Information Structure and Statistical Information in Discrete Response Models[1]

Shakeeb Khan

Department of Economics, Duke University

Denis Nekipelov

Department of Economics, UC-Berkeley

This version: September 2011

Preliminary and Incomplete.

ABSTRACT

Discrete response models are of high interest in economics and econometrics as they encompass treatment effects, social interaction and peer effect models, and discrete games. We study the impact of the structure of information sets of economic agents on the *Fisher information* of (strategic) interaction parameters in such models. While in *complete* information models the information sets of participating economic agents coincide, in *incomplete* information models each agent has a type, which we model as a payoff shock, that is not observed by other agents. We allow for the presence of a payoff component that is common knowledge to economic agents but is not observed by the econometrician (representing unobserved heterogeneity) and have the agents' payoffs in the incomplete information model approach their payoff in the complete information model as the heterogeneity term approaches 0. We find that in the complete information models, there is *zero Fisher information* for interaction parameters, implying that estimation and inference become nonstandard. In contrast, positive Fisher information can be attained in the incomplete information models with any non-zero variance of player types, and for those we can also find the semiparametric efficiency bound with unknown distribution of unobserved heterogeneity. The contrast in Fisher information is illustrated in two important cases: treatment effect models, which we model as a triangular system of equations, and static game models. In static game models we show this result is not due to *equilibrium refinement* with an increase in incomplete information, as our model has a fixed equilibrium selection mechanism. We find that the key factor in these models is the relative tail behavior of the unobserved component in the economic agents' payoffs and that of the observable covariates.

JEL Classification: C35, C14, C25, C13.

Keywords: Endogeneity, semiparametric efficiency, optimal convergence rate, strategic response.

---

1 Introduction

Endogenous regressors are frequently encountered in econometric models, and failure to correct for endogeneity can result in incorrect inference. With the availability of appropriate instruments, two-stage least squares (2SLS) yields consistent estimates in linear models without the need for making parametric assumptions on the error disturbances. Unfortunately, it is not theoretically appropriate to apply 2SLS to non-linear models, as the consistency of 2SLS depends critically upon the orthogonality conditions that arise in the linear-regression context.

Until recently, the standard approach for handling endogeneity in certain non-linear models has required parametric specification of the error disturbances (see, e.g.,Heckman (1978), Blundell and Smith (1989), and Rivers and Vuong (1988)). A more recent literature in econometrics has developed methods that do not require parametric distributional assumptions, which is more in line with the 2SLS approach in linear models. In the context of the model considered in this paper, existing approaches depend critically upon the form of the endogenous regressor(s).[1]

For continuous endogenous regressors, a "control-function approach" has often been been proposed by Blundell and Powell (2004) for many nonlinear models, and, without linear-index and separability restrictions, Imbens and Newey (2009). With these approaches, often a linear model specifies a relationship between the continuous endogenous regressors and the full set of exogenous covariates (including the instruments). The first-stage estimation yields estimates of the residuals from this model, which are then plugged into a second-stage estimation procedure to appropriately "control" for the endogenous regressors. The control-function approach, however, requires the endogenous regressors to be continuously distributed.

Consequently, these approaches are inapplicable to the models we study in this paper, which focuses on simultaneous discrete response models with discrete endogenous variables. Identification and inference in these models becomes much more complicated than in the continuous case, as illustrated in the important work in Cheshire (2005), who considers a general class of nonlinear, nonseparable models. Given his findings of lack of point identification in the discrete setting, he adopts a partial identification approach, and finds that sharper bounds can be found the more support points the endogenous variable has.

---

[1]Several papers have considered estimation in the presence of endogeneity under additional assumptions. These include Lewbel (1998), Hong and Tamer (2003)

The class of models we consider will include many important special cases that have received a great deal of attention in both theoretical and empirical work. Important examples include treatment effect models, models of social interactions, and game theory models. In these models the parameter of interest will be the coefficient on the discrete endogenous variables. In one class of models we consider, where the system of equations is triangular this parameter is directly related to the average treatment effect (ATE). In the other class we study, which nests social interaction and game theory models, this parameter is often referred to as the "interaction" term.

In this paper we also consider both the identification of such parameters, and what we are particularly interested in here is the *Fisher information* for such parameters in this class of models. We find a fundamental relationship between the choice-theoretic information sets of the agents (reflecting their knowledge regarding the types of their opponents) and Fisher information for the strategic interaction parameters. To demonstrate our finding we consider a complete information model where the agents have perfect knowledge regarding payoffs and then we consider a incomplete information model by introducing the types of economic agents as random shocks to their payoffs. In this setting the incomplete information model is embedding[2] the complete information model, given that the payoffs of agents in the incomplete information model will converge to their payoffs in the complete information model if the variance of their types approaches zero. The payoff shocks in this setting is private information of the agents.

In the incomplete information setting private information of economic agents can be treated as an additional source of unobserved heterogeneity in the model. The presence of private information adds the term that is not observable both to competing economic agents and to the econometrician. As we will show, it the incomplete information models positive Fisher information is possible for the parameter interest (i.e. treatment effect or interaction term), whereas for the complete information models with common knowledge of the agents information is 0.

In recent econometrics work on inference in static game models (see, for instance, Bajari, Hong, Krainer, and Nekipelov (2010b), Aradillas-Lopez (2010) and de Paula and Tang (2011)) demonstrate why identification of the interaction parameters can be easier in games with incomplete information because regions of the data can be found where a unique equilibrium exists. However, in this paper we argue that the increased information is not due to equilibrium refinement. This fundamental point is demonstrated in two ways.

---

[2]In the terminology of Kajii and Morris (1997)

First, we consider a triangular system of discrete response models, which is often to model treatment effects in policy evaluation programs. These models, which are generally coherent, do not suffer from multiple equilibria. Nonetheless, we find a stark contrast in the complete and incomplete cases. For the complete case model, which has been studied in many papers, including Vytlacil and Yildiz (2007), Klein, Shan, and Vella (2011), Abrevaya, Hausman, and Khan (2011), Jun, Pinkse, and Xu (2011), we establish 0 information for the treatment effect parameter. This implies that inference becomes nonstandard and difficult for the treatment effect . In that sense the weak identification of this parameter is similar to its identification in the model introduced in Lewbel (1998), for which 0 information was shown in Khan and Tamer (2010). Therefore, for the triangular system we measure information levels in terms of optimal rates of convergence for the treatment effect parameter. This method of quantifying information is often used for nonstandard models such as those in Stone (1982), Horowitz (1993), and more recently, Menzel and Morgantini (2009) and Linton, Komarova, and Srisuma (2011). As we show, optimal rates will be directly related to the relative tail behavior of unobservable and observable variables.

In these treatment effect models the incomplete information structure involves an additional additive unobserved heterogeneity term. This is a new model, and corresponds to the agent deciding to comply or not with the treatment *before* the treatment is assigned. The treatment assignment remains uncertain to the agent until after he or she makes the decision to comply. In that sense, the additional noise plays a loosely analogous role to what a "placebo" usually plays in the natural sciences in aiding with inference on treatment effects. In the game-theoretic terms, this model can be described as a game between the two players, the case player one is the treated individual and player 2 is the one assigning treatment. However, the "type" player 2 is not completely known to player 1. Then the decision problem of player 1 can be considered in the same Bayesian framework as in the standard static game of incomplete information. Our main finding is that positive information for the treatment effect parameter can be attained, and we derive the semiparametric efficiency bound for this parameter. Modeling the incomplete information model this way, we can show that the complete information model can be viewed as a limiting case of the incomplete information model, where the variance of the unobserved heterogeneity term converges to 0. We demonstrate this by showing that the Fisher information of the incomplete information model, when expressed as a function of this variance, converges to 0 as the heterogeneity variance does.

Consideration of the triangular model allows us to build a case for the strategic interac-

tion models, which involve nontriangular systems, arguing that our result regarding Fisher information for strategic interaction parameters is not an artifact of equilibrium refinement. These interaction models, which include static game theory models as a leading case, (see, e.g. Bjorn and Vuong (1985), Bresnahan and Reiss (1990), Bresnahan and Reiss (1991b), Bresnahan and Reiss (1991a), Tamer (2003), Andrews, Berry, and Jia (2004), Berry and Tamer (2006), Pakes, Ostrovsky, and Berry (2007), Ciliberto and Tamer (2009), Bajari, Hong, and Ryan (2010), Beresteanu, Molchanov, and Molinari (2011)), are well known to suffer from problems of incoherency and multiple equilibria. In this class of models our complete and incomplete characterization correspond to agents playing complete and incomplete information games, respectively.

Fundamental work in game theory, for instance, in Fudenberg, Kreps, and Levine (1988), Dekel and Fudenberg (1990), Kajii and Morris (1997) has focused on studying convergence of equilibria in the game of incomplete information to those in the limiting the game of complete information with the reduction of the dispersion in the private information of players. It was established that only particular equilibria in the complete information game will be robust to adding noise to the players' payoffs.[3] As a result, introduction of incomplete information may result in equilibrium refinement.

However, we find that in this paper, while positive Fisher information for the interaction parameter can be attained for the interaction parameter in the incomplete information case, and 0 Fisher information is found for the complete information case, this is not because of equilibrium refinement. We do so by assuming the simplest *equilibrium selection rule*, so that the model is coherent and multiple equilibria is not a problem in both the complete and incomplete cases.[4] Nonetheless, we find that in the complete case the information for the interaction parameters is 0, whereas for the incomplete case can be positive. So as with the triangular system, for the complete case, where identification is irregular, we quantify information with optimal rates of convergence, and for the incomplete case we derive the semiparametric efficiency bound for the interaction parameter. Furthermore, we show that the Fisher information in the complete information model can be viewed as the limiting case of the Fisher information in the incomplete information models. We recognize that this result *does not* imply the convergence of equilibria in the incomplete information game to those in the incomplete information games. Instead, we claim that the statistical information

---

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

[4]As a result, the zero Fischer information is not a result of failure of the rank condition as in Bajari, Hahn, Hong, and Ridder (2010), but is rather a determined by the distribution of unobserved payoff components.

regarding strategic interaction parameters contained in observed distribution of equilibrium actions in the incomplete information game converges to the statistical information contained in the data of equilibrium actions in the complete information games.

