents coincide, and the incomplete information case where each agent’s actions are privately observable. We find that in models with complete information, the strategic interaction parameters are more difficult to identify than they are in incomplete information models. We show this by exploring the Fisher information (from standard statistics literature) for the strategic interaction parameters in each of these models. Our findings are that in complete information models, the statistical (Fisher) information for the interaction parameters is zero, implying the difficulty in identifying them from observed data. In contrast, for incomplete information models, the Fisher information for the interaction parameters is positive, indicating that not only can these parameters be identified from observed data, but standard inference can be conducted on them. This finding is illustrated in two cases: treatment effect models (expressed as a triangular system of equations), and static game models.

JEL Classification: C35, C14, C25, C13.

Keywords: Endogenous discrete response, treatment effects, 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. However, correcting for endogeneity can be particularly difficult in nonlinear models. Recent important work in econometrics has studied the identification and estimation of some nonlinear models with endogenous regressors by adopting what is referred to as a “control function” approach. Seminal papers include Newey, Powell, and Vella (1999), Blundell and Powell (2004), Imbens and Newey (2009). Their control function approach, however, requires the endogenous regressors to be continuously distributed.

In this paper the class of models we consider will focus on simultaneous discrete response models with discrete endogenous variables. This class includes many important special cases that have received a great deal of attention in both theoretical and empirical work. Examples include strategic compliance models, models of social interactions, and the simultaneous move discrete game model. For these models we are specifically interested in identifiability of the coefficients of the discrete endogenous variable(s).

We first look at the case of the triangular system of equations, where by triangular we mean that the binary outcome in one equation is an explanatory variable in the other equation. Here the parameter of interest is the coefficient on that binary endogenous variable and is directly related to the treatment effect in that literature. In the other class of models we study, a system of simultaneous discrete equations with feedback effects, each of the two binary outcomes in the two equations is an explanatory variable in the other equation, such as those that appear in social interaction and static game models, and the parameter of interest characterizes the degree of strategic interaction.

We will be primarily interested in quantifying and comparing the identifiability of these parameters of interest. Our approach will be to do so by analyzing what is referred to in classical statistics as the Fisher information of these parameters. The Fisher information can be considered an important indicator of the “quantity” of identification, and the relationship between the two concepts of identification and information dates back to the seminal paper in Rothenberg (1971), who considered identification in parametric models.

The aim of this paper will be to relate the quantity of identification of the parameters of interest to the observable information available to the economic agents in the model. For both the triangular system and the simultaneous system, we consider first what we

---

1 Identification and inference in these models is more complicated than in the continuous endogenous variable case, as exhibited in the important work in Cheshire (2005), who considers partial identification of general classes of nonlinear, nonseparable models, and derives sharp identification regions for parameters of interest. See also Klein, Shan, and Vella (2011) for estimation in this area.
call a complete information model where the agents have perfect knowledge regarding all the observed variables, or more generally speaking the information sets of participating economic agents coincide. Then we consider an incomplete information model, where each agent’s own actions are only privately observable.

Our main finding will be that, generally speaking, incomplete information models have more identifying power for the parameters of interest than complete information models do. We reach this conclusion by systematically evaluating the Fisher information for the parameters of interest in each of the four cases, corresponding to each of the two types of information available in each of the two types of systems of equations.

The rest of the paper is organized as follows. In the following section we introduce what we will call a triangular system of equations with complete information. This will be a basic binary choice model with a binary endogenous variable determined by a reduced-form model. This model has often been seen in the treatment effects literature, and we will quantify the identifying power for the coefficient on the endogenous variable, by deriving the Fisher information for that parameter.

In Section 3 we then consider the triangular system with incomplete information, where the agent has uncertainty of the value of the endogenous variable affecting his/her decision. The parameter of interest is the coefficient of on that variable. A treatment effect interpretation of this model could be that the recipient of treatment is uncertain about the type of treatment being offered. As we explain, this model has a loose analogy to the introduction of a "placebo" in a controlled experiment. That interpretation intuitively suggests increased identifying power and we confirm this by evaluating the Fisher information and comparing it what we found in the complete information case.

In Section 4 we explore models of simultaneous discrete systems, using the example of a two-player simultaneous move game with particular equilibrium selection rule. Here the parameters of interest are the two interaction parameters in this game, and we explore the informational content for these parameters by evaluating the Fisher information.

In Section 5 we consider a game of incomplete information in which each player has uncertainty of the binary decision of the other player, making this endogenous variable incompletely, or noisily observed. Of interest here is the identifying power of the uncertainty on the parameter of interest, in this case the coefficient on that incompletely observed endogenous variable. Analogous to what we find in Section 3, the additional layer of uncertainty increases identifying power for the parameter of interest, as measured by Fisher information being higher.

Finally, Section 6 concludes the paper by summarizing and suggesting areas for future research. An appendix collects the proofs of the main theorems.
2 Discrete response model

2.1 Information in discrete response model

In this section we will examine the information for the interaction parameter in a triangular system of binary equations. 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 $Y_2$.

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

$$Y_1 = 1\{Z_1'\gamma_0 + \alpha_0 Y_2 - U > 0\}. \quad (2.1)$$

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:

$$Y_2 = 1\{Z_2'\delta_0 - V > 0\}, \quad (2.2)$$

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.

This type of model fits into the class of models considered in Vytlacil and Yildiz (2007). See also Klein, Shan, and Vella (2011). In this section of the paper we are interested in the parameter $\alpha_0$, which is related to a treatment effect. To simplify exposition, we will assume the parameters $\delta_0$ and $\gamma_0$ are known. What this part of the paper will focus on is the information for $\alpha_0$ (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 $X = Z_2'\delta_0$.

The discrete response model can then be written as

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

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) The 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;

Although additional complications may arise if covariates $Z$ are not truly exogenous, under our assumptions one can regularly identify parameters $\gamma_0$ and $\delta_0$. The issues of inference for these parameters are discussed, for instance in Abrevaya, Hausman, and Khan (2011).
(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 $G(\cdot, \cdot)$;

(iii) For each $t \in \mathbb{R}$ and fixed $\gamma_0$ and $\delta_0$, 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 some of these assumptions are stronger than assumptions that are often 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 $\alpha_0$ 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 $\alpha_0$ 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 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 formally state this result in the following theorem:

**Theorem 2.1** Under Assumption 1, the Fisher information associated with parameter $\alpha_0$ in model (2.3) is zero.

We find that under our conditions the parameter $\alpha_0$ cannot be estimated at the parametric rate. Even though this result has similar conclusions to the impossibility theorems in Chamberlain (1986) and Cosslett (1987), who studied selection and binary choice models, respectively, the issue of zero information of payoff parameters in strategic models was not well understood so far. The conditions of the theorem imply that for 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.

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).

---

3 The proof of this and all subsequent theorems is provided in Appendix A.
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 $\alpha_0$. Thus these results can be regarded as analogous to those found in Manski (1985) and Horowitz (1992) for a binary choice model under conditional median restrictions, where best achievable rates were found to be slower than parametric.

We take a constructive approach to establish the optimal convergence rate for the estimator for $\alpha_0$. We begin with a definition of the optimal rate following Ibragimov and Has’minskii (1978). Let $G$ characterize a class of joint densities of error terms $(U,V)$ (denoted $g(\cdot,\cdot)$) and single indices $X_1$ and $X$ (with density function $f(\cdot,\cdot)$). By $P_{f,g}$ we denote the probability measure associated with the product of two densities $f$ and $g$. Suppose that $\hat{\alpha}$ is a consistent estimator for parameter $\alpha_0$. First, we recall that for the class of distributions $G$, we define the risk using a positive (rate) sequence $r_n$ and a constant $L$ as

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

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 $G$ if there exists $L > 0$ such that

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

for some constant $p_0$.

(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.\footnote{We give concrete examples of such distributions including normal and logistic in Appendix C}

**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 $\nu(\cdot)$ such that for any $|t| < \infty$

$$0 < \lim_{c \to +\infty} \frac{1}{\nu(c)} \sup_{f,g \in \mathcal{G}} E_{f,g} \left[ \left( \frac{\partial G(X_1 + t, X)}{\partial t} \right)^2 G(X_1 + t, X)^{-1} (G_v(X) - G(X_1 + t, X))^{-1} \right] < \infty$$

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

$$0 < \lim_{c \to +\infty} \beta(c) \sup_{f,g \in \mathcal{G}} E_{f,g} \left[ \log \left( G(X_1 + t, X) \right) (G_v(X) - G(X_1 + t, X)) \right]^{-1} < \infty$$

where $E_{f,g}$ denotes the expectation operator with respect to densities $f, g$.

This assumption can alternatively be considered a definition for functions $\nu(\cdot)$ and $\beta(\cdot)$. The value of function $\nu(c)$ corresponds to the variance of the score of the semiparametric likelihood of the triangular model with respect to parameter $\alpha$ evaluated on a bounded subset of the support of $(X_1, X)$ (with diameter $c$). The function $\nu(\cdot)$ then characterizes the rate at which the variance diverges to infinity as the score is integrated over larger and larger sets. Function $\beta(\cdot)$ corresponds to the bias in the expected value of the semiparametric likelihood function that arises when the expectation of the likelihood is taken over trimmed values of $(X_1, X)$.

The following theorem outlines our main result regarding the optimal rate for the interaction parameter in the triangular model.

**Theorem 2.2** Suppose that Assumptions\footnote{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.} and\footnote{We give concrete examples of such distributions including normal and logistic in Appendix C} 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 $\alpha_0$ in model (2.3).
Remark 2.1 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 term $\nu(\cdot)$ 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.

Finally we note that in this section and in the paper we primarily omit the discussion of estimation procedures that may achieve these optimal rates under stated conditions. The main purpose of this paper is to consider models which can and cannot be estimated at standard rates. For work on estimation and inference procedures in discrete triangular systems, see for example Klein, Shan, and Vella (2011).

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 $\alpha_0$. 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 consider a new model which can be arbitrarily “close” to the classical triangular model but have positive information. 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

\footnote{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.}
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

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

We express our assumption regarding the additional noise component $\eta$ formally:

**Assumption 3** Suppose that $\eta \perp (U, V)$ and $\eta \perp (X_1, X)$. The distribution of $\eta$ 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.

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

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

where the parameter of interest is $\alpha_0$ for which we wish to derive the information.\footnote{We show further in the paper that this choice is a normalization.} 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 be written as

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

$$Y_2 = 1\{X - V + \sigma \eta > 0\}. \tag{3.1}$$

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. \footnote{We note that here $\alpha_0$ 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 $\sigma$) is not small, the new parameter will have the same sign as the old one.}
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 Vylacil (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.