The rest of the paper is organized as follows. The following section introduces the base model we consider- a binary choice with one endogenous binary variable, for which to complete the specification of the (triangular) model, a reduced-form model is utilized for the endogenous regressor. Our first finding is the zero information for the parameter of interest, which is a result similar to that found in Khan and Tamer (2010) for the binary choice model with endogeneity in Lewbel (1998), and related to that found in Chamberlain (1986) and Chen and Khan (1999) for heteroskedastic binary choice models. As this result implies the difficulties with inference for the parameter of interest, we further explore possible asymptotic properties for conducting inference in this model. We then consider the triangular system with incomplete information, which is the strategic behavior model with the agent playing against nature where nature has a type that is not observed by the agent. For the incomplete information framework we show in Section 3 that Fisher information for the parameter of interest is positive, and as such inference becomes more standard. So we derive the semiparametric efficiency bound for the interaction parameter. In Section 4 we explore nontriangular systems, and we begin with simplest models: a 2 equation simultaneous game where the incoherency/multiplicity of equilibria is resolved via an equilibrium selection rule. Nonetheless, we show this "simplified" model has 0 information for the parameter of interest. As with the triangular system, inference becomes complicated, even though the parameter is point identified, so we explore this issue further. Then in Section 5 we consider the model where each player has a type, represented by a random shock to her payoff. The types of opponents are not observed by the players, which produces the setup of the game of incomplete information. This incomplete information game embeds the complete information game in Section 4 provided that the payoffs of players in the incomplete information game will be identical to those in the complete information game if the variance of types of players is set equal to zero. The presence of random payoff perturbations does not completely resolve the problem of multiple equilibria, but by introducing an equilibrium selection mechanism (as we did for the complete information game), we can now attain positive information for the interaction parameter. The contrast illustrates that positive information has to do with more than equilibrium refinement, as neither of the models was multiple equilibria an issue. Finally, Section 6 concludes the paper by summarizing and suggesting areas for future research. An appendix collects all the proofs of the theorems.

2 Discrete response model

2.1 Information in discrete response model

Let y_1 denote the dependent variable of interest, which is assumed to depend upon a vector of covariates z_1 and a single endogenous variable y_2.

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

$$y_1 = \mathbf{1}\{z_1'\beta_0 + \alpha_0 y_2 - u > 0\}. \tag{2.1}$$

Turning to the model for the endogenous regressor, the binary endogenous variable y_2 is assumed to be determined by the following reduced-form model:

$$y_2 = \mathbf{1}\{z'\delta_0 - v > 0\}, \tag{2.2}$$

where $z \equiv (z_1, z_2)$ is the vector of "instruments" and (u, v) is an error disturbance. The z_2 subcomponent of z provides the exclusion restrictions in the model. z_2 will only required to be nondegenerate conditional on $z_1'\beta_0$. We assume that (u, v) is independent of z. Endogeneity of y_2 in (2.1) arises when u and v are not independent of each other. Estimation of the model in (2.2) is standard. When dealing with a binary endogenous regressor, we will use the common terminology "treatment effect" rather then referring to the "causal effect of y_2 on y_1." Thus, for example, a positive treatment effect would correspond to the case of equation (2.1) where y_2 can take on only two values.

This type of model fits into the class of models considered in Vytlacil and Yildiz (2007) In this paper of particular interest in this section is the parameter α_0 which is related to a treatment effect parameter. Thus to simplify exposition, we will assume the parameters δ_0, β_0 are known. What this paper will focus on is the *information* for α_0- see, e.g. (Ibragimov and Has'minskii 1981), Newey (1990), Chamberlain (1986) for the relevant definitions. Our first result is that there is 0 information for α_0 which we state in the following theorem:

Theorem 2.1 *Suppose the model is characterized by the two equations above, and suppose that w.l.o.g., z has full support on R^k, then Fisher information associated with parameter α_0 is zero.*

Thus we can see, that under our conditions the parameter α_0 cannot be estimated at the parametric rate. This result is analogous to impossibility theorems in Chamberlain (1986).

Remark 2.1 *This result was alluded to, but not formally proven in Abrevaya, Hausman, and Khan (2011), whey they conducted inference on the sign of α_0, and indicated why the positive information found in Vytlacil and Yildiz (2007) was due to a relative support condition on unknown parameters that they imposed. The delicacy of point identification was also made apparent in Shaikh and Vytlacil (2011), who partially identified this parameter. As we will see later in this paper, this zero information result can be overturned by introducing a little more uncertainty in the model, by reducing the information available to the treated agent (Player 1) in the game.*

The fact that the information associated with the "interaction" parameter is zero, does not mean that the parameter cannot be estimated consistently. Now we will describe the set of results regarding rates of the semiparametric estimator for α_0.

2.2 Optimal rate for estimation of the interaction parameter

In the subsequent discussion we will not be interested in estimating parameters of linear indices δ_0 and β_0. To simplify the notation, we introduce the notation for the single indices $x_1 = z_1'\beta_0$ and $x = z'\delta_0$. Then the discrete response model can be written in the simplified form as

$$y_1 = \mathbf{1}\{x_1 + \alpha_0 y_2 - u \geq 0\},$$
$$y_2 = \mathbf{1}\{x - v \geq 0\}.$$

We take a constructive approach to establishing the rate result for the estimator of α_0 and before starting the analysis we define the optimal rate following the definition in Ibragimov and Has'minskii (1978). Let $\mathcal{G}$ characterize a class of joint densities of error terms (u, v). First, we recall that for the class of distributions $\mathcal{G}$, we define the maximal risk as

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

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

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

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

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

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

(iii) *Then r_n is the minimax (or optimal) rate if it is both a lower and an upper rate.*

We choose to derive the upper convergence rate using the constructive approach and ac-tually provide an estimator whose associated risk satisfies the property in Definition 2.1(ii). The rate of the resulting estimator relies on the tail behavior of the joint density of error distribution. To be more specific regarding the class of considered error densities, we for-mulate assumptions that restrict the "thickness" of tails of the error distribution and the smoothness of the density in the mean-square norm.

Our leading assumption highlights the class of densities that has "appropriate" tail be-havior. As we demonstrate in Appendix C, this class includes many commonly used error distributions such as normal and logistic.

Assumption 1 *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 $|t| < \infty$*

$$\lim_{c \to \infty} \frac{1}{\nu(c)} \int_{-c}^{c} \int_{-c}^{c} \left[G(x_1 + t, x)^{-1} + (1 - G(x_1 + t, x))^{-1} \right] f(x_1, x) \, dx_1 dx < \infty$$

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

$$\lim_{c \to \infty} \frac{1}{\beta(c)} \int_{|x_1| > c} \int_{|x| > c} \left[\log G(x_1 + t, x) + \log (1 - G(x_1 + t, x)) \right] f(x_1, x) \, dx_1 dx < \infty$$

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

Assumption 1(ii) requires the population likelihood function of the model to be finite (provided that $\beta(\cdot)$ is a non-increasing function). In addition, if the support of the indices x_1 and x is restricted to a square with the edge of some large length c, the resulting restricted likelihood will be sufficiently close to the true population likelihood.

In the next assumption we impose restrictions on the joint density of errors. First, we require the density to be sufficiently smooth in the L_2 norm. Second, we require the density to have an approximation in a relatively simple Hilbert space. Both these will assure that the estimator for the non-parametric element of the model (the cdf of the joint distribution of errors) to have a sufficiently high convergence rate that will not interfere with the asymptotic properties of the interaction parameter. Following Kim and Pollard (1990), we refer to the class of densities satisfying our assumptions *uniformly manageable*. We give the formal definition of this class in the Appendix A.2.

The class of uniformly manageable densities of errors satisfying Assumption 1 characterize the class of error distributions that we will be considering in our analysis. This is a large class of functions admitting irregularities such as discontinuities in the density and we allow the supports of covariates to be "large" relative to the tails of the error distributions. In light of our constructive approach to derivation of the rates we first propose the estimator that turns out to attain the upper convergence rate.

We consider the following procedure to estimate α_0 which is formalized in Appendix B.0.1. First, we look at the probability of the outcome $(0,0)$ conditional on linear indices x_1 and x. This probability does not depend on the interaction parameter and its derivative with respect to linear indices will be equal to the joint error density. For instance, we can fit the orthogonal polynomials to the joint probability of the $(0,0)$ outcome and then directly differentiate it with respect to the arguments.

Second, when the density of errors is known, it can be directly substituted into the expression for the probabilities containing the interaction parameter which correspond to the outcomes $(1,1)$ and $(0,1)$. Then we can form the quasi-likelihood of the model which we trim using the trimming sequence c_n depending on the sample size. We do so to avoid the divergence of the Hessian of the log-likelihood function at large values of the argument. We define the estimator as the maximizer of quasi-likelihood

$$\hat{\alpha}_{0,n}^* = \operatorname{argmax}_\alpha \frac{1}{n} \sum_{i=1}^n l\left(\alpha; y_{1i}, y_{2i}, x_{1i}, x_i\right). \tag{2.3}$$

It turns out that for an appropriately selected trimming sequence, the maximizer of the constructed quasi-likelihood function will converge to the interaction parameter α_0 at an

optimal rate.

Next we establish the result regarding the convergence rate of the constructed estimator, where K is the number of terms in the orthogonal polynomial expansion of the density.

Theorem 2.2 *Suppose that sequence c_n is selected such that $\nu(c_n)/n \to 0$, $K^\alpha/\nu(c_n) \to 0$, $\nu(c_n)K^d/n \to \infty$. Then for any sequence $\hat{\alpha}_n$ with the function $\hat{l}(\alpha)$ corresponding to the maximant of (2.3) such that $\hat{l}(\hat{\alpha}_{0,n}) \geq \sup_\alpha \hat{l}(\alpha) - o_p\left(\sqrt{\frac{\nu(c_n)}{n}}\right)$ we have*

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

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

It turns out that the obtained estimator $\hat{\alpha}_{0,n}^*$ based on the optimal trimming of the distribution tails of error terms u and v delivers the upper convergence rate for the estimators of α_0. We formalize this result in the following theorem.

Theorem 2.3 *Suppose that $c_n \to \infty$ is a sequence such that $\frac{n\beta^2(c_n)}{\nu(c_n)} = O(1)$ with $n/\nu(c_n) \to \infty$, $\nu(c_n)/K \to \infty$ and $\nu(c_n)K^d/n \to \infty$. Then for this sequence $\sqrt{\frac{n}{\nu(c_n)}}$ is the upper rate for the estimator for α_0*

Having established the upper convergence rate, we need to find the lower rate in order to determine whether our procedure delivers the optimal convergence rate for α_0. To derive the lower rate of convergence rate we use the result from Koroselev and Tsybakov (1993). Denote the likelihood ratio $\Lambda(\mathcal{P}_1, \mathcal{P}_2) = \frac{dP_{\mathcal{P}_1}^n}{dP_{\mathcal{P}_2}^n}$. Then the following lemma is the result given in Koroselev and Tsybakov (1993).

Lemma 2.1 *Suppose that $\alpha_0^1 = \alpha(\mathcal{P}_1)$ and $\alpha_0^2 = \alpha(\mathcal{P}_1)$, and $\lambda > 0$ be such that*

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

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

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

10

Theorem 2.4 *Suppose that $c_n \to \infty$ is a sequence such that $\frac{n\beta^2(c_n)}{\nu(c_n)} = O(1)$ with $n/\nu(c_n) \to \infty$, $\nu(c_n)/K \to \infty$ and $\nu(c_n)K^d/n \to \infty$. Then for this sequence $\sqrt{\frac{n}{\nu(c_n)}}$ is the lower rate for the estimator for α_0*

We can notice that both upper and lower rates result from balancing the bias introduced by trimming and the degree of "explosiveness" of the inverse cdf of the errors at the tails. Using the results of Theorems 2.3 and 2.4 by definition we write the following corollary.

Corollary 2.1 *Suppose that $c_n \to \infty$ is a sequence such that $\frac{n\beta^2(c_n)}{\nu(c_n)} = O(1)$ with $n/\nu(c_n) \to \infty$, $\nu(c_n)/k \to \infty$ and $\nu(c_n)K^d/n \to \infty$. Then for this sequence $\sqrt{\frac{n}{\nu(c_n)}}$ is the optimal rate for the estimator for α_0.*

The above result helps illustrate how widely the rates can vary, depending on the tail propeties of both the observed indexes, and section C illustrates this by considering widely used parametric distributions such as the normal and logistic distributions.

3 Triangular model with incomplete information

3.1 Identification and information of the model

In the previous section we considered a classical triangular discrete response model and demonstrated that, in general, that model has zero Fisher information for interaction parameter α_0. Our results suggested that the optimal convergence rate for the estimator of the interaction parameter will be sub-parametric and its convergence rate depends on the relative tail behavior of the error terms (u, v). In this section we set up the model which can be arbitrarily "close" to the classical triangular model but have positive information.[5] We construct this model by adding small noise to the second equation in the triangular system. It turns out that adding arbitrarily small but positive noise to that equation discontinuously changes the optimal rate to the standard parametric rate. The motivation for this approach could be adding artificial noise to the treatment assignment in a controlled experiment. In that case the experimental subjects do not know the specific realizations of the experimental noise but know its distribution. As a result, they will be responding to the *expected* treatment instead of the actual treatment. Incorporating expectations as explanatory models is

---

[5]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),where they were interested in the sign of the treatment effect.

similar in spirit to work considered in Ahn and Manski (1993). In this way, we were 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).

Consider the model where the endogenous variable is defined by

$$y_2 = \mathbf{1}[x - v - \sigma\eta > 0].$$

We assume that η is strictly orthogonal to u, v, x_1 and x and comes from a known distribution with a cumulative distribution function $\Phi(\cdot)$. Variable y_1 reflects the response of agent who does not observe the realization of noise η but observes the error term v. As a result, the response in the first equation can be characterized as:

$$y_1 = \mathbf{1}[x_1 + \alpha_0 E_\eta[y_2|x, v] - u > 0]$$

as before the parameter of interest is α_0 for which we wish to derive the information. Therefore we can express the conditional expectation in the above term as $E_\eta[y_2|x, v] = \Phi((x - v)/\sigma)$. Thus, the constructed discrete response model can be written as

$$
\begin{aligned}
y_1 &= \mathbf{1}\{x_1 - u + \alpha_0 E[y_2|x, v] > 0\}, \\
y_2 &= \mathbf{1}\{x - v + \sigma\eta > 0\}.
\end{aligned}
\tag{3.1}
$$

This model has features of the continuous treatment model considered in Hirano and Imbens (2004), Florens, Heckman, Meghir, and Vytlacil (2008) and Imbens and Wooldridge (2009). However, while in the latter 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. In this respect, we consider even the triangular model in the previous section from the behavioral perspective, characterizing the optimal choice of an economic agent. This approach has been proven useful in the modern treatment effect literature such as Abadie, Angrist, and Imbens (2002), Heckman and Navarro (2004), Carneiro, Heckman, and Vytlacil (2010). Outside of the treatment effect setting, analysis of binary choice models with a continuous endogenous variable was also studied in Blundell and Powell (2004), who demonstrated the attainability of positive information for the coefficient on the endogenous variable.

We can illustrate the structure of the model using Figure 1. Panel (a) in Figure 1 corresponds to the classic binary triangular system and panes (b)-(d) correspond to the triangular system with incomplete information. 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 given quantile q, the probability of y_2 equal zero or one exceeds $1 - q$. With a decrease in the variance of noise in the second equation, for given q panels b to c will approaching to the figure on panel a.
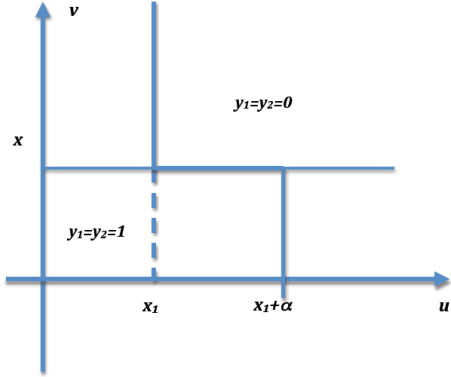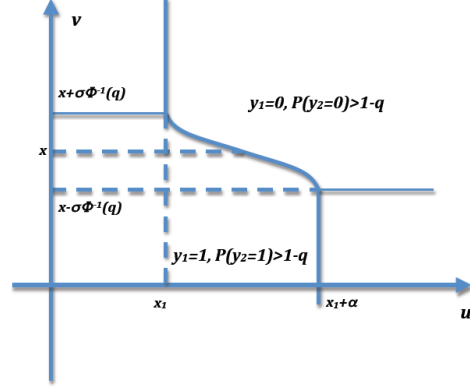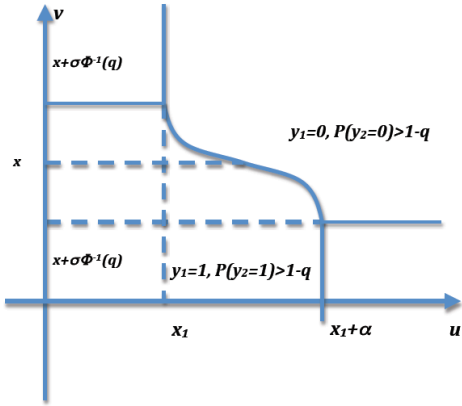

Figure 1.a

Figure 1.b

Figure 1.c
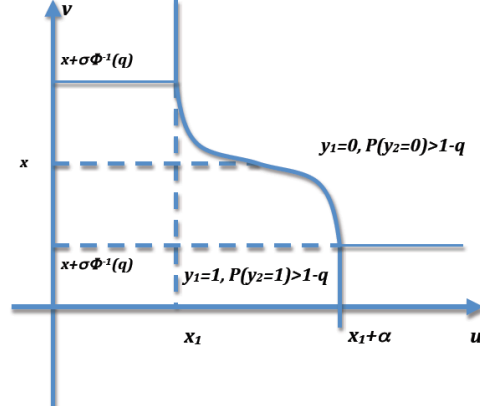
Figure 1.d

This discrete response model that is inspired by the payoff perturbation technique used in game theory. 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 payoff perturbation, but treatment perturbation model. Treatment perturbation can be considered in the experimental settings where the subjects can be exposed to the placebo treatment with some fixed probability, but they do not observe whether they get placebo or not. In that case they will respond to the expected treatment. The error terms u and v

in this setup can be interpreted as unobserved heterogeneity of the economic agent and the endogenous covariate.

Given that this is a new model, we will need to establish first that the model is identified from the data. The following theorem considers the identification of the interaction parameter α_0 along with the density of error terms $g(\cdot, \cdot)$.