In the practical settings, the model that we analyze is closely related to the setting of A/B testing that is used for experimentation by large Internet companies like Google or Microsoft on their online advertising platforms. In each experiment that companies measure the response of advertisers to changes in the advertising platform such as pricing, or the rules for allocating online ads. Then the ads of each advertiser with a very small probability (usually below 1-5%) will be exposed to the new platform settings and otherwise it will be exposed to the status quo setting. Then the platform measures the advertiser’s response of advertisers in this experiment.

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, Snowberg, et al. (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 the figure on panel (a).

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 \( Y_2 \)).

We note that in the case of the incomplete information triangular model identification includes both parameter \( \alpha_0 \) and the distribution of error terms \( U \) and \( V \). Provided that identification of such a model was previously established (with the exception of (Heckman and Navarro-Lozano 2004) where the focus is on the treatment effect parameter and not on the distributional components), we will need to establish first that the model is identified from the data.

**Theorem 3.1** Under Assumptions [1] and [3], the interaction parameter \( \alpha_0 \) and the distribution of \((U, V)\) in model (3.1) is identified.

We note that parameter \( \alpha_0 \) is identified under conditions that are weaker than those that we needed to establish the optimal rate of convergence in the complete information triangular model (2.3). The proof of identification for parameter \( \alpha_0 \) is based on the idea to relate the expectations of \( Y_1 \) to the expectation of \( Y_2 \). First of all note that

\[
\int \frac{\partial}{\partial x} E[Y_1 | X_1 = x_1, X = x] \, dx_1 = \alpha_0 \int \int g(x_1 + \alpha_0 \Phi\left( \frac{x-v}{\sigma}\right), v) \frac{1}{\sigma} \phi\left( \frac{x-v}{\sigma}\right) \, dv \, dx_1
\]

which identifies the interaction parameter. Then, for sufficiently small variance of the noise \( \sigma^2 \) we note that

\[
\alpha_0 = \int \frac{\partial}{\partial x} E[Y_1 | X_1 = x_1, X = x] \, dx_1 \frac{\partial}{\partial x} E[Y_2 | X = x],
\]

which identifies the interaction parameter. Then, for sufficiently small variance of the noise \( \sigma^2 \) we note that

\[
E[Y_1 Y_2 | X_1 = x_1, X = x] \approx G(x_1 + \alpha_0, x),
\]

which allows us to identify the joint distribution of the error terms \((U, V)\). We provide a more careful argument for identification of the distribution of error terms based on the deconvolution technique in Appendix [E].

Having established the identification of the parameter of interest, we analyze its Fisher information. We find that for any finite variance \( \sigma^2 \) (which can be arbitrarily small) the information for \( \alpha_0 \) in the incomplete information triangular model is strictly positive. Moreover, the Fisher information of the strategic interaction parameter \( \alpha_0 \) 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 \( \alpha_0 \)._
Theorem 3.2  Suppose that Assumptions [7] 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.

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 $\alpha_0$, 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), which also suggests an efficient method of moments estimator. Our efficiency result provides the semiparametric efficiency bound for the new discrete response model.

Our final result expresses 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 ($\sqrt{n}$) 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:

Theorem 3.3 Under Assumptions [7] and [3] letting $\hat{\alpha}_{0,n}$ denote the efficient method of moments estimator of the interaction parameter,

\[ \sqrt{n}(\hat{\alpha}_{0,n} - \alpha_0) \Rightarrow N(0, \Omega) \quad (3.2) \]

where $\Omega$ is the semiparametric efficiency bound\footnote{We provide an explicit expression for $\Omega$ in terms of the primitives of the model in Appendix F.1}.
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:

$$Y_1^* = \gamma_0 + \alpha_1 Y_2 - U,$$

and the payoff of player 2 from choosing action $Y_2 = 1$ is characterized as

$$Y_2^* = \delta_0 + \alpha_2 Y_1 - V.$$

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 is the mapping from the observable variables into actions: $(u,v,x_1,x_2) \rightarrow 0,1$. 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:

$$Y_1 = 1\{X_1 + \alpha_1 Y_2 - U > 0\},$$

$$Y_2 = 1\{X_2 + \alpha_2 Y_1 - V > 0\},$$

where errors $U$ and $V$ are correlated with each other with an unknown distribution\(^{10}\). In

\(^{10}\)We note that in this formulation for the case of the complete information game we only allow the players to use pure strategies. The mixed strategy equilibria can be represented in this model by a distribution over pure strategy equilibria. Our arguments, however, can be extended to allow pure strategy equilibria in model (4.1).
particular, we are interested in determining when the strategic interaction parameters $\alpha_1, \alpha_2$ 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 $G(\cdot, \cdot)$;

(iii) For each $t_1, t_2 \in \mathbb{R}$, there exist functions $q_1(\cdot, \cdot)$ and $q_2(\cdot, \cdot)$ with $E[q_1(X_1, X_2)^2] < \infty$ $E[q_2(X_1, X_2)^2] < \infty$ which dominate $\frac{\partial G(x_1+t_1, x_2+t_2)}{\partial x_1}$ and $\frac{\partial G(x_1+t_1, x_2+t_2)}{\partial x_2}$, 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 *incoherence* in the model. We select this simple setup to emphasize that the complete information model has zero information even when there is no incoherence.

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.

---

11 By incoherence, here we mean the presence of multiple equilibria or non-existence of the pure strategy equilibria.
We now prove identification of strategic interaction parameters, arguing that the zero information result is not a consequence of unidentifiability. 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 but assumes that the distribution of $(U, V)$ is normal) while imposing a linear index structure on the payoffs.\footnote{The proof of identification can be found in the companion paper “Information Bounds and Impossibility Theorems for Simultaneous Discrete Response Models”.
}

**Theorem 4.1** Suppose that Assumptions 4 and 5 are satisfied. Then the interaction parameters $\alpha_1$ and $\alpha_2$ and the distribution of error terms $(U, V)$ in model (4.1) are identified.

The identification for parameters of interest in this case requires a form of the “identification at infinity” arguments:

\[
\lim_{x_2 \to +\infty} E[Y_1|X_1 = x_1, X_2 = x_2] = G_u(x_1 + \alpha_1),
\]

\[
\lim_{x_2 \to -\infty} E[Y_1|X_1 = x_1, X_2 = x_2] = G_u(x_1 + \alpha_1),
\]

which identifies the marginal distribution of $U$ and, therefore, parameter $\alpha_1$. By similar considerations, we can identify the marginal distribution of $V$ and parameter $\alpha_2$. The identification of the joint distribution requires a more careful mathematical argument and we use deconvolution technique to recover it. 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 insight 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 $\alpha_1$ and $\alpha_2$ 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 outcomes 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 the interaction parameters by deriving their optimal convergence rates like we did before for the triangular system.
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 $|s| < \infty$

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

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

$$
\lim_{c \to \infty} \beta(c) \sup_{f,g \in \mathcal{G}} \left( E_{f,g} \left[ \log \left( G(X_1 + t, X_2 + s) \right) \left( 1 - G(X_1 + t, X_2 + s) \right) \right] |X_1, |X_2| > c \right)^{-1} < \infty
$$

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. Consequently, we leave the detailed definitions and conditions to the appendix, only stating the main results here in a theorem.

**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 $\alpha_1$ and $\alpha_2$.

One of the important conclusions 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. 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 \( \eta_1 \) and \( \eta_2 \). 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 \( \eta_1 \) and \( \eta_2 \) 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 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 \( \eta_1 \) and \( \eta_2 \), whose distributions are normalized. An alternative interpretation for this information structure is that the payoff components \( u \) and \( v \) are a priori known to the players but not to the econometrician. The interaction of the players is considered in the experimental settings where the payoff noise \((\eta_1, \eta_2)\) is introduced artificially by the experiment designer. For this reason its distribution is known both to the players and to the econometrician.

\(^{13}\)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).
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 $\eta_1$ and $\eta_2$ 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

$$Y_1 = 1\{E[Y_1^* \mid X_1, X_2, U, V, \eta_1] > 0\}$$

$$= 1\{X_1 - U + \alpha_1 P_2(X_1, X_2, U, V) - \sigma \eta_1 > 0\}.$$  \hspace{1cm} (5.1)

Similarly, the strategy of player 2 can be written as

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

We note the resemblance 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 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:

$$\sigma \Phi^{-1}(P_1) = x_1 - u + \alpha_1 P_2$$

$$\sigma \Phi^{-1}(P_2) = x_2 - v + \alpha_2 P_1.$$  \hspace{1cm} (5.3)

Our informational assumption regarding the independence of the unobserved heterogeneity components $U$ and $V$ from payoff perturbations $\eta_1$ and $\eta_2$ 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.\footnote{Sweeting (2009) considers a $2 \times 2$ game of incomplete information and gives examples of multiple equilibria in that game. Bajari, Hong, Krainer, and Nekipelov (2010a) develop a class of algorithms for efficient computation of all equilibria in incomplete information games with logistically distributed noise components.} 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 \(\sigma\) 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 \(\eta_1\) and \(\eta_2\)). 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.
First, we establish the fact that the strategic interaction parameters $\alpha_1$ and $\alpha_2$ 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 $(\eta_1, \eta_2)$ 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 $E[Y_1Y_2|x_1,x_2]$. 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. Although, with continuous distribution of the noise \( \eta_1 \) and \( \eta_2 \) 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 \( E[Y_1Y_2|x_1, x_2] \) can be expressed as

\[
E[Y_1Y_2|x_1, x_2] = \int \bar{P}_1(x_1 - u, x_2 - v)\bar{P}_2(x_1 - u, x_2 - v)g(u, v) \, du \, dv
\]

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 strategic 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 \( \alpha_1 \) and \( \alpha_2 \). 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 \( \alpha_1 \) and \( \alpha_2 \) 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 \( \sigma^2 \) (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 \( (\alpha_1, \alpha_2) \) 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 \( (\alpha_1, \alpha_2) \) 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 extends the result of Theorem 5.2. This theorem states that the optimal convergence rate for the estimator for the strategic interaction parameters in the incomplete information game is parametric with a limiting normal distribution 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

\[
\sqrt{n}(\hat{\alpha}_n - \alpha_0) \Rightarrow N(0, \Omega) \tag{5.4}
\]

where \(\hat{\alpha}_n\) denotes the efficient method of moments estimator and \(\Omega\) is the semiparametric efficiency bound.\(^{15}\)

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 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

\(^{15}\)We provide an explicit result for both the efficient estimator and the semiparametric efficiency bound in Appendix F.2.
Kosenok (2008) for the games with quantal response equilibria considered in Palfrey (1985), with the further generalization 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, if any advantage to estimating the parameter in this model relative to estimating the less structural model proposed in, for example, 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 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