Theorem 3.1 *Suppose that the joint density of error terms (u, v) has a characteristic function that is non-vanishing on its support, the error term η has a known distribution with absolutely continuous density and the probability of at least one pair of outcomes y_1 and y_2 conditional on the indices x_1 and x, has Fourier transformation that is non-zero on some compact set. Then the interaction parameter α_0 is identified.*

One more notable thing is that there is a tradeoff between identification of the marginal distribution of the error term v and the distribution of noise η. The observable conditional probability of the indicator of the second equation equal to one can be written as

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

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

Our identification argument for the fixed cdf of noise η is based on transforming the observed joint probabilities of indicators in both equations and the marginal probability of the indicator in the second equation. So, in addition to equation (3.2) we construct conditional probabilities $\mathbf{P}_{11}(x_1, x) = P(y_1 = y_2 = 1 \mid x_1, x)$ and $\mathbf{P}_{01}(x_1, x) = P(1 - y_1 = y_2 = 1 \mid x_1, x)$ determined by the distribution of treatment noise η and the unobserved components u and v. Then if $\mathcal{F}_{ij}(t_1, t_2)$ is the two-dimensional Fourier transform of the corresponding conditional probability and $\mathcal{F}_1(t_2)$ is the Fourier transform of the probability $\mathbf{P}_1(x)$ in equation (3.2), then we can explicitly express the interaction parameter α_0 as

$$\alpha_0 = \frac{\mathcal{F}_{11}(0, t_2) - \mathcal{F}_{01}(0, t_2)}{\mathcal{F}_1(t_2)} \frac{\int e^{-i\sigma t_2 x} \Phi(x) \, dx}{\int e^{-i\sigma t_2 x} \Phi(x)^2 \, dx} \tag{3.3}$$

We will further use this expression to derive the efficiency bound for the interaction parameter.

After establishing the identification of the model, we analyze its Fisher information. We summarize our result in the following theorem. It turns out that for any finite variance of noise η (which can be arbitrarily small) the information in the model of the incomplete information triangular model is strictly positive.

Theorem 3.2 *Suppose that the distribution of errors (u, v) has a characteristic function that is non-vanishing on its support and the conditional probability of at least one of the pairs of observable outcomes has a non-vanishing Fourier transformation. Then for any $\sigma > 0$ the information in the triangular model of incomplete information (3.1) is strictly positive.*

The last result of of this section demonstrates that from the inference viewpoint the incomplete information model has the information approaching zero when the variance of noise shrinks to zero. In other words, the less the informational asymmetry the economic agents have, the smaller the Fisher information corresponding to the interaction parameter α_0.

Theorem 3.3 *Suppose that the distribution of errors (u, v) has a characteristic function that is non-vanishing on its support and the conditional probability of at least one of the pairs of observable outcomes has a non-vanishing Fourier transformation. Then as $\sigma \to 0$ the information in the triangular model of incomplete information (3.1) converges to zero.*

3.2   Efficiency and convergence rate for the interaction parameter

We proved that the triangular model with incomplete information has positive Fisher information for any amount of noise added to the second equation. Now we consider derivation of the semiparametric efficiency bound for estimation of α_0. The calculations are partially based on the result for the semiparametric efficiency bound in conditional moment systems provided in (Ai and Chen 2003).

Theorem 3.4 *Denote*

$$a(t_2) = \left(\int e^{-i\sigma t_2 x} \Phi(x)^2 \, dx \right)^{-1} \int e^{-i\sigma t_2 x} \Phi(x) \, dx,$$

$$\Omega(x_1, x) = \begin{pmatrix} \mathbf{P}_{11}(1 - \mathbf{P}_{11}) & 0 \\ 0 & \mathbf{P}_{10}(1 - \mathbf{P}_{10}) \end{pmatrix}, \quad and \ T = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Finally, let

$$\zeta(x_1, x) = \left(f(x_1, x)^{-1} E\left[\mathbf{P}_1(X)\rho(X) \int e^{it_2(X-x)} a(t_2)\, dt_2 \right], \right.$$

$$\left. \rho(x) \int e^{it_2 x} a(t_2) \left(\mathcal{F}_{11}(0, t_2) - \mathcal{F}_{01}(0, t_2) \right)\, dt_2 \right).$$

Then the semiparametric efficiency bound for estimation of α_0 can be expressed as

$$\Sigma = E \left(\min_{\rho(\cdot)} \zeta(x_1, x)' T^{-1} \Omega(x_1, x)^{-1} T^{-1} \zeta(x_1, x) \right)^{-1}$$

Our efficiency result provides the semiparametric efficiency bound for the new discrete response model. In this model we allow agent-specific and treatment specific unobserved heterogeneity components to be fully nonparametric at a cost of parametrizing the noise distribution.

Our final result is related to the optimal convergence rate for the interaction parameter. Given that the information of the model is positive, the optimal convergence rate will be parametric. We formalize this in the following theorem which is a direct corollary of Theorem 3.4 and Theorem IV.1.1 in Ibragimov and Has'minskii (1981).

Theorem 3.5 *Under conditions of Theorem 3.4 for any sub-convex loss function $w(\cdot)$ and standard Gaussian element $\mathbb{G}$:*

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

where $g(\cdot, \cdot)$ is the distribution of errors u and v and $f(\cdot, \cdot)$ is the distribution of covariates.

This theorem establishes the parametric optimal convergence rate for the estimates of interaction parameters in the incomplete information game model.

4 Nontriangular Systems: A Static game of complete information

4.1 Information in the complete information game

In this section we consider information of parameters of interest in a simultaneous discrete system of equations where we no longer impose the triangular structure as the previous

section. A leading example of this type of system is a 2-player discrete game with complete information. See, e.g. the seminal papers in Bjorn and Vuong (1985), and Tamer (2003). We will later extend this model to one with incomplete information, in a way analogous to what we did for the triangular system.

A simple binary game of complete information is characterized by the players' deterministic payoffs, strategic interaction coefficients and the random payoff components u and v. Then the payoff of player 1 from choosing action $y_1 = 1$ can be characterized as a function of action of player 2

$$y_1^* = z_1' \beta_0 + \alpha_1 y_2 - u,$$

and the payoff of player 2 is characterized as

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

Further, for convenience of analysis we change notation to $x_1 = z_1' \beta_0$ and $x_2 = z_2' \delta_0$. We normalize the payoff from action $y_i = 0$ to zero. We make the following assumption regarding the information structure in the game.

Assumption 2 *The random payoff components (u, v) are mutually dependent with a distribution that has an absolutely continuous density with a full support. They are commonly observed by the players but not observed by the econometrician and $(u, v) \perp (x_1, x_2)$.*

Under this information structure the pure strategy of each player is the mapping from the observable variables into actions: $(u, v, x_1, x_2) \mapsto 0, 1$. A pair of pure strategies constitute a Nash equilibrium if they reflect the best responses to the rival's equilibrium actions. As a result, we can characterize the equilibrium by a pair of binary equations:

$$
\begin{aligned}
y_1 &= 1[x_1 + \alpha_1 y_2 - u > 0], \\
y_2 &= 1[x_2 + \alpha_2 y_1 - v > 0],
\end{aligned}
\tag{4.1}
$$

assuming that the errors u and v are correlated with each other with unknown distribution. In particular, we are interested in determining under which conditions those two parameters α_1, α_2 can be estimated at the parametric rate, and in situations where they cannot be, and which functions of the parameters can be.

As noted in Tamer (2003), this system of simultaneous discrete response equations 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.

17

Assumption 3 *Denote* $S_1 = [\alpha_1 + x_1, \, x_1] \times [\alpha_2 + x_2, \, x_2]$
and $S_2 = [x_1, \, \alpha_1 + x_1] \times [x_2, \, \alpha_2 + x_2]$. *Note* $S_1 = \emptyset$ *iff* $\alpha_1 > 0, \alpha_2 > 0$, *and* $S_2 = \emptyset$ *iff* $\alpha_1 < 0, \alpha_2 < 0$.

 (i) If $S_1 \neq \emptyset$ *then* $Pr\left(y_1 = y_2 = 1 | (\varepsilon, \eta) \in S_1\right) \equiv \frac{1}{2}$

 (ii) If $S_2 \neq \emptyset$ *then* $Pr\left(y_1 = (1 - y_2) = 1 | (\varepsilon, \eta) \in S_2\right) \equiv \frac{1}{2}$.

Assumption 3 suggests that when the system of binary responses has multiple solutions, then the realization of a particular solution is resolved over a symmetric coin flip. We select this simple setup *to emphasize our finding that the complete information has zero information holds even where there is no incoherency issue and equilibrium selection is fixed.* In principle, one can generalize this condition to the cases where the distribution over multiple outcomes depends on some additional covariates. However, given that the structure of results with that extension remains the same, we will not consider it in this paper.

First of all, we provide the result of identification of strategic interaction parameters, to argue that the zero information result is not a consequence of poor identifiability. Our identification result, generally speaking, is new. We leave the distribution of unobserved payoff components to be fully non-parametric (and non-independent, unlike Bajari, Hong, and Ryan (2010), who assume independence and normality of unobserved components u and v) while imposing a linear index structure on the payoffs.[6]

Theorem 4.1 *Suppose that the characteristic function of the unobserved payoff components is non-vanishing on its support, linear indices* x_1 *and* x_2 *have full support, and Assumptions 2 and 3 are satisfied. Then the interaction parameters* α_1 *and* α_2 *are identified.*

Having established the identifiability of the parameters of interest, we now study the information associated with the strategic interaction parameters. The following result establishes that the information associated with the interaction parameters in the static game of complete information is zero. The important note here is that in the light of 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.