Appendix

A Proofs

A.1 Proof of Theorem 2.1

To simplify our argument, we assume that coefficients $\beta_0$ and $\delta_0$ 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 the parameter $\alpha$ is equal to zero. Suppose that $\Gamma$ 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 $E[q(X_1, X)^2] < \infty$ for $q(\cdot, \cdot)$ 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:

$$
\begin{align*}
P_{11}(t_1, t) &= \Pr (U \leq t_1, V \leq t) = G(t_1, t), \\
P_{01}(t_1, t) &= \Pr (U > t_1, V \leq t), \\
P_{10}(t_1, t) &= \Pr (U \leq t_1, V > t), \\
P_{00}(t_1, t) &= \Pr (U > t_1, V > t).
\end{align*}
$$

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 $\mu$ 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 a given compact set $S$ 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 $\bar{A}$ as the collection of paths through the original...
We can define the measure on Borel sets in \( \mathbb{R} \) from below.

\[
\lambda^{11}(t_1, t; \delta) = P^{11}(t_1 + \delta(h(t_1, t) + 1), t),
\]

\[
\lambda^{01}(t_1, t; \delta) = P^{01}(t_1 + \delta(h(t_1, t) + 1), t),
\]

\[
\lambda^{10}(t_1, t; \delta) = P^{10}(t_1, t),
\]

\[
\lambda^{00}(t_1, t; \delta) = P^{00}(t_1, t),
\]

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 \( \delta \), 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 \( l_\lambda(y_1, y_2, x_1; \alpha, \delta) \). 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

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

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. 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 \( \lambda \) can be expressed as

\[
\psi_\delta(x_1, x) = \frac{1}{2} \{ y_1 y_2 P^{11}(x_1 + \alpha_0, x)^{-1/2} - (1 - y_1) y_2 P^{01}(x_1 + \alpha_0, x)^{-1/2} \}
\]

\[
\times \frac{\partial G(x_1 + \alpha_0, x)}{\partial u} (h(x_1 + \alpha_0, x) + 1).
\]

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

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

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

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

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

allowing us to characterize

\[
I_{\lambda, \alpha} \leq 4 \| h \|^2_{L^2(\pi)}
\]
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 $S_{\epsilon^*}$ to be a compact subset of $S$ such that the Euclidean distance of the boundary of $S$ from the boundary of $S_{\epsilon^*}$ is at least $\epsilon^*$, where $\epsilon^*$ 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 $a \in C^2(\mathbb{R}^2)$ 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 $S_{\epsilon^*}$ and, due to its compactness it is uniformly continuous there. As a result, there exists $M = \sup_{S_{\epsilon^*}} \left| \frac{\partial a(u,v)}{\partial u} \right|$. There also exists $M' = \sup_{S} \left| a \right|$. Then we pick the direction $h^*$ as function with support on $S$ such that $h^* = \frac{a}{2} (a/M)$ in $S_{\epsilon^*}$. Then we note that

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

As a result, $I_{\lambda,\alpha} \leq \frac{B^2(M' + 1)^2}{M^2} \epsilon$. As the choice of $\epsilon$ was arbitrary, this proves that $\inf_{\lambda \in \Lambda} I_{\lambda,\alpha} = 0$. 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 $\mathcal{H}$ with the basis of normalized Hermite polynomials $\{h_l\}_{l=0}^{\infty}$ such that

(a) For any sufficiently large $K \in \mathbb{N}$ and $H_K = \{h_l\}_{l=0}^{K} \sup_{g \in \mathcal{G}} \inf_{\mu \in \mathfrak{M}_K} \left\| g - \sum_{l=0}^{K} \mu^l h_l \right\|_{L_2} = O(K^{-r})$, for $r > 0$

(b) $|h_l(\cdot)| \leq C$ and $\int (-1) |h_l(z)|^2 dz \leq C$

(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 $\sup_{Q} \sum_{N(\epsilon, \mathcal{F}_{h,l}, L_2(Q))} < A \epsilon^{-\gamma}$, for some $\gamma > 0$ and probability measures $Q$.

(ii) Consider functions $f(\cdot, t, s) = \int^{+t} \int^{+s} g(u, v) du dv$. For each $K \in \mathbb{N}$ the class of projections of these functions on each basis vector $\mathcal{F}_k = \{\text{proj}(f(\cdot, t, s), h_k), g \in \mathcal{G}, |t|, |s| < \infty\}$ has envelope $F_k$ such that $E \left[ \left\| F_k \right\|_{L_2} \right] < \infty$ and it has at most exponential covering number, i.e. there exist constants $A'$ and $\gamma'$ such that:

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

Next, we prove the following lemma.

**Lemma A.1** Suppose that the choice probability functions are estimated via an orthogonal sequence $\mathcal{H}^{(K)}(\cdot, \cdot) = \{H_k(x, x)\}_{k=0}^{K}$ and $\inf_{\mu \in \mathfrak{M}_K} \|P(y_1, y_2 | x_1, x) - \mu \mathcal{H}^{(K)}(x_1, x)\| = O(K^{-r})$. 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 $\{|x_1|, x_1 \leq c_n\}$. 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 (B.1) such that $\hat{l}_{K,c_n}(\hat{\alpha}_n) \geq \sup_{\alpha} \hat{l}_{K,c_n}(\alpha) - o_p\left(\sqrt{\nu(c_n)/n}\right)$ we have
\[
\sqrt{\frac{n}{\nu(c_n)} |\hat{\alpha}_n^* - \alpha_0|} = O_p(1).
\]

**Proof:** We introduce the “uncensored” objective function
\[
l(\alpha; y_1, y_2, x_1, x) = y_1y_2 \log \hat{P}_{n11}^1(x_1 + \alpha, x) + (1 - y_1)y_2 \log \hat{P}_{n11}^0(x_1 + \alpha, x),
\]
with $Q(\alpha) = E[l(\alpha; y_1, y_2, x_1, x)]$, and $\hat{P}_{n11}$ defined in Appendix B. Denote $\hat{l}(\alpha) = \frac{1}{n} \sum_{i=1}^n l(\alpha; y_1, y_2, x_1, x_i)$. Also use the censoring function $\omega_n(\cdot) = 1\{|\cdot| \leq c_n\}$
\[
\hat{l}(\alpha; y_1, y_2, x_1, x) = y_1y_2 \omega_n(x_1 + \alpha) \omega_n(x) \log \mathcal{P}^{11}(x_1 + \alpha, x)
+ (1 - y_1)y_2 \omega_n(x_1 + \alpha) \omega_n(x) \log \mathcal{P}^{01}(x_1 + \alpha, x),
\]
and $\hat{l}(\alpha) = \frac{1}{n} \sum_{i=1}^n \hat{l}(\alpha; y_{1i}, y_{2i}, x_{1i}, x_i)$. Now consider the following decomposition of the objective function:
\[
\hat{l}(\alpha) - \hat{l}(\alpha_0) = R_1 + R_2 + R_3 + R_4 + R_5 + R_6,
\]
where
\[
R_1 = \hat{l}(\alpha) - \hat{l}(\alpha_0) - E[\hat{l}(\alpha)] + E[\hat{l}(\alpha)]
+ R_2 = \hat{l}(\alpha) - \hat{l}(\alpha_0) - E[\hat{l}(\alpha)] + E[\hat{l}(\alpha_0)]
+ R_3 = E[\hat{l}(\alpha)] - E[\hat{l}(\alpha)]
+ R_4 = E[\hat{l}(\alpha)] - Q(\alpha)
+ R_5 = -E[\hat{l}(\alpha_0)] + Q(\alpha_0)
+ R_6 = Q(\alpha) - Q(\alpha).
\]

**Term $R_1$**

For convenience, we introduce new notation denoting
\[
p^K_{\mathcal{H}}(z) = \omega_n(x_1) \omega_n(x) [\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x_1)] [\mathcal{H}_{l_2}(c_n) - \mathcal{H}_{l_2}(x)]
\]
and introduce vectors $p^K(z) = (p^{K1}(z), \ldots, p^{KK}(z))^\prime$. Also let $d^{00}_i = (1 - y_{1i})(1 - y_{z_i})$ and $d^{00} = (d^{00}_1, \ldots, d^{00}_n)^\prime$. Let $\Delta(z) = E[d^{00}|z]$ and $\Delta = (\Delta(z_1), \ldots, \Delta(z_n))^\prime$. We can project this function of $z$ on $K$ basis vectors of the sieve space. Let $\beta$ be the vector of coefficients of this projection. As demonstrated in Newey (1997), for $P = (p^K(z_1), \ldots, p^K(z_n))^\prime$ and $\hat{Q} = P'P/n$
\[
\|\hat{Q} - Q\| = O_p\left(\sqrt{\frac{K}{n}}\right),
\]
where his $\zeta_0(K) = C$. 

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

$$m^{Kk}(z) = \omega_n(x_1)\omega_n(x) [H_{t_1}(x_1) - H_{t_1}(-c_n)] [H_{t_2}(x) - H_{t_2}(-c_n)].$$

We then can write the estimator

$$\hat{P}^{11}(x_1, x) = m^K(z)\hat{Q}^{-1}Pd^0/n$$

Note that

$$m^{K'}(z) (\beta - \beta) = m^{K'}(z) \left(\hat{Q}^{-1}P' (d^0 - \Delta) / n + \hat{Q}^{-1}P' (\Delta - P\beta) / n\right).$$

We can evaluate the component in the second term as

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

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 $\lambda > 0$. Then application of the Cauchy-Schwartz inequality leads to

$$\left|m^{K'}(z)\hat{Q}^{-1}P' (d^0 - \Delta)\right| \leq \left\|\hat{Q}^{-1}m^K(z)\right\| \left\|P' (d^0 - \Delta)\right\|.$$ 

Then $\left\|\hat{Q}^{-1}m^K(z)\right\| \leq \frac{C}{2}\sqrt{K}$, and

$$\left\|P' (d^0 - \Delta)\right\| = \sqrt{\sum_{k=1}^{K} \left(\sum_{i=1}^{n} p^{Kk}(z_i) (d^0_i - \Delta(z_i))\right)^2} \leq \sqrt{K} \max_k \left|\sum_{i=1}^{n} p^{Kk}(z_i) (d^0_i - \Delta(z_i))\right|$$

Thus,

$$\left|m^{K'}(z)\hat{Q}^{-1}P' (d^0 - \Delta) / n\right| \leq \frac{CK}{\Delta} \max_k \left|\frac{1}{n} \sum_{i=1}^{n} p^{Kk}(z_i) (d^0_i - \Delta(z_i))\right|.$$ 