---

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

Theorem 4.2 *Suppose that the characteristic function of the unobserved payoff components is non-vanishing on its support, linear indices x_1 and x_2 have full support, and Assumptions 2 and 3 are satisfied. Then Fisher information associated with parameters α_1 and α_2 is zero.*

This fully illustrates why zero interaction of the interaction parameter is a problem that is distinct from the multiple equilibria problem in these models. That is because we have explicitly completed the model in an ad-hoc way so that it is coherent , yet we still cannot attain positive information. Therefore, even for the simplified model, estimation and inference of the interaction parameters is complicated, and inference becomes nonstandard, which is analogous to what we found for the triangular system in the previous section. This suggests alternative inference methods or different(less informative) parameters to be estimated. Here we aim to quantify the information of the model in the same we did before, by attaining optimal rates for this parameter. As we show in the next section, provided that the identification in this case relies on the full support of linear indices, the optimal rate of convergence for the estimator of the interaction parameters will be sub-parametric and reflect the relative tail behavior of the distribution of the unobserved payoff components.

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 1 to account for the presence of the interaction between both discrete response equations.

Assumption 4 *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 $|s| < \infty$

$$\lim_{c \to \infty} \frac{1}{\nu(c)} \int\limits_{-c}^{c} \int\limits_{-c}^{c} \left[G(x_1 + t, x_2 + s)^{-1} + (1 - G(x_1 + t, x_2 + s))^{-1} \right] f(x_1, x_2) \, dx_1 dx_2 < \infty$$

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

$$\lim_{c \to \infty} \frac{1}{\beta(c)} \int\limits_{|x_1|>c} \int\limits_{|x_2|>c} \left[\log G(x_1 + t, x_2 + s) \right.$$

$$\left. + \log\left(1 - G(x_1 + t, x_2 + s)\right) \right] f(x_1, x_2) \, dx_1 dx_2 < \infty$$

In principle, we can consider a generalized version of Assumption 4 where we allow different behavior of the distribution tails in the strategic responses of different players. In that case we will need to select the trimming sequences differently for each equation. This will come at a cost of more tedious algebra. However, the conceptual result will be very similar.

We will use the assumption regarding the class of unobserved payoff components u and v with minimal modifications and we will not reproduce it from Section 1. As in that assumption we require the density of errors to be sufficiently smooth in the L_2 sense. Thus, we require that density to belong to "uniformly manageable" class of functions (as per definition in Kim and Pollard (1990)). In our case this boils down to the function being representable in relatively simple Hilbert space with the approximation error having a polynomial decay in the number of used basis functions. In addition, we need to assume that the interaction parameter will not be interfering "too much" with the estimated distribution of errors. The following assumption is closely related with the structural assumption in Chernozhukov, Chen, Lee, and Newey (2010)

Assumption 5 *Consider the population log-likelihood $\mathcal{L}(\alpha_1, \alpha_2; g)$ which is maximized at $(\alpha_{1,0}, \alpha_{2,0}; g_0)$. Then there exists $\delta > 0$ such that for any $0 < \epsilon < \delta$ and $\max\{|\alpha_1 - \alpha_{1,0}|, |\alpha_2 - \alpha_{2,0}|\} < \epsilon$ and $\|g - g_0\|_{L_2} < \epsilon$:*

$$\mathcal{L}(\alpha_1, \alpha_2; g) - \mathcal{L}(\alpha_{1,0}, \alpha_{2,0}; g_0) = -C_1|\alpha_1 - \alpha_{1,0}|^2 - C_2|\alpha_2 - \alpha_{2,0}|^2 - C_3\|g - g_0\|_{L_2}^2 + o(\delta).$$

Assumptions 4, 5 that are uniformly manageable characterize distributions of errors that we will consider in our model. The error distributions used in empirical analysis of games such as normal and logistic fall into the considered function class.

As in the case of triangular model, we propose a constructive approach to analyzing the optimal rate for the estimators of the interaction parameters. The idea behind the estimation procedure in the case of triangular system was to use the case where both indicators are equal to zero which allows one to directly observe the cumulative distribution of errors. This approach will not be immediately available in case of the complete information game. In

fact, we we noted above, the outcome probability

$$\mathcal{P}_{00}(x_1, x_2; \alpha_1, \alpha_2) = \int\limits_{x_1}^{+\infty} \int\limits_{x_2}^{+\infty} g(u,v)\, du\, dv - \frac{1}{2} \int\limits_{x_1}^{x_1+\alpha_1} \int\limits_{x_2}^{x_2+\alpha_2} g(u,v)\, du\, dv,$$

which depends on the unknown parameters α_1 and α_2. We modify the estimator by substituting the two-step procedure by the iterated procedure where one can "profile out" the the unknown density of errors at each step of maximizing the likelihood with respect to the interaction parameter. Defining the sample log-likelihood

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

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

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

We provide the formal discussion of this estimator in Appendix B.0.2

Next we establish the result regarding the convergence rate of the constructed estimator.

Theorem 4.3 *Suppose that sequence c_n is selected such that $\nu(c_n)/n \to 0$, $K^\alpha/\nu(c_n) \to 0$, $\nu(c_n)K^d/n \to \infty$. Then for any sequence $\hat{\alpha}_n$ with the function $\hat{l}(\alpha)$ corresponding to the maximant of (4.2) such that $\hat{l}^p(\hat{\alpha}_n) \geq \sup_\alpha \hat{l}^p(\alpha) - o_p\left(\sqrt{\frac{\nu(c_n)}{n}}\right)$ we have*

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

This obtained result is analogous to the result regarding the rate of the semiparametric two-stage estimator that we propose for the case of triangular system. Provided that we assumed identical tail behavior for both error terms, the resulting rates for the interaction parameters are the same. As we discussed previously, if the tails of the error distributions were differently, the rate result can established by choosing different trimming sequences for x_1 and x_2.

Using the same arguments as in triangular system case, we can prove that the iterative estimator $(\hat{\alpha}_{1n}^*, \hat{\alpha}_{2n}^*)$ attains the optimal rate. We given this result in the following theorem which replicates Theorem 2.3 for the case of complete information game.

Theorem 4.4 *Suppose that $c_n \to \infty$ is a sequence such that $\frac{n\beta^2(c_n)}{\nu(c_n)} = O(1)$ with $n/\nu(c_n) \to \infty$, $\nu(c_n)/K \to \infty$ and $\nu(c_n)K^d/n \to \infty$. Then for this sequence $\sqrt{\frac{n}{\nu(c_n)}}$ is the upper rate for the estimator for $(\alpha_{1,0}, \alpha_{2,0})$*

Using the technique proposed in Koroselev and Tsybakov (1993) we can further prove that the derived rate also corresponds to the lower rate of convergence for the strategic interaction parameters. We conclude the argument by formulating the following theorem, which we give without proof which is completely analogous to the case of the triangular system.

Theorem 4.5 *Consider the model of the game of complete information in which the error distribution satisfies Assumptions 4, 7 and 5. Suppose that $c_n \to \infty$ is a sequence such that $\frac{n\beta^2(c_n)}{\nu(c_n)} = O(1)$ with $n/\nu(c_n) \to \infty$, $\nu(c_n)/k \to \infty$ and $\nu(c_n)K^d/n \to \infty$. Then for this sequence $\sqrt{\frac{n}{\nu(c_n)}}$ is the optimal rate for the estimator for strategic interaction parameters α_1 and α_2.*

One of the important takeaways from this result is that, contrary to the previous point of view, the optimal rate for estimating strategic interaction parameters is 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 as a special case of a familiar model of a static game of incomplete information. Theoretical results demonstrate that introduction of payoff perturbations leads to reduction in the number of equilibria.[7] Here we attain regular identification for the interaction parameter as well, but our argument is not one of equilibrium refinement. That is because, 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.

---

[7]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).

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

Assumption 6   *(i) Suppose that the error terms u and v are commonly observed by the players, but not observed by the econometrician, along with linear indices x_1 and x_2 which are observed both by the players and the econometrician. We assume that $(u, v) \perp (x_1, x_2)$. However u and v are not independent from each other.*

  (ii) The payoff perturbations are private information of each player, such that player 1 observes η_1 but not η_2 and player 2 observes η_2 but not η_1. Private information assumption means that $\eta_1 \perp \eta_2$ and both perturbations are independent from linear indices and error terms u and v.

  (iii) Finally, we assume that η_1 and η_2 have the same distribution with continuous density having full support with cdf $\Phi(\cdot)$ which is fixed and known both to the players and to the econometrician.

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 players participate in repeated realizations of the static game. In that case 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 component of the payoff shock that is correlated with their shock. The remaining element that cannot be learned from the replications of the game is the remaining noise components η_1 and η_2 whose distribution is normalized.

An alternative interpretation for this information structure is the that the payoff components u and v are *a priori* known to the players, but not to the econonometrician. The interaction of the players is considered in the experimental settings where the payoff noise (η_1, η_2) is introduced artificially by the experiment designer. For this reason its distribution is known both to the players and to the econometrician.