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

$$P \left(\max_i Y_i > Kc\right) \leq \sum_{i=1}^{K} P (Y_i > c).$$
Using this observation, we can find that
\[
P \left( \sup_z \frac{1}{n} \| m^{K}(z) \hat{Q}^{-1} P' (d^{00} - \Delta) \| > K\mu_n \right) \leq \sum_{k=1}^{K} P \left( \left\| \frac{1}{n} \sum_{i=1}^{n} p^{K}(z_i) (d^{00}_i - \Delta(z_i)) \right\| > \gamma_n \right),
\]
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
\[
\mathcal{P}_{n,k} = \{ p^{K}(\cdot) (d^{00} - \Delta(\cdot)) \}.
\]
Then we can apply the inequality from Theorem 37 in (Pollard 1984) to obtain
\[
P \left( \frac{1}{n} \left\| \sum_{i=1}^{n} p^{K}(z_i) (d^{00}_i - \Delta(z_i)) \right\| > \gamma_n \right) \leq 2 \exp \left( -\frac{2n\gamma_n^2}{C^2} + A'\gamma_n^{\gamma'} \right).
\]
As a result, we find that
\[
P \left( \sup_z \frac{1}{n} \| m^{K}(z) \hat{Q}^{-1} P' (d^{00} - \Delta) \| > K\mu_n \right) \leq 2 K \exp \left( -\frac{2n\gamma_n^2}{C^2} + A'\gamma_n^{\gamma'} \right).
\]
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
\[
\sup_{(x_1, x) \in \mathcal{X}} \| \hat{P}^{11}(x_1, x) - \text{proj} (\mathcal{P}^{11}(x_1, x) | H_{K}) \| = o_p \left( n^{\frac{\gamma'}{2}} \right).
\]
From the second term we provide the evaluation
\[
\sup_{\mathcal{P}^{11} \in H_{x_1, x}} \sup \| \text{proj} (\mathcal{P}^{11}(x_1, x) | H_{K}) - \mathcal{P}^{11}(x_1, x) \| = O(K^{-r})
\]
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 attain 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
\[
E \left[ \hat{l}(\alpha) \right] = E \left[ \omega_n(x_1 + \alpha) \omega_n(x) \left( \mathcal{P}^{11}(x_1 + \alpha, x) \log \hat{P}^{11}_n(x_1 + \alpha, x) + \mathcal{P}^{01}(x_1 + \alpha, x) \log \hat{P}^{01}_n(x_1 + \alpha, x) \right) \right].
\]
Similarly, we can express

\[
E \left[ \hat{\ell}(\alpha) \right] = E \left[ \omega_n(x_1 + \alpha) \omega_n(x) \left( P_{11}^1(x_1 + \alpha, x) \log P_{11}^1(x_1 + \alpha, x) 
+ P_{01}^1(x_1 + \alpha, x) \log P_{01}^1(x_1 + \alpha, x) \right) \right].
\]

One can attain a uniform rate

\[
\sup_{x_1, x} \left\| \hat{P}_n^1(x_1 + \alpha, x) - P_{11}^1(x_1 + \alpha, x) \right\| = O_p \left( \sqrt{\frac{K}{n}} + K^{-r} \right),
\]
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 R₄ and R₅**

Consider term R₄. We can evaluate this term as

\[
|E \left[ \hat{\ell}(\alpha) - Q(\alpha) \right| \leq 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_{11}^1(x_1 + \alpha, x) \log P_{11}^1(x_1 + \alpha, x) f(x_1, x) \, dx_1 \, dx.
\]

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

\[
|E \left[ \hat{\ell}(\alpha) - Q(\alpha) \right| \leq 4E \left[ y_1 y_2 \right] \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \log P_{11}^1(x_1 + \alpha, x) f(x_1, x) \, dx_1 \leq C \beta(c_n).
\]

from Assumption 2

**Term R₂**

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 \( \alpha_0 \) with a negative definite Hessian.

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

\[
\ell(\alpha, y_1, y_2, x_1, x) = [y_1 y_2 \log P_{11}^1(x_1 + \alpha, x) + (1 - y_1) y_2 \log P_{01}^1(x_1 + \alpha, x)] \omega_n(x_1 + \alpha) \omega_n(x)
\]

\[
\mathcal{F}_{n, \delta} = \{ f = \ell(\alpha, \cdot) - \ell(\alpha_0, \cdot), \ |\alpha - \alpha_0| \leq \delta \}.
\]
Provided that the density of errors is twice differentiable in mean square with bounded mean square derivatives, there exist bounded functions $\hat{P}^{11}$ and $\hat{P}^{01}$ such that functions in class $F_{n,\delta}$ have envelope

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

Then, by Assumption 2, we can evaluate

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

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 $\eta$ 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 $\alpha_0$ in Assumption 10. We can evaluate

$$P \left( \sup_{h \in S_{n,j}} \left( \hat{l}(\alpha_0 + \frac{h}{r_n}) - \hat{l}(\alpha_0) \right) \geq 0 \right) \leq P \left( \sup_{h \in S_{n,j}} |R_2| \geq |R_1| + |R_3| + |R_4| + |R_5| \right) \leq \frac{2^{2j-2}}{r_n^2} + O \left( \sqrt{\frac{K}{n}} + K^{1-r} + \beta(c_n)^{-1} \right),$$

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

$$P \left( \sup_{h \in S_{n,j}} \left( \hat{l}(\alpha_0 + \frac{h}{r_n}) - \hat{l}(\alpha_0) \right) \geq 0 \right) \leq E \left[ \sup_{h \in S_{n,j}} \left| \hat{l}(\alpha_0 + \frac{h}{r_n}) - \hat{l}(\alpha_0) - E \left[ \hat{l}(\alpha_0 + \frac{h}{r_n}) \right] + E \left[ \hat{l}(\alpha_0) \right] \right| \right] + \frac{2^{2j-2}}{r_n^2} + O \left( \sqrt{\frac{K}{n}} + K^{1-r} + \beta(c_n)^{-1} \right)$$
Using the empirical process notation, we define the covering integral as
\[ J(\delta, F) = \sup_Q \int_0^\delta \sqrt{1 + \log N(\epsilon \|F\|_{L^2(Q)}, L^2(Q))} \, d\epsilon, \]
where \( Q \) is the probability measure, \( F \) 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 \( F_{n, \delta} \), we can use the maximum inequality to evaluate
\[
P\left( \sup_{h \in S_{n,j}} \sqrt{n} \left( \hat{\ell}(\alpha_0 + \frac{h}{r_n}) - \hat{\ell}(\alpha_0) \right) - E \left[ \hat{\ell}(\alpha_0 + \frac{h}{r_n}) \right] + E \left[ \hat{\ell}(\alpha_0) \right] \right) \leq J(1, F_{n,h/r_n}) E \left[ F_{n,h/r_n} \right]^{1/2} = O \left( \|c_n\|^{1/2} \frac{2^j}{r_n} \right).
\]
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
\[
P\left( \sup_{h \in S_{n,j}} \left( \hat{\ell}(\alpha_0 + \frac{h}{r_n}) - \hat{\ell}(\alpha_0) \right) \geq 0 \right) \leq O \left( 2^{-j+2} r_n \sqrt{n} \right).
\]
This implies that
\[
P \left( |\hat{h}| > 2^M \right) \leq O \left( 2^{-M+3} r_n \sqrt{n} \right).
\]
The right-hand side converges to zero for \( M \to \infty \) if \( r_n = \frac{n}{\sqrt{\nu(c_n)}} \).

Q.E.D.

### A.3 Proof of Theorem 2.2

First, consider the following evaluation from the proof of Lemma 2.1
\[
P\left( \sup_{h \in S_{n,j}} \left( \hat{\ell}(\alpha_0 + \frac{h}{r_n}) - \hat{\ell}(\alpha_0) \right) \geq 0 \right) \leq E \left[ \sup_{h \in S_{n,j}} \left| \hat{\ell}(\alpha_0 + \frac{h}{r_n}) - \hat{\ell}(\alpha_0) \right| \right] \leq \frac{2^{2j-2} \nu(c_n)}{r_n^2}.
\]
Using the maximum inequality as before we can conclude that the ratio can be evaluated as
\[
P\left( \sup_{h \in S_{n,j}} \left( \hat{\ell}(\alpha_0 + \frac{h}{r_n}) - \hat{\ell}(\alpha_0) \right) \geq 0 \right) \leq O \left( 2^{-j+1} \frac{\nu(c_n)^{1/2}}{\sqrt{n}} \right).
\]
We note that evaluation here is different, because, unlike in Lemma 2.1 here we allow \( r_n \beta(c_n) = O(1) \). This allows us to obtain
\[
P \left( \sqrt{n} \nu(c_n)^{1/2} |\hat{h}| > 2^M \right) \leq O \left( 2^{-M+2} r_n \sqrt{n} \right).
\]
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
\[
\lim_{L \to \infty} \limsup_{n \to \infty} R \left( \alpha_0 + \frac{h}{r_n}, r_n, L \right) = 0.
\]
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 \( \Lambda(P_1, P_2) = \frac{dP_1}{dP_2} \). Then the following lemma is the result given in Koroselev and Tsybakov (1993).

**Lemma A.2** Suppose that \( \alpha_0^1 = \alpha(P_1) \) and \( \alpha_0^2 = \alpha(P_2) \), and let \( \lambda > 0 \) be such that
\[
P_{P_2}(\Lambda(P_1, P_2) > \exp(-\lambda)) \geq p > 0,
\]
and \( |\alpha_0^1 - \alpha_0^2| \geq 2s_n \). Then for any estimator \( \hat{\alpha}_{0,n} \) we have \( \max_{P_1, P_2} P (|\hat{\alpha}_{0,n} - \alpha_0| > s_n) \geq p \exp(-\lambda/2) \).