Assumption 6 lays the groundwork for the coherent characterization of the structure

of equilibrium in this game of incomplete information. First, the strategy of player i is a mapping from the observable variables into actions: $(x_1, x_2, u, v, \eta_i) \mapsto \{0, 1\}$. Second, player i forms the beliefs regarding the action of the rival. Provided that η_1 and η_2 are independent, the beliefs will only be functions of u, v and linear indices. Thus, if $P_i(x_1, x_2, u, v)$ are beliefs regarding actions of players 1 and 2 correspondingly, then the strategy, for instance, of player 1 can be characterized as

$$
\begin{aligned}
y_1 &= \mathbf{1}\{E\left[y_1^* \mid x_1, x_2, u, v, \eta_1\right] > 0\} \\
&= \mathbf{1}\{x_1 - u + \alpha_1 P_2(x_1, x_2, u, v) - \sigma\eta_1 > 0\}.
\end{aligned}
\tag{5.1}
$$

Similarly, the strategy of player 2 can be written as

$$
y_2 = \mathbf{1}\{x_2 - v + \alpha_2 P_1(x_1, x_2, u, v) - \sigma\eta_2 > 0\}.
\tag{5.2}
$$

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

To characterize the Bayes-Nash equilibrium in the considered simultaneous 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. In other words

$$
\begin{aligned}
P_1\left(x_1, x_2, u, v\right) &= E\left[\mathbf{1}\{x_1 - u + \alpha_1 P_2(x_1, x_2, u, v) - \sigma\eta_1 > 0\} \mid x_1, x_2, u, v\right], \quad \text{and} \\
P_2\left(x_1, x_2, u, v\right) &= E\left[\mathbf{1}\{x_2 - v + \alpha_2 P_1(x_1, x_2, u, v) - \sigma\eta_2 > 0\} \mid x_1, x_2, u, v\right].
\end{aligned}
$$

Taking into consideration independence of the noise terms η and the fact that their cdf is known, we can characterize the pair of equilibrium beliefs as a solution of the system of nonlinear equations:

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

Our informational assumption regarding the independence of the unobserved heterogeneity components u and v from payoff perturbations η_1 and η_2 was crucial to define a game with a coherent equilibrium structure. If we allow the 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 noise components. This would not allow for 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 types of players.

System of equations (5.3) can have multiple solutions.[8] To resolve the uncertainty over equilibria and maintain the symmetry with our discussion of the games of complete information, we assume that uncertainty over multiple possible equilibrium beliefs is resolved over the independent coin flips.

We note that the incomplete information model that we constructed embeds the complete information model in the previous section. In fact, when $\sigma \equiv 0$ then the payoffs in the incomplete information model are identical to those in the complete information model and there no elements that are not commonly observable by both players. We illustrate the transition from the complete to the incomplete information environment on Figure 2.When $\sigma = 0$, the actions of the players will be determined by u and v only. Figure 2.a. shows the regions with four possible pairs of actions. There is a region in the middle where both pairs of actions are optimal, leading to multiple equilibria. Then, with the introduction of uncertainty, we can only plot the probabilistic picture of actions of players (integrating over the payoff noise η_1 and η_2). Then we can represent the areas where specific action pairs are chosen with probability exceeding certain 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 on Figure 2.a.

---

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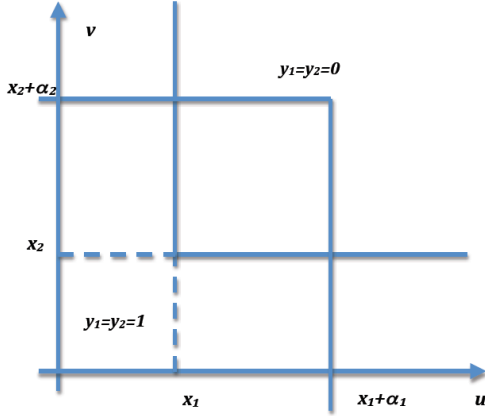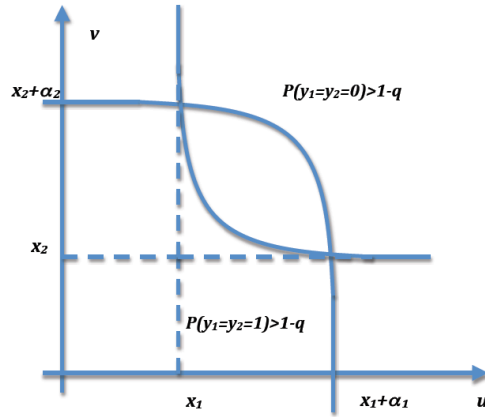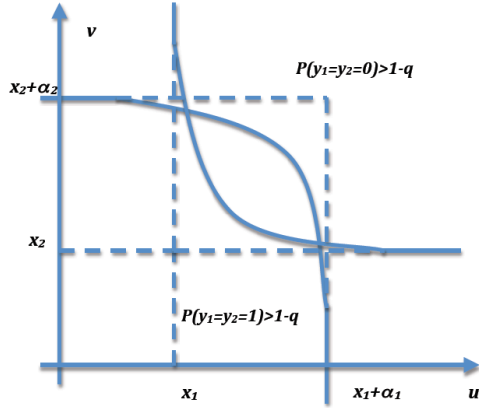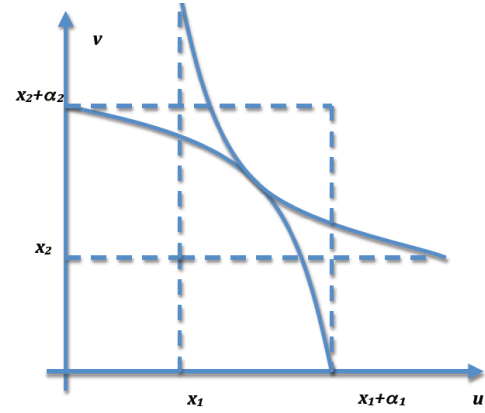
Figure 2.a

Figure 2.b

Figure 2.c

Figure 2.d

First, we establish the fact that the strategic interaction parameters α_1 and α_2 are identified in the given model along with the distribution of errors (u, v). Note that x_1, x_2, u and v enter the system of equations of interest in a special form. This means that the equilibrium beliefs will be functions of $x_1 - u$ and $x_2 - v$. We note that conditional on x_1, x_2, u and v the choices of two players are independent. On the other hand, given that the errors u and v are not observable to the econometrician conditional on x_1 and x_2 the choice will be correlated. As a result we define the object of interest as probabilities of observed pairs of outcomes $y_1 = y_2 = 1$ as well as $y_1 = 1 - y_2$. Denote $\mathbf{P}_{11}(x_1, x_2) = E[y_1 y_2 \mid x_1, x_2]$, $\mathbf{P}_{10}(x_1, x_2) = E[y_1(1 - y_2) \mid x_1, x_2]$, and $\mathbf{P}_{01}(x_1, x_2) = E[(1 - y_1)y_2 \mid x_1, x_2]$. We note that the moments depend on the unknown function $g(\cdot, \cdot)$, which, along with the distribution of covariates, is the infinite-dimensional parameter of the model.

We can show that this model is identified in a constructive way. Our approach will be

26

to eliminate the unknown density by "inverting" one of three observable expectations to get $g(\cdot, \cdot)$. To perform such an "inversion" we realize that the observed choice probabilities are represented by convolutions of the error density with the equilibrium beliefs. Recalling that Fourier transforms of convolutions are equal to the products of Fourier transforms of their components, we can recover the characteristic function of errors. Once the distribution of u and v is known, identification of the strategic interaction coefficients reduces to solving a standard parametric conditional moment-based problem.

Suppose that $\mathcal{F}_{11}(t_1, t_2)$ is a two-dimensional Fourier transform of the product of equilibrium beliefs $P_1(x_1 - u, x_2 - v)P_2(x_1 - u, x_2 - v)$. Then we note that the right-hand side of observable conditional probability

$$E[y_1 y_2 \mid x_1, x_2] = \int P_1(x_1 - u, x_2 - v)P_2(x_1 - u, x_2 - v)g(u, v)\, du\, dv$$

is a convolution of the joint error density and the choice probabilities. Performing standard deconvolution, can express the joint density of the unobserved heterogeneity components:

$$\bar{g}(u, v) = \int \int E\left[e^{i(t_1(u - x_1) + t_2(v - x_2))} \frac{\mathbf{P}_{11}(x_1, x_2)}{f(x_1, x_2)\mathcal{F}_{11}(t_1, t_2)}\right] dt_1\, dt_2,$$

where we use $\bar{g}(\cdot, \cdot)$ to emphasize that this is the error density recovered from the observed conditional expectation.

We used the first conditional moment equation to identify the density of errors in terms of the equilibrium choice probabilities and observed joint probability of the outcome (1,1). Now we use the remaining equation with the recovered joint error density to identify the interaction parameters of interest. The system of identifying equations can then be written as

$$\mathbf{P}_{10}(x_1, x_2) = \int P_1(x_1 - u, x_2 - v)\left(1 - P_2(x_1 - u, x_2 - v)\right)\bar{g}(u, v)\, du\, dv,$$
$$\mathbf{P}_{01}(x_1, x_2) = \int \left(1 - P_1(x_1 - u, x_2 - v)\right)P_2(x_1 - u, x_2 - v)\bar{g}(u, v)\, du\, dv.$$

$$(5.4)$$

Note that as the recovered density $\bar{g}(\cdot, \cdot)$ is computed also using the equilibrium choice probability, it will also depend on α.

In the following theorem we summarize our identification result.

Theorem 5.1 *Suppose that the distribution of errors (u, v) has a characteristic function that is non-vanishing on its support. Moreover, at least one outcome pair y_1 and y_2 with $Pr(y_1 = i, y_2 = j \mid q_1 = x_1 - u, q_2 = x_2 - v)$ has a Fourier transform that is not equal to zero*

in some compact set. Then strategic interaction terms α_1 and α_2 along with the joint density of error terms (u, v) 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 model. It turns out that for any finite variance of noise η (which can be arbitrarily small) the information in the model of the incomplete information game is not zero.

Theorem 5.2 *Suppose that the distribution of errors (u, v) has a characteristic function that is non-vanishing on its support and the conditional probability of at least one of the pairs of observable outcomes has a non-vanishing Fourier transformation. Then for any $\sigma > 0$ the information in the model of incomplete information game defined by (5.1) and (5.2) is strictly positive.*

The system of equilibrium choice probabilities can have multiple solutions. We can approach those cases by resolving the uncertainty regarding the equilibria via coin flips. 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 substituted by the mixture of possible equilibrium choice probabilities. The rest of the argument will remain unchanged.

New we provide the result that shows the behavior of Fisher information for the strategic interaction parameters as the variance of privately observed payoff shocks of players approaches zero. As in the case of incomplete information triangular model, in this case Fisher information of those parameters will approach to zero.

Theorem 5.3 *Suppose that the distribution of errors (u, v) has a characteristic function that is non-vanishing on its support and the conditional probability of at least one of the pairs of observable outcomes has a non-vanishing Fourier transformation. Then for as $\sigma \to 0$ the Fisher information in the model of incomplete information game defined by (5.1) and (5.2) approaches zero.*

5.2 Efficiency and convergence rate in the incomplete information game

Now we analyze the efficiency bound for the considered strategic interaction model. Provided that this is a semiparametric conditional moment model (with strategic interaction parame-

ters and unknown distribution of unobserved heterogeneity and covariates), the freamework of Ai and Chen (2003) directly applies here.

Theorem 5.4 *Let $J(\cdot, \cdot)$ correspond to the Jacobi matrix of the system of conditional moments (5.4), and $\Omega(x_1, x_2) = diag\left(\mathbf{P}_{10}(1 - \mathbf{P}_{10}), \mathbf{P}_{01}(1 - \mathbf{P}_{01}), \mathbf{P}_{11}(1 - \mathbf{P}_{11})\right)$. Introduce a two-dimensional $\rho(x_1, x_2)$ and denote*

$$\zeta_\rho(x_1, x_2) = (\rho(x_1, x_2), \quad E\left[\rho(X_1, X_2)\varphi(x_1 - X_1, x_2 - X_2)\right])$$

with

$$\varphi(x_1 - X_1, x_2 - X_2) = \int e^{i(t_1(x_1 - X_1) + t_2(x_2 - X_2))} \left(\frac{\mathcal{F}_1(t) - \mathcal{F}_{11}(t_1, t_2)}{\mathcal{F}_{11}(t_1, t_2)}, \quad \frac{\mathcal{F}_2(t) - \mathcal{F}_{11}(t_1, t_2)}{\mathcal{F}_{11}(t_1, t_2)}\right) dt_1 dt_2$$

Then, the semiparametric efficiency bound for strategic interaction parameters can be expressed as

$$\Sigma = \left(\min_{\rho \in \mathbf{H}} E\left[J(X_1, X_2)'\zeta_\rho(X_1, X_2)'\Omega(X_1, X_2)^{-1}\zeta_\rho(X_1, X_2)J(X_1, X_2)\right]\right)^{-1}.$$

Our efficiency result provides the semiparametric efficiency bound for the generalization of the class of static games of incomplete information in Bajari, Hong, Krainer, and Nekipelov (2010b) as well as (Haile, Hortaçsu, and Kosenok 2008) for games with quantal response equilibria considered in (Palfrey 1985). In addition to idiosyncratic errors η_1 and η_2 we allow for player-specific unobserved heterogeneity represented by u and v. 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. Notably 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. This structure allows us to analyze the game without additional assumptions regarding the formation of equilibria (except the assumption on equilibrium selection).

We conclude the analysis by the following theorem which is a direct corollary of Theorem 5.4 and Theorem IV.1.1 in Ibragimov and Has'minskii (1981).

Theorem 5.5 *Under conditions of Theorem 5.4 for any sub-convex loss function $w(\cdot)$ and standard Gaussian element $\mathbb{G}$:*

$$\liminf_{n \to \infty} \inf_{\hat{\alpha}_1^n, \hat{\alpha}_2^n} \sup_{f, g} E_{f, g}\left[w\left(\sqrt{n}(\hat{\alpha} - \alpha(f, g))\right)\right] \geq E[w\left(\Sigma^{1/2}\mathbb{G}\right)],$$

where $g(\cdot)$ is the distribution of errors u and v and $f(\cdot)$ is the distribution of covariates.

This theorem establishes the parametric optimal convergence rate for the estimates of inter-action parameters in the incomplete information game model.

6   Conclusions

This paper considers identification and inference in simultaneous equation models with discrete endogenous variables. The models studies include triangular systems where the parameter of interest, the coefficient on a particular discrete endogenous variable, related to the treatment effect in certain settings, and nontriangular systems, which include peer effect models and models of simultaneous discrete choice games. In these cases, the parameter of interest is often referred to as the interaction parameter. We divide the two sets into two classes, which we refer to as incomplete information and complete information, which are distinguished by the presence of an additional unobserved heterogeneity term. Our main findings are that the complete information models have 0 information under our conditions, whereeras the incomplete information model can have positive information. Our findings have important implications for both the triangular and nontriangular systems. In the triangular case, both the 0 information and the optimal rates we attain indicate little, if any advantage of estimating the parameter in this model when compared to estimating the simpler model proposed in Lewbel (1998). In the nontriangular case our 0 information result establishes that the difficulty in identifying the interaction parameter is not solely based on incoherency and multiple equilibria, as we find this result even after introducing an arbitrary equilibrium selection rule. What appears to drive positive information in the incomplete versions of the model is the support of the endogenous variable, which is convexifyed by the additional uncertainty introduced in the incomplete models.

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

BLUNDELL, R., AND R. SMITH (1989): "Estimation in a Class of Simultaneous Equation Limited Dependent Variable Models," *Review of Economic Studies*, 56(1), 37–57.

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

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

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

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

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

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

CHERNOZHUKOV, V., Z. CHEN, S. LEE, AND W. NEWEY (2010): "Local Identification for Semi- Parametric and Non-parametric Models," Working Paper.

CHESHIRE, A. (2005): "Nonparametric Identification under Discrete Variation," *Econometrica*, 73, 1525–1550.

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

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

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

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

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

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

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

HECKMAN, J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931–960.

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

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

HONG, H., AND E. TAMER (2003): "Inference in Censored Models with Endogenous Regressors," *Econometrica*, 71(3), 905–932.

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

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

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

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

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

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

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

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

KIM, J., AND D. POLLARD (1990): "Cube Root Asymptotics," *Annals of Statistics*, 18, 191–219.

KLEIN, R., C. SHAN, AND F. VELLA (2011): "Semi-Parametric Estimation of Sample Selection Models with Binary Selection Rules and Binary Outcomes," Georgetown University working paper.

KOROSELEV, A., AND A. TSYBAKOV (1993): *Minimax Theory of Image Reconstruction.* Springer-Verlag.

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

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

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

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

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

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

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

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

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

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

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

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

RIVERS, D., AND Q. VUONG (1988): "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, 49, 347–366.

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

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

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

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

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

Appendix

A   Proofs

A.1   Proof of Theorem 2.1

To simplify our arguments, we will assumed the regression coefficients β_0 and δ_0 are known. Consequently we will refer to the indexs in each equation as x_1, x respectively. We follow the approach in, e.g. Chamberlain (1986) by projecting the score with respect to the parameter of interest α_0 on the score with respect to a finite dimensional parameter in a *path-* i.e. a parameterized arc passing through the infinite dimensional parameter, in this case the bivariate density function of ϵ and η. We begin by characterizing the space of functions that the unknown bivariate density function, denoted here by g is assumed to lie in:

Definition A.1 *Let Γ consist of all densities g with respect to Lebesgue measure on $\mathbf{R}^2$ such that*

1. *$g : \mathbf{R}^2 \to \mathbf{R}$ is a positive, bounded continuously differentiable function.*

2. *$\int \int g(u, v) du dv = 1$. For each $v \in \mathbf{R}$, there is a function $q : \mathbf{R} \to \mathbf{R}$ such that $g(u, s) \leq q(u)$ for s in a neighborhood of v and $\int q(u) du < \infty$.*

3. *$\int |\partial g(u, v)/\partial u| du dv < \infty$*

Having defined the space of unctions we next define the set of paths we will work with:

Definition A.2 *Λ consists of the paths:*

$$\lambda(\delta_1, \delta_2) = g_0[1 + (\delta_1 - \delta_{10})h_1][1 + (\delta_2 - \delta_{20})h_2] \tag{A.1}$$

where g_0 is the "true" density function, assumed to lie in Γ and h_1, h_2 are each $:\mathbf{R}^1 \to \mathbf{R}$, continuously differentiable function that equal 0 outside some compact set and

$$\int \int g_0(u, v) h_1(u) h_2(v) du dv = 0 \tag{A.2}$$

With these definitions it will follow that $\lambda(\delta_1, \delta_2)$ will lie in Γ for δ_1, δ_2 in neighborhoods of δ_{10}, δ_{20}, respectively- see Chamberlain (1986). We proceed by expressing the likelihood function, noting the bivariate dependent variable can be one of four categories, (1,1),(1,0),(0,1)(0,0). We will denote each of those out comes by the the indicators $d_{ij}, i, j = 0, 1$. So, for example d_{11} denotes $I[y_1 = 1, y_2 = 1]$. We let $\mathcal{P}_{ij}$ $i, j = 0, 1$ denote the conditional probabilities of out comes as functions of parameters

and indexes, so for example $\mathcal{P}_{11}(\alpha, \delta_1, \delta_2) = \int \int I[u < x_1 + \alpha]I[v < x]g(u,v)dudv$. Note that this probability depends on δ_1, δ_2 because of our definition of $\lambda(\delta_1, \delta_2)$. Thus for a single observation our log likelihood can be expressed as

$$\sum_{i,j=0,1} d_{ij} \log \mathcal{P}_{ij}(\alpha, \delta_1, \delta_2) \tag{A.3}$$

We can then take the derivative of the above term with respect to α, evaluated at $\alpha = \alpha_0$ and $\delta_1 = \delta_{10}, \delta_2 = \delta_{20}$. To begin we first do this for one term in the summation, corresponding to $i = j = 1$. Conditioning on the indexes x_1, x this derivative can be expressed as

$$d_{11}\mathcal{P}_{11}(\alpha_0, \delta_{10}, \delta_{20})^{-1} \int \int \delta(u - x_1 - \alpha_0)I[v < x]dudv \tag{A.4}$$

where $\delta(\cdot)$ above denotes the Dirac delta function. The derivative with respect to δ_1 evaluated at $\alpha_0, \delta_{10}, \delta_{20}$ is of the form

$$d_{11}\mathcal{P}_{11}(\alpha_0, \delta_{10})^{-1} \int \int I[u < x_1 + \alpha_0]I[v < x]g_0(u,v)h_1(u)dudv \tag{A.5}$$