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
\[
n\hat{L}(\alpha) = n\hat{\ell}(\alpha) + n\hat{e}(\alpha)
\]
with
\[
\hat{\ell}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \left\{ y_{1i} \log P_1^{11}(x_{1i} + \alpha, x_i) + (1 - y_{1i}) \log P_0^{01}(x_{1i} + \alpha, x_i) \right\}
\]
\[
y_{2i} \mathbf{1}\{|x_{1i}| > c_n, |x_i| > c_n\}
\]
Note that we use the same distribution of covariates \( x_1 \) and \( x \). For \( c_n \to \infty \), pick\(^{16}\)
\[
P_2(\cdot, \cdot) = P(\cdot, \cdot), \text{ and } P_1(\cdot, \cdot) = P(\cdot, \cdot)\omega_n(\cdot)\omega_n(\cdot).
\]
Following from our previous analysis for such choices of \( P_1(\cdot) \) and \( P_2(\cdot) \), the corresponding likelihood maximizers satisfy
\[
|\alpha_1 - \alpha_2| = O(\beta(c_n)).
\]
\(^{16}\) The selected \( P_1 \) 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.
We can then express
\[
\Lambda(\mathcal{P}_1, \mathcal{P}_2) = \exp \left( n\hat{\mathcal{L}}_1(\alpha_1) - n\hat{\mathcal{L}}_2(\alpha_2) \right) \\
= \exp \left( n\hat{\ell}(\alpha_1) - n\hat{\ell}(\alpha_2) - n\hat{\epsilon}(\alpha_2) \right) \\
= \exp \left( n\left[ \ell(\alpha_1) - \ell(\alpha_2) - \ell(\alpha_1) + \ell(\alpha_2) \right] - n\hat{\epsilon}(\alpha_2) - n(\ell(\alpha_2) - \ell(\alpha_1)) \right)
\]

We note that \( \hat{\ell}(\alpha_1) - \hat{\ell}(\alpha_2) - \ell(\alpha_1) + \ell(\alpha_2) = o_p(1) \) and \( \hat{\epsilon}(\alpha_2) = o_p(1) \). 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 \( \alpha_1 \) 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.

### 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 \( E[Y_1Y_2|x_1, x] \). In Appendix E we derive the Fourier transformation of each of these probabilities and demonstrate the identification of the distribution of unobserved heterogeneity. Here we provide a simple argument that demonstrates the identification of parameter \( \alpha \). Take points \( x > x' \) in the support of \( X \) and consider
\[
E[Y_2|X_1 = x_1, X = x] - E[Y_2|X_1 = x_1, X = x']
= \int 1 \left\{ u - \alpha\Phi \left( \frac{x - v}{\sigma} \right) \leq x_1 \leq u - \alpha\Phi \left( \frac{x' - v}{\sigma} \right) \right\} g(u, v) \, du \, dv.
\] (A.2)

For each fixed pair of \( x > x' \) this function is absolutely integrable inside any interval \([-c, c]\). In fact, the set of points \((u, v) \in S_c = \{(u, v) : x_1 + \alpha\Phi \left( \frac{x - v}{\sigma} \right) \leq u \leq x_1 + \alpha\Phi \left( \frac{x' - v}{\sigma} \right), -c \leq x_1 \leq c\}\) is a closed connected subset of \( \mathbb{R}^2 \) and thus has finite measure with respect to \( g(\cdot, \cdot) \). Moreover, for any sequence \( c \to \infty \) the limit of measures of sets is well-defined and is bounded by 1. Therefore, the improper integral of (A.2) on \( \mathbb{R} \) over \( x_1 \) is well-defined. Taking this integral, we note that
\[
\int \int 1 \left\{ u - \alpha\Phi \left( \frac{x - v}{\sigma} \right) \leq x_1 \leq u - \alpha\Phi \left( \frac{x' - v}{\sigma} \right) \right\} g(u, v) \, du \, dv \, dx_1
= \alpha \int \left( \Phi \left( \frac{x - v}{\sigma} \right) - \Phi \left( \frac{x' - v}{\sigma} \right) \right) g(u, v) \, du \, dv.
\]
Recall that
\[ E[Y_1|X = x] = \int \Phi \left( \frac{x-v}{\sigma} \right) g(u,v) \, du \, dv. \]

Therefore, we can write the final expression that identifies parameter \( \alpha \) as
\[ \alpha = \int \frac{E[Y_2|X_1 = x_1, X = x] - E[Y_2|X_1 = x_1, X = x']}{E[Y_1|X = x] - E[Y_1|X = x']} \, dx_1. \tag{A.3} \]

Therefore, parameter \( \alpha \) 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.3). To compute the information corresponding to the parameter of interest, we construct the log-likelihood of the model by explicitly expressing the probabilities:
\[
\begin{align*}
P_{11}(x_1, x; \alpha, g) &= \int 1\{x_1 - u + \alpha \Phi \left( \frac{x-v}{\sigma} \right) > 0\} \Phi \left( \frac{x-v}{\sigma} \right) g(u,v) \, du \, dv, \\
P(x_1, x; \alpha, g) &= \int \Phi \left( \frac{x-v}{\sigma} \right) g_v(v) \, dv, \\
Q(x_1, x; \alpha, g) &= \int 1\{x_1 - u + \alpha \Phi \left( \frac{x-v}{\sigma} \right) > 0\} g(u,v) \, du \, dv.
\end{align*}
\]

We can then express all probabilities of interest as
\[
\begin{align*}
P_{01}(x_1, x; \alpha, g) &= P(x_1, x; \alpha, g) - P_{11}(x_1, x; \alpha, g), \\
P_{10}(x_1, x; \alpha, g) &= Q(x_1, x; \alpha, g) - P_{11}(x_1, x; \alpha, g), \\
P_{00}(x_1, x; \alpha, g) &= 1 - Q(x_1, x; \alpha, g) - P(x_1, x; \alpha, g) + P_{11}(x_1, x; \alpha, g),
\end{align*}
\]

and the derivatives of the probabilities of interest as
\[
\begin{align*}
\frac{\partial P_{11}(x_1, x; \alpha, g)}{\partial \alpha} &= \int \Phi \left( \frac{x-v}{\sigma} \right) g \left( x_1 + \alpha \Phi \left( \frac{x-v}{\sigma} \right), v \right) \, dv \equiv D_1(x_1, x; \alpha, g), \\
\frac{\partial Q(x_1, x; \alpha, g)}{\partial \alpha} &= \int g \left( x_1 + \alpha \Phi \left( \frac{x-v}{\sigma} \right), v \right) \, dv \equiv D_2(x_1, x; \alpha, g).
\end{align*}
\]

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:
\[
\begin{align*}
r(y_1, y_2, x_1, x; \alpha, g)^{1/2} &= y_1 y_2 P_{11}(x_1, x; \alpha, g)^{1/2} + y_1(1 - y_1) P_{10}(x_1, x; \alpha, g)^{1/2} \\
&\quad+ (1 - y_1) y_2 P_{01}(x_1, x; \alpha, g)^{1/2} + (1 - y_1)(1 - y_1) P_{00}(x_1, x; \alpha, g)^{1/2}.
\end{align*}
\]
We can express the mean square derivative with respect to $\alpha$ 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 $\alpha$ as

$$
I_\alpha = 4 \int (\psi_\alpha)^2 d\mu.
$$

If $\nu$ 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 $\mathbb{R}^2$

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

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

We can then express the information of the model as

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

We construct the measure $\pi^*$ which minorizes the Radon-Nikodym density of measures $\pi_1$ and $\pi_2$ meaning that: $\frac{d\pi^*}{d\nu} = \min\{\frac{d\pi_1}{d\nu}, \frac{d\pi_2}{d\nu}\}$. Based on this structure of the measure, we can write:

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

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

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

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

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 $M(t) = \inf_{(x_1, x) \in S} |g(x_1 + \alpha w(t), x - \sigma t)|$ 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

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

Therefore, the information corresponding to parameter $\alpha$ is strictly positive whenever $\sigma > 0$. 

Q.E.D.
A.5.2 Proof for part (ii)

Now we provide the upper bound on information with the notice that our bound will be conservative. However, it will be sufficient to deliver the claim in the theorem. Consider the expression for the information in the incomplete information triangular model expressed in (A.4):

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

We construct the measure \( \pi^{**} \) (it may be not a probability measure) that is constructed as an integral over:

\[ d\pi^{**} d\nu = \max \{ d\pi_1 d\nu, d\pi_1 d\nu \}, \]

where the maximum is considered in the pointwise sense over all regular points of measures \( \pi_1 \) and \( \pi_2 \) and where \( d\pi_1 d\nu \) is the Radon-Nykodim density with respect to the \( \sigma \)-finite measure \( \nu \). Then we note that \( \pi^{**}(\mathbb{R}^2) < \Pi < \infty \), assuming that both measures are defined on the entire \( \mathbb{R}^2 \). We denote \( w(t) = \Phi(t) \) and \( t = (x - v)/\sigma \) and express

\[ D_1(x_1, x; \alpha_0, g) = \sigma^2 \int w(t) g(x_1 + \alpha_0 w(t), x - \sigma t) \, dt \leq \sigma^2 \int_{w \in [0,1]} \max g(x_1 + \alpha_0 w, x - \sigma t) \, dt \]

and

\[ D_2(x_1, x; \alpha_0, g) - D_1(x_1, x; \alpha_0, g) = \sigma^2 \int (1 - w(t)) g(x_1 + \alpha_0 w(t), x - \sigma t) \, dt \leq \sigma^2 \int_{w \in [0,1]} \max g(x_1 + \alpha_0 w, x - \sigma t) \, dt \]

As a result, we find that

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

\[ \leq 2\sigma^2 \left\| \max_{w \in [0,1]} g_u(x_1 + \alpha_0 w) \right\|_{L_2(\pi^{**})}^2. \]

Note that \( g_u(\cdot) \) is a probability density which we assumed was absolutely continuous and square-integrable. Then, we find that

\[ \left\| \max_{w \in [0,1]} g_u(x_1 + \alpha_0 w) \right\|_{L_2(\pi^{**})}^2 \leq \left( \sup_x g_u(x) \right)^2 = g_u^2, \]

given that \( g(\cdot, \cdot) \) is twice continuously differentiable with finite moments.

As a result, we provided an upper bound \( I_\alpha \leq 2\sigma^2 g_u^2 \). As \( \sigma \to \to 0 \) this upper bound converges to zero, meaning that \( I_\alpha \to 0 \).

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 $\alpha_1$ and $\alpha_2$ is equal to zero.

Suppose that $\Gamma$ 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 $\alpha_1$ and $\alpha_2$ coincide. We construct the probability mass corresponding to the region with multiple equilibria 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

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

with respect to the measure $\mu$ 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 $h_1 : \mathbb{R}^2 \to \mathbb{R}$ and $h_2 : \mathbb{R}^2 \to \mathbb{R}$ 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 $\Lambda$ as the collection of paths through the original model which we design as

$$\lambda^{11}(t_1, t_2; \delta_1, \delta_2) = P^{11}(t_1 + \delta_1(h_1(t_1, t_2) + 1), t_2 + \delta_2(h_2(t_1, t_2) + 1)),$$
$$\lambda^{01}(t_1, t_2; \delta_1, \delta_2) = P^{01}(t_1 + \delta_1(h_1(t_1, t_2 + \alpha_2) + 1), t_2),$$
$$\lambda^{10}(t_1, t_2; \delta_1, \delta_2) = P^{11}(t_1, t_2 + \delta_2(h_2(t_1 + \alpha_1, t_2 + 1)), t_2 + \delta_2(h_2(t_1 + \alpha_1, t_2 + \alpha_2) + 1)),$$
$$\lambda^{00}(t_1, t_2; \delta_1, \delta_2) = P^{11}(t_1, t),$$
$$\gamma(t_1, t_2; \alpha_1, \alpha_2, \delta_1, \delta_2) = \Pr\left( t_1 < U \leq t_1 + \alpha_1 + \delta_1(h_1(t_1 + \alpha_1, t_2 + \alpha_2) + 1), t_2 < V \leq t_2 + \alpha_2 + \delta_2(h_2(t_1 + \alpha_1, t_2 + \alpha_2) + 1) \right)$$

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
\( \delta \), 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

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

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 \( \lambda \) can be expressed as

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

\[
\times \frac{\partial G(x_1 + \alpha_1, x_2 + \alpha_2)}{\partial u} (h_1(x_1 + \alpha_1, x_2 + \alpha_2) + 1)
\]

\[
- \frac{1}{2} (1 - y_1) y_2 P^{-+}(x_1, x_2; \alpha)^{-1/2} \frac{\partial G(x_1 + \alpha_1, x_2)}{\partial u} (h_1(x_1 + \alpha_1, x_2 + \alpha_2) + 1),
\]
Suppose that a sequence $c_n$ is selected such that $\nu(c_n)/n \to 0$, $K^n/\nu(c_n) \to 0$, $\nu(c_n)K^2/n \to \infty$. Then for any sequence $\hat{\alpha}_n$ with the function $\hat{l}(\alpha)$ corresponding to the maximand of (B.2) such that $\hat{l}(\alpha_n) \geq \sup_{\alpha} \hat{l}(\alpha) - o_p\left(\sqrt{\nu(c_n)/n}\right)$ we have

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

We note that the corresponding score has mean zero.

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

$$I_{\lambda,\alpha_1} \leq 4 \int_{\Omega} (\psi_{\alpha_1} - \psi_{\beta,1})^2 d\mu$$

$$= \int_{\Omega} \frac{1}{4} \left( [P^{+\lambda,\beta}(x_1, x_2; \alpha_0) - 1 + P^{-(x_1, x_2; \alpha_0)} - 1] \left( \frac{\partial G(x_1 + \alpha_1, x_2 + \alpha_2)}{\partial u} \right)^2 \right) d\nu(x_1, x_2)$$

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

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

allowing us to characterize

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

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 $\inf_{\lambda \in \Omega} I_{\lambda,\alpha_1} = 0$. Similarly, we can also show that $\inf_{\lambda \in \Omega} I_{\lambda,\alpha_2} = 0$.

Q.E.D.

### A.7 Convergence rate for the iterative estimator

In this subsection we establish the convergence rate for the iterative estimator which we use to establish the upper convergence rate for strategic interaction parameters in the complete information game. 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 $\nu(c_n)/n \to 0$, $K^n/\nu(c_n) \to 0$, $\nu(c_n)K^2/n \to \infty$. Then for any sequence $\hat{\alpha}_n$ with the function $\hat{l}(\alpha)$ corresponding to the maximand of (B.2) such that $\hat{l}(\alpha_n) \geq \sup_{\alpha} \hat{l}(\alpha) - o_p\left(\sqrt{\nu(c_n)/n}\right)$ we have

$$\sqrt{\frac{n}{\nu(c_n)}}|\hat{\alpha}^{*}_{1n} - \alpha_{1,0}| = O_p(1), \text{ and } \sqrt{\frac{n}{\nu(c_n)}}|\hat{\alpha}^{*}_{2n} - \alpha_{2,0}| = O_p(1).$$
Proof:
For simplicity of notation, denote $y = (y_1, y_2)$ and $x = (x_1, x_2)$. Let $\hat{P}_K^{ij}(\cdot; \alpha)$ 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

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

with $Q(\alpha) = E[l(\alpha; y, x)]$, and $\hat{P}_n^{11}$ 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

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

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

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

where

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

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

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

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

$$R_5 = -E[\hat{l}(\alpha_0)] + Q(\alpha),$$

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

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 $\delta$ 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 parameters \((\alpha_1, \alpha_2)\). 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 \(\alpha_1\) and \(\alpha_2\). In other words, if \((P_1(\cdot, \cdot), P_2(\cdot, \cdot))\) correspond to \((\alpha_1, \alpha_2)\) and \((P'_1(\cdot, \cdot), P'_2(\cdot, \cdot))\) correspond to \((\alpha'_1, \alpha'_2)\) 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 \(\alpha_1\) and \(\alpha_2\). We note that provided the differentiability of the distribution of perturbations \(\eta_1\) and \(\eta_2\), 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 \(a_1 = \phi(\Phi^{-1}(P_1)) / \sigma\) and \(a_2 = \phi(\Phi^{-1}(P_2)) / \sigma\). We note that this matrix is non-singular if and only if \(a_1 a_2 \neq 0\) and \(a_1 a_2 \alpha_1 \alpha_2 \neq 1\). 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 $J^q 1\{(q_1, q_2) \notin S^m\} + \bar{J}^q 1\{(q_1, q_2) \in S^m\}$. The matrix defined this way is \textit{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-deficient.

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 $a_1 \alpha_1 \alpha_1 \alpha_2 = 1$ 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 Lebesgue measure zero, its level set also has Lebesgue 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

$$J^\alpha = \begin{pmatrix} \frac{\partial P_1}{\partial q_1} & \frac{\partial P_1}{\partial q_2} \\ \frac{\partial P_2}{\partial q_1} & \frac{\partial P_2}{\partial q_2} \end{pmatrix} = \frac{a_1 a_2}{1 + \alpha_1 \alpha_2 a_1 a_2} \begin{pmatrix} P_1 / a_2 & \alpha_1 P_1 \\ \alpha_2 P_2 & P_2 / a_1 \end{pmatrix}.$$ 

This matrix is non-singular whenever $a_1 a_2 \neq 0$ and $a_1 a_2 \alpha_1 \alpha_2 \neq 1$. Thus $J^\alpha$ 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, $J^\alpha$ 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.

\textbf{Step 2} Second, we show that from observed conditional expectations $E [Y_1|x_1, x_2]$, $E [Y_1|v_1, x_2]$, and $E [Y_1 Y_2|x_1, x_2]$ we can uniquely recover the corresponding equilibrium 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

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

We now use this characteristic function to express the first derivative of player’s beliefs which we further use to express the strategic interaction parameter. To do so we consider the following limit:

$$\lim_{t_1 \to 0} \frac{\chi_{P_1}(t_1, t_2)}{it_1} = - \int e^{-it_2 q_2} q_1 \frac{\partial^2 P_1(q_1, q_2)}{\partial q_1 \partial q_2} dq_1 dq_2 = \int e^{-it_2 q_2} \frac{\partial P_1(q_1, q_2)}{\partial q_2} dq_1 dq_2.$$
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 \( \alpha_1 \) and \( \alpha_2 \), 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 \( Q(t_1, t_2) \) is the Fourier transform of \( E[Y_1|x_1, x_2] \) and \( F(t_1, t_2) \) is the Fourier transform of \( E[Y_2|x_1, x_2] \) then

\[
Q(0, t_2) = \frac{\alpha_2}{it_2} A_1(t_2) \chi_v(t_2) (1 + \pi it_2 \delta(t_2)),
\]

\[
F(t_1, 0) = \frac{\alpha_1}{it_1} A_2(t_1) \chi_u(t_1) (1 + \pi it_1 \delta(t_1)).
\]

Performing an inverse Fourier transform, we find that

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

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

(A.5)

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.5).

Q.E.D.

A.10 Proof of Theorem 5.2

A.10.1 Proof of result (i)

We start the proof with the following lemma that demonstrates that the addition of our equilibrium selection mechanism does not affect the smoothness properties of the semiparametric likelihood function.
Lemma A.4 The set of values of strategic interaction parameters $\alpha_1$ and $\alpha_2$ in the static game of incomplete information for which the game has multiple equilibria is closed connected set with a differentiable boundary $S^m(\alpha_1, \alpha_2)$

Proof: 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:

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

For given parameters $\alpha_1$, $\alpha_2$, 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 $S^m(\alpha_1, \alpha_2)$.

Q.E.D.

We now use the constructed set of parameters leading to multiple equilibria to form the likelihood function of the models. The likelihood of the model can then be characterized by four objects:

$$E \left[ Y_1 Y_2 | x_1, x_2 \right] = P_{11}(x_1, x_2; \alpha) = \int \Phi \left( \frac{x_1 - u + \alpha_1 P_2(x_1 - u, x_2 - v)}{\sigma} \right) \times \Phi \left( \frac{x_2 - v + \alpha_2 P_2(x_1 - u, x_2 - v)}{\sigma} \right) g(u, v) du dv,$$

$$E \left[ Y_1 | x_1, x_2 \right] = Q_1(x_1, x_2; \alpha) = \int \Phi \left( \frac{x_1 - u + \alpha_1 P_2(x_1 - u, x_2 - v)}{\sigma} \right) g(u, v) du dv,$$

$$E \left[ Y_2 | x_1, x_2 \right] = P_1(x_1, x_2; \alpha) = \int \Phi \left( \frac{x_2 - v + \alpha_2 P_1(x_1 - u, x_2 - v)}{\sigma} \right) g(u, v) du dv,$$

$$\Pr (((X_1 - U, X_2 - V) \in S^m(\alpha_1, \alpha_2) | x_1, x_2) = \Delta(x_1, x_2; \alpha)$$

$$= \int 1 \{(x_1 - u, x_2 - v) \in S^m(\alpha_1, \alpha_2) \} g(u, v) du dv.$$
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 $D_1(x_1, x_2; \alpha) = \frac{\partial}{\partial \alpha} (P_{11}(x_1, x_2; \alpha) - \frac{1}{2} \Delta(x_1, x_2; \alpha))$, $D_2(x_1, x_2; \alpha) = \frac{\partial}{\partial \alpha} P^{++}(x_1, x_2; \alpha)$, 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 $\nu$ 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

$$\pi_2(A) = \int A \frac{1 - Q_1(x_1, x; \alpha_0)}{P^{-+}(x_1, x; \alpha_0) P^{--}(x_1, x; \alpha_0)} d\nu(x_1, x),$$
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
\[
I_\alpha \geq \|D_1(x_1, x_2; \alpha_0)\|_{L_2(\pi_1)}^2 + \|D_2(x_1, x_2; \alpha_0)\|_{L_2(\pi_2)}^2 + \|D_3(x_1, x_2; \alpha_0)\|_{L_2(\pi_3)}^2.
\]

Then we can construct the measure \(\pi^*\) which minorizes the Radon-Nikodym density of measures \(\pi_1\) and \(\pi_2\) meaning that:
\[
\frac{d\pi^*}{d\nu} = \min\{\frac{d\pi_1}{d\nu}, \frac{d\pi_2}{d\nu}\}.
\]
Based on this structure of the measure, we can write:
\[
I_\alpha \geq \|D_1(x_1, x_2; \alpha_0)\|_{L_2(\pi^*)}^2 + \|D_2(x_1, x_2; \alpha_0)\|_{L_2(\pi^*)}^2 + \|D_3(x_1, x_2; \alpha_0)\|_{L_2(\pi^*)}^2.
\]

By combining the triangle inequality and taking into account the non-negativity of the square, we can evaluate
\[
I_\alpha \geq \|D_1(x_1, x_2; \alpha_0) + D_2(x_1, x_2; \alpha_0)\|_{L_2(\pi^*)}^2.
\]

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

We denote \(t_1 = (x_1 - u)/\sigma\) and \(t_2 = (x_2 - v)/\sigma\). Then
\[
D_1(x_1, x_2; \alpha_0) + D_2(x_1, x_2; \alpha_0) = \sigma^2 \int \phi \left( \frac{1}{\sigma} \Phi^{-1}(P_1(\sigma t_1, \sigma t_2)) \right) \left( \frac{\alpha_1}{1} \frac{\partial P_2(\sigma t_1, \sigma t_2)}{\partial \alpha} + (P_2(\sigma t_1, \sigma t_2), 0) \right) g(\sigma t_1 + x_1, \sigma t_2 + x_2) dt_1 dt_2.
\]

Denote \(w(t_1, t_2) = \phi \left( \frac{1}{\sigma} \Phi^{-1}(P_1(\sigma t_1, \sigma t_2)) \right) \left( \frac{\alpha_1}{1} \frac{\partial P_2(\sigma t_1, \sigma t_2)}{\partial \alpha} + (P_2(\sigma t_1, \sigma t_2), 0) \right) g(\sigma t_1 + x_1, \sigma t_2 + x_2)\). Then we can express
\[
D_1(x_1, x_2; \alpha_0) + D_2(x_1, x_2; \alpha_0) = \sigma^2 \int w(t_1, t_2) g(x_1 + \sigma t_1, x_2 + \sigma t_2) dt_1 dt_2.
\]

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 \(M(t_1, t_2) = \inf_{(x_1, x_2) \in S} \|g(x_1 + \sigma t_1, x_2 + \sigma t_2)\|\) which is not equal to zero for at least some \((t_1, t_2) \in \mathbb{R}^2\) which is not equal to zero for at least some \((t_1, t_2) \in \mathbb{R}^2\). We take \(\sqrt{\kappa} = \sup_{t \in [-B, B] \times [-B, B]} |M(t)|\), 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 \(M(t) > \sqrt{\kappa}/2\). 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 \(\alpha_1 \frac{\partial P_2}{\partial \alpha} > 0\)). Moreover this function has a well-defined limit as \(\sigma \to 0\). Thus this function attains its lower bound in every compact set and that lower bound is above zero
\[
\inf_{(t_1, t_2) \in B_R(t^*_1, t^*_2)} \|w(t_1, t_2)\| = A \sqrt{\kappa} > 0.
\]
We substitute our evaluations into the bound for the information:

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

\[ \geq CA^2\sigma^2\sqrt{\epsilon} \left\| \int_{\mathbb{R}} M(t_1, t_2) \, dt \right\|^2 I_{2x2} \geq \frac{1}{2} CA^2 R^2 \sigma^2 \epsilon I_{2x2} > 0, \]

where \( I_{2x2} \) is the identity matrix. Therefore the information corresponding to parameters \( \alpha_1 \) and \( \alpha_2 \) is strictly positive.

### A.10.2 Proof of result (ii)

Consider the measures introduced in the proof of Theorem 5.2. Suppose that measure \( \pi^{**} \) is such its Radon-Nikodym density is constructed as: \( \frac{d\pi^{**}}{d\nu} = \max\{\frac{d\pi_1}{d\nu}, \frac{d\pi_2}{d\nu}\} \). Then we can see that

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

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

Consider the change of variables \( t_1 = \Phi^{-1}(P_1(x_1 - u, x_2 - v)) \) and \( t_2 = \Phi^{-1}(P_2(x_1 - u, x_2 - v)) \).

Thus we can write

\[ |D_1(x_1, x_2; \alpha)| \leq \int \left( \frac{1 + \alpha_2 \alpha_1}{1 + \alpha_1 \alpha_0} \right) |a_1 a_2 \alpha_1^2 \alpha_2^2 - 1| P_1 P_2 g(x_1 + \alpha_1 P_1 - \sigma t_1, x_2 + \alpha_2 P_2 - \sigma t_2) \, dt_1 \, dt_2 \]

\[ \leq \sigma^2 \left( \frac{\alpha_2}{\alpha_1} \right) \int \phi(\Phi^{-1}(P_1)) \phi(\Phi^{-1}(P_2)) P_1 P_2 g(x_1 + \alpha_1 P_1 - \sigma t_1, x_2 + \alpha_2 P_2 - \sigma t_2) \, dt_1 \, dt_2 + o(\sigma^2) \]

\[ \leq \sigma^2 \tilde{\sigma}^2 \left( \frac{\alpha_2}{\alpha_1} \right) + o(\sigma^2), \]

provided that \( \phi(\cdot) \leq \tilde{\phi} \) and \( P_1, P_2 \leq 1 \). The same evaluation can be written for other components \( D_i(x_1, x_2; \alpha) \) with \( i = 1, 2, 3 \). We evaluate the information as

\[ I_\alpha \leq \sigma^2 \tilde{\sigma}^2 A + o(\sigma^2), \]

for a fixed matrix \( A \) (determined by coefficients \( \alpha_1 \) and \( \alpha_2 \)). We note that this is a conservative upper bound. When \( \sigma \to 0 \) this upper bound approaches zero. Thus, 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 \( h_t(x) = (\sqrt{2\pi t})^{-1/2} e^{-x^2/4t} H_t(x) \), where \( H_t(\cdot) \) is the \( l \)-th degree Hermite polynomial. Also denote \( \mathcal{H}_l(x) = \int_{-\infty}^x h_t(z) \, dz \). 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 \( c_n \rightarrow \infty \), and define the function \( \omega_n(x) = 1 \{ |x| \leq c_n \} \) and estimate the probability of both indicators are equal to zero (\( y_1 = y_2 = 0 \)) as

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

The estimates can be obtained via a regression of \( \omega_n(x_1)[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x_1)] \omega_n(x)[\mathcal{H}_{l_2}(c_n) - \mathcal{H}_{l_2}(x)] \) 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

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

and

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

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

\[
l(\alpha; y_1, y_2, x_1, x) = y_1 y_2 \omega_n(x_1 + \alpha)\omega_n(x) \log \hat{\mathcal{P}}_{n}^{11}(x_1 + \alpha, x)
+ (1 - y_1) y_2 \omega_n(x_1 + \alpha)\omega_n(x) \log \hat{\mathcal{P}}_{n}^{01}(x_1 + \alpha, x).
\]

Then we can express the empirical score as

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

This expression can be rewritten as

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

Setting the empirical score equal to zero, we obtain the estimator for \( \alpha_0 \) as

\[
\hat{\alpha}_n = \arg\max_{\alpha} \frac{1}{n} \sum_{i=1}^{n} l(\alpha; y_{1i}, y_{2i}, x_{1i}, x_i).
\]

(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 \( \omega_n(x) = 1\{ |x| \leq c_n \} \). 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

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

Similarly, we approximate the remaining probabilities

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

and

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

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

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

We can consider the sample profile log-likelihood

\[
\hat{l}(\alpha_1, \alpha_2) = \sup_{\alpha_1, \ldots, \alpha_K} \frac{1}{n} \sum_{i=1}^{n} l(\alpha_1, \alpha_2; y_{1i}, y_{2i}, x_{1i}, x_{2i}).
\]

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

\[
(\hat{\alpha}^{*}_{1n}, \hat{\alpha}^{*}_{2n}) = \arg\max_{\alpha_1, \alpha_2} \hat{l}(\alpha_1, \alpha_2).
\]
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

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

Therefore, we can select $\beta(c) = e^{-c}$ and the optimal rate will be $\sqrt{n/c^2}$ 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

\[
nc_n^A e^{-c_n^2} = O(1).
\]

The convergence rate will correspond to \( \sqrt{nc_n^A} \). 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 \( x_1 = x + \sqrt{u} \phi(x_1 + \alpha_0) / \Phi(x_1 + \alpha_0) - \alpha_0 \) 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: $x_1 + \alpha_0 < x$

\[
y_1 = 1, \quad y_2 = 1 \\
y_1 = 0, \quad y_2 = 1 \\
y_1 = 0, \quad y_2 = 0
\]

\[u\]

Case 2: $x_1 \geq x$

\[
y_1 = 0, \quad y_2 = 1 \\
y_1 = 0, \quad y_2 = 0 \\
y_1 = 0, \quad y_2 = 0
\]

\[u\]

The score will be above zero only in Case 1:

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

The variance of the score can be expressed as $E \left[ \left( \Phi(x_1 + \alpha_0) (\Phi(x_1 + \alpha_0) - \Phi(x)) \right)^{-1} \phi(x_1 + \alpha_0)^2 \right]$. 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) (\Phi(x_1 + \alpha_0) - \Phi(x))} \right] \geq 2 \int_0^1 \frac{du}{\sqrt{u}} \int_{-K}^0 dx_1 \to +\infty, \quad \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 $\alpha_0$ in this parametric model is equal to zero.

### E Identification of the distribution of unobserved heterogeneity.

In this section we demonstrate how we attained identification of the distribution of the unobserved heterogeneity terms in both the triangular and non triangular systems. Our results will be based on using the deconvolution methods.

**Theorem E.1** The distribution of unobserved heterogeneity terms $(U, V)$ $G(\cdot)$ in the triangular model with incomplete information (3.1) under Assumptions 1 and 3 is non-parametrically identified from the observed distribution of the data $P(Y_1, Y_2, X_1 \leq x_1, X \leq x)$.

**Proof:** 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 \( \alpha \). We note that

\[
E [Y_1 Y_2 | x, x] = P_{11}(x, x) = \int 1 \{ x_1 - u + \alpha \Phi \left( \frac{x - v}{\sigma} \right) \geq 0 \} \Phi \left( \frac{x - v}{\sigma} \right) g(u, v) \, du \, dv.
\]

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

\[
\int e^{-i t_1 y_1 - i t_2 y_2} 1 \{ y_1 + \alpha \Phi(y_2) \geq 0 \} \, dy_1 \, dy_2.
\]

We notice that function \( 1 \{ y_1 + \alpha \Phi(y_2) \geq 0 \} - \frac{1}{2} \) 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

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

Then we can consider the Fourier transform

\[
\int e^{-i t_1 z_1 - i t_2 z_2} \left( 1 \{ z_1 + \alpha \left( \Phi(z_2) - \frac{1}{2} \right) \geq 0 \} - \frac{1}{2} \right) \, dz_1 \, dz_2.
\]

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

\[
\int e^{-i t_1 z_1 - i t_2 z_2} \left( 1 \{ z_1 + \alpha \left( \Phi(z_2) - \frac{1}{2} \right) \geq 0 \} - \frac{1}{2} \right) \, dz_1 \, dz_2 = \frac{1}{it_1} \int e^{i t_1 \alpha (\Phi(z_2) - \frac{1}{2}) - i t_2 z_2} \, dz_2.
\]

The generalized function \( \Gamma(t_1, t_2; \alpha) = \int_{-\infty}^{\infty} e^{i t_1 \alpha (\Phi(z) - \frac{1}{2}) - i t_2 z} \, dz \) 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 \([-\frac{1}{2}, \frac{1}{2}]\), 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 \( \frac{2 \sin(\frac{\pi}{2} t_1)}{t_2} + 2 \cos\left(\frac{\pi}{2} t_1\right) \delta(t_2) \). We note that

\[
\lim_{t_1 \to 0} \frac{1}{it_1} [\Gamma(t_1, t_2; \alpha) - 2 \pi \delta(t_2)] = \alpha \int \left( \Phi(z) - \frac{1}{2} \right) e^{-i t_2 z} \, dz = \alpha \frac{\chi_{\Phi(t_2)}}{it_2}.
\]

The last useful property of this generalized function is that

\[
\int t_2 \Gamma(t_1, t_2; \alpha) \, dt_2 = 2 \pi i \int e^{i t_1 \alpha (\Phi(z) - \frac{1}{2})} \delta'(z) \, dz = -2 \pi t_1 \alpha \phi(0),
\]

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

\[
Q(t_1, t_2) = \left( \frac{\sigma e^{-\frac{1}{2} it_1 \alpha}}{it_1} \Gamma(t_1, \sigma t_2) + 2 \pi^2 \delta(t_1) \delta(t_2) \right) \chi_{uv}(t_1, t_2).
\]
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 \( R(\cdot, \cdot) \). 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 \( R(\cdot, \cdot) \), we can express the Fourier transform

\[
\int e^{-it_1y_1-it_2y_2} \mathbf{1}\{y_1 + \alpha \Phi(y_2) \geq 0\}\Phi(y_2)dy_1 dy_2
= \int e^{-it_1y_1-it_2y_2} \mathbf{1}\{z_1 \leq y_1\} \mathbf{1}\{z_2 \leq y_2\}dR(z_1, z_2) = -\frac{\chi_R(t_1, t_2)}{t_1t_2} (1 + \pi it_1 \delta(t_1))(1 + \pi it_2 \delta(t_2)).
\]

Considering the expectation \( P_{11}(x_1, x) = E[Y_1Y_2 | x_1, x] \) with the corresponding Fourier transform \( \mathcal{F}_{11}(t_1, t_2) \), we can express

\[
\mathcal{F}_{11}(t_1, t_2) = -\frac{\chi_R(t_1, \sigma t_2)}{t_1t_2} (1 + \pi it_1 \delta(t_1))(1 + \pi i\sigma t_2 \delta(t_2))
\chi_{uv}(t_1, t_2).
\]

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

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

This concludes our description of the Fourier transformations of the system of identifying equations for the triangular model with incomplete information. \textit{Q.E.D.}

\textbf{Theorem E.2} \textit{The distribution of the unobserved heterogeneity \((U, V)\) in the incomplete information game defined by (5.1) and (5.2) under Assumptions 4 and 7, and 8 is non-parametrically identified from the distribution of the data.}

\textit{Proof:} Consider 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

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

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

$$
\Delta_1(t_1, t_2) = \Phi\left(\frac{1}{\sigma} [t_1 + \alpha_1 \Delta_2(t_1, t_2)] \right) - \frac{1}{2},
$$
$$
\Delta_2(t_1, t_2) = \Phi\left(\frac{1}{\sigma} [t_2 + \alpha_2 \Delta_1(t_1, t_2)] \right) - \frac{1}{2},
$$

and if for some $(t_1, t_2)$ the pair $(\Delta_1, \Delta_2)$ 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

$$
\int e^{-it_1q_1-it_2q_2}P_1(q_1, q_2), dq_1 dq_2
$$
$$
= \int e^{-it_1q_1-it_2q_2}1\{z_1 - q_1 \leq 0\}1\{z_2 - q_2 \leq 0\}p_1(z_1, z_2)dz_1 dz_2 dq_1 dq_2
$$
$$
= -\chi p_1(t_1, t_2) \frac{1}{t_1 t_2} (1 + \pi it_1 \delta(t_1)) (1 + \pi it_2 \delta(t_2)).
$$

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

$$
\int e^{-it_1q_1-it_2q_2}P_1(q_1, q_2)P_2(q_1, q_2), dq_1 dq_2
$$
$$
= -\chi p_1 p_2 (t_1, t_2) \frac{1}{t_1 t_2} (1 + \pi it_1 \delta(t_1)) (1 + \pi it_2 \delta(t_2)).
$$

We then consider the Fourier transform of the expectation $E[Y_1Y_2 | x_1, x_2]$, which we denote $\mathcal{F}_{11}(t_1, t_2)$, leading to the following expression for the density of the distribution of unobserved heterogeneity:

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

Q.E.D.
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:

\[ P_{11}(x_1, x) = E[y_1 y_2 | x_1, x] = \int 1\{x_1 - u + \alpha \Phi \left( \frac{x - v}{\sigma} \right) > 0\} \Phi \left( \frac{x - v}{\sigma} \right) g(u, v) \, du \, dv, \]

\[ P(x_1, x) = E[y_2 | x_1, x] = \int \Phi \left( \frac{x - v}{\sigma} \right) g_v(v) \, dv, \]

\[ Q(x_1, x) = E[y_1 | x_1, x] = \int 1\{x_1 - u + \alpha \Phi \left( \frac{x - v}{\sigma} \right) > 0\} g(u, v) \, du \, dv. \]

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

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

\[ = E[y_1 - \int \Phi \left( \frac{x - v}{\sigma} \right) g(u, v) \, du \, dv | x_1, x], \]

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

\[ = E[\rho_1(y, x; \alpha, g) | x_1, x], \]

\[ m_3(x_1, x; \alpha, g) = E[y_2 - \int \Phi \left( \frac{x - v}{\sigma} \right) g_v(v) \, dv | x_1, x] = E[\rho_2(y, x; \alpha, g) | x_1, x]. \]

Consider the derivatives of these moment equations with respect to parameter \(\alpha\):

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

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

\[ \frac{dm_3}{d\alpha} = 0. \]
Then considering the space of densities that are uniformly manageable and satisfy Assumption [4], we take a direction in this space $h$ and

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

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

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

We introduce the vector with elements

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

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

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

and denote

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

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

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

and

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

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

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

We express

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

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

$$
\Sigma(x_1, x) = P_{11} \left( I - \begin{pmatrix} P_{11} & Q \frac{Q(1-Q)}{P_{11} P Q} \\ P & \frac{P Q}{P_{11}} \frac{P Q}{P_{11} (1-P)} \end{pmatrix} \right) (F.11)
$$
The semiparametric efficiency bound will be associated with the “least favorable” direction $h$. To find this direction one needs to solve the minimization problem

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

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

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

where $\lambda$ 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 countable 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. The 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: $E[Y_1|x_1, x_2]$, 

[...省略部分内容...]
derivatives of the equilibrium beliefs with respect to the parameters as:

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

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

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

Then considering the space of densities that are uniformly manageable we take a direction in this space \( h \) and obtain

\[ \frac{d m(x_1, x_2; \alpha, g)}{d\alpha'}[h] = \int \psi(x_1 - u, x_2 - v)h(u, v)\, du\, dv, \]

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

\[ m_1(x_1, x_2; \alpha, g) = E\left[Y_1Y_2 - \int P_1(X_1 - u, X_2 - v)P_2(X_1 - u, X_2 - v)\times g(u, v)\, du\, dv \bigg| x_1, x_2 \right] = E[p_1(Y, \alpha, g)|x_1, x], \]

\[ m_2(x_1, x_2; \alpha, g) = E\left[Y_1 - \int P_1(X_1 - u, X_2 - v)g(u, v)\, du\, dv \bigg| x_1, x_2 \right] = E[p_2(Y, \alpha, g)|x_1, x_2], \]

\[ m_3(x_1, x_2; \alpha, g) = E\left[Y_2 - \int P_2(X_1 - u, X_2 - v)g(u, v)\, du\, dv \bigg| x_1, x_2 \right] = E[p_3(Y, \alpha, g)|x_1, x], \]

Under our assumption regarding the distribution of errors \( \eta_1 \) and \( \eta_2 \), equilibrium beliefs are monotone functions of the parameters. Previously, we derived the Jacobian matrix corresponding to the derivatives of the equilibrium beliefs with respect to the parameters as:

\[
J^\alpha = \begin{pmatrix}
\frac{\partial P_1}{\partial \alpha_1} & \frac{\partial P_2}{\partial \alpha_1} \\
\frac{\partial P_1}{\partial \alpha_2} & \frac{\partial P_2}{\partial \alpha_2}
\end{pmatrix}
= \frac{\alpha_1\alpha_2}{1 + \alpha_1\alpha_2\alpha_1\alpha_2}
\begin{pmatrix}
P_1/a_2 & \alpha_1P_1 \\
\alpha_2P_2 & P_2/a_1
\end{pmatrix},
\]

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

We can express the Jacobian matrix of the moment vector \( m(\cdot) \) with respect to the finite-dimensional parameters \( \alpha_1 \) and \( \alpha_2 \) as:

\[
\frac{d m(x_1, x_2; \alpha, g)}{d\alpha'} = \int \begin{pmatrix} P_2 & P_1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = (\mu_1, \mu_2). \]

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

\[ \frac{d m(x_1, x_2; \alpha, g)}{d\alpha'}[h] = \int \psi(x_1 - u, x_2 - v)h(u, v)\, du\, dv, \]
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 \( D_{h_i}(x_1, x_2) = \frac{dm(x_1,x_2;\alpha,g)}{da_i} - \frac{dm(x_1,x_2;\alpha,g)}{dg} \left[ h_i \right] \) 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

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

where \( \lambda \) 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

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

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

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