We next take the conditional expectation of the squared difference of the above two terms, which is of the form:

$$\mathcal{P}_{11}(\alpha_0, \delta_0)^{-1}\left(\int \int \delta(u - x_1 - \alpha_0)I[v < x]g_0(u,v)dudv \right.$$
$$\left. - \int \int I[u < x_1 + \alpha_0]I[v < x]g_0(u,v)h_1(u)dudv\right)^2 \tag{A.6}$$

To show our impossibility result we need to find an $h_1^*(u)$ that sets (A.6) to 0. Informally, this can be accomplished by setting

$$h_1^*(u) = \delta(u - x_1 - \alpha_0) \tag{A.7}$$

However, this would violate the smoothness conditions in Λ. Nonetheless, a smoothed version of $h_1^*(u)$ can make (A.6) arbitrarily small, but positive, and still satisfy the smoothness conditions in the definition of Λ. For example, one could replace the delta function in $h_1^*(u)$ with a **nascent delta function**. Typically a nascent delta function can be constructed in the following manner. Let ϕ be any continuously differentiable density function with support on the real line- e.g. that of the standard normal distribution. Then one would define the nascent delta function as: $\delta_\epsilon(x) = \phi(x/\epsilon)/\epsilon$. Note that $\lim_{\epsilon \to 0} \delta_\epsilon(x) = \delta(x)$. Thus we can take a mean value expansion of $\delta_\epsilon(x)$, around $\epsilon = 0$. This yields (using that $\phi(x) = -x\phi(x)$):

$$\delta_\epsilon(x) = \delta(x) + \phi(x/\epsilon^*)/(\epsilon^*)^4(x^2 + 1) \cdot \epsilon$$

where ϵ^* denotes an intermediate value between 0 and ϵ. Thus this remainder can be made as small as desirable for ϵ small enough. Therefore one could set $h_1^{*\epsilon}(u) = \delta_\epsilon(u - x_1 - \alpha_0)$. This would

make (A.6) positive but arbitrarily small and not violate our smoothness conditions. Note we have established this result without even considering the score with respect to δ_2. This because we can effectively set $h_2^*(v) = 0$.

So to make the squared distance 0, we set $h_1^*(u) = \delta(u - x_1 - \alpha_0), h_2^*(v) = 0$. But to complete the proof we need to show these choices of $h_1^{(}u), h_2^*(v)$ can also be used to the other two choice probabilities. We next turn attention to outcome $y_2 = y_1 = 0$. Here we have $\left[\mathcal{P}_{00}(\alpha, \delta_1, \delta_2) = \int \int I[u > x_1 + \alpha] I[v > x] g(u, v) du dv\right]$ where as before the terms δ_1, δ_2 are in the definition of $g(u, v)$.

Conditioning on the indexes x_1, x this derivative now with respect to α at $\alpha = \alpha_0, \delta_1 = \delta_{10}$ can be expressed as

$$d_{11}\mathcal{P}_{11}(\alpha_0, \delta_{10}, \delta_{20})^{-1} \int \int -\delta(u - x_1 - \alpha_0) I[v > x] du dv \tag{A.8}$$

where $\delta(\cdot)$ again denotes the Dirac delta function. The score with respect to δ_1 at $\delta_1 = \delta_{10}, \delta_2 = \delta_{20}, \alpha = \alpha_0$ is

$$d_{11}\mathcal{P}_{11}(\alpha_0, \delta_{10})^{-1} \int \int I[u > x_1 + \alpha_0] I[v > x] g_0(u, v) h_1(u) du dv \tag{A.9}$$

So as we see again we can set $h_1^*(u) = -\delta_\epsilon(u - x_1 - \alpha_0)$. The same arguments can be used to deal with the outcome $y_1 = 0. y_2 = 1$.

Q.E.D.

A.2   Proof of Theorem 2.2

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

Assumption 7   *(i) The joint density of errors is continous almost everywhere in the L_2 norm*

(ii) *Functions $\mathcal{P}_{11}(x_1 + t, x) = \int_{-\infty}^{x_1+t} \int_{-\infty}^{x} g(u, v) \, du \, dv$ and $\mathcal{P}_{01}(x_1 + t, x) = \int_{x_1+t}^{-\infty} \int_{-\infty}^{x} g(u, v) \, du \, dv$ are differentiable in mean square in t with the mean-square derivative $\dot{\mathcal{P}}_{k1}$, $k = 0, 1$ such that $E\left[|\dot{\mathcal{P}}_{k1}|^2\right] < \infty$.*

(iii) *There exists a Hilbert space $\mathbf{H}$ with the basis $\{h_l\}_0^\infty$ such that*

    (a) *For any sufficiently large K and $\mathbf{H}_K = \{h_l\}_0^K$ and the orthogonal projection of the density $proj(g \,|\, \mathbf{H}_K) = O(K^{-\alpha})$, for $\alpha > 0$*

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

    (c) *For any finite K the closure of the linear space based on $\mathbf{H}_K$ with coefficients in the ball of radius ϵ has uniform entropy that is linear in ϵ and at most linear in K.*

We introduce the "uncensored" objective function

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

with

$$Q(\alpha) = E\left[q\left(\alpha; y_1, y_2, x_1, x\right)\right].$$

Denote

$$\hat{l}\left(\alpha\right) = \frac{1}{n}\sum_{i=1}^{n} l\left(\alpha; y_{1i}, y_{2i}, x_{1i}, x_i\right).$$

Also denote

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

and

$$\hat{\ell}\left(\alpha\right) = \frac{1}{n}\sum_{i=1}^{n}\ell\left(\alpha; y_{1i}, y_{2i}, x_{1i}, x_i\right).$$

Now consider the following decomposition of the objective function:

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

where

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

Term R_1

For convenience, we introduce new notation denoting

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

and introduce vectors $p^K(z) = \left(p^{K1}(z), \ldots, p^{KK^2}(z)\right)'$. Also let $d_i^{00} = (1 - y_{1i})(1 - y_{2i})$ and $d^{00} = \left(d_1^{00}, \ldots, d_n^{00}\right)'$. Let $\Delta(z) = E[d^{00}|z]$ and $\Delta = (\Delta(z_1), \ldots, \Delta(z_n))'$. We can project this

function of z on K basis vectors of the sieve space. Let β be the vector of coefficients of this projection. As demonstrated in Newey (1997), for $P = \left(p^K(z_1), \ldots, p^K(z_n)\right)'$ and $\hat{Q} = P'P/n$

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

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

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

Then we can write the estimate

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

Note that

$$m^{K\prime}(z)\left(\hat{\beta} - \beta\right) = m^{K\prime}(z)\left(\hat{Q}^{-1}P'\left(d^{00} - \Delta\right)/n + \hat{Q}^{-1}P'\left(\Delta - P\beta\right)/n\right). \tag{A.10}$$

For the first term in (A.10), we can use the result that smallest eigenvalue of $\hat{Q}$ is converging to $\underline{\lambda} > 0$. Then application of the Cauchy-Schwartz inequality leads to

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

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

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

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

Thus,

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

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

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

40

Using this observation, we can find that

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

This inequality allows us to substitute the tail bound for the class of functions $\mathcal{P}_n^{11}(\cdot,\cdot)$ by a tail bound for fixed functions

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

Then we can apply the Hoeffding exponential inequality to obtain

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

As a result, we find that

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

Then, provided that $n/\log K \to \infty$, we prove that the right-hand side of this inequality converges to zero. Application of the delta-method allows us to conclude that for any given $\delta > 0$ and $n/\log K \to \infty$

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

Term R_3

Consider the approximation bias term. Note that we can express

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

Similarly, we can express

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

Noting that one can attain a uniform rate

$$\sup_{x_1,x}\left\|\hat{\mathcal{P}}_n^{11}(x_1, x) - \mathcal{P}^{11}(x_1 + \alpha, x)\right\| = O_p\left(\sqrt{\frac{K}{n}} + K^{-(d+1)/2}\right),$$

given the quality of approximation by Hermite polynomials and d mean square derivatives of the density of interest. We can then evaluate the entire term

$$|R_3| = O\left(\sqrt{\frac{K}{n}} + K^{-(d+1)/2}\right).$$

Terms R_4 and R_5

Consider term R_4. We can evaluate this term as

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

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

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

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

from Assumption 1.

Term R_2

We use the following assumption regarding the population likelihood function.

Assumption 8 *The population likelihood function $Q(\cdot)$ is twice continuously differentiable and uniquely maximized at α_0 with a negative definite Hessian.*

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

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

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

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

$$F_{n,\delta} = \mathbf{1}\{|x_1 + \alpha_0| \leq c_n + \delta\}\omega_n(x)$$

$$\times \left[\frac{y_1 y_2 \dot{\mathcal{P}}^{11}}{\mathcal{P}^{11}} + \frac{(1 - y_1)y_2 \dot{\mathcal{P}}^{01}}{\mathcal{P}^{01}}\right]\delta.$$

Then, by Assumption 1, we can evaluate

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

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 η and split the interval $[0, \eta]$ 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 can use the results on the evaluation of the terms R_1, $R_3 - R_5$, taking into consideration that

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

for some $H > 0$ due to differentiability of $Q(\cdot)$ and the restriction on its Hessian at α_0 in Assumption 8. Then we can evaluate

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

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

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

Then we use the Markov inequality to obtain that

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

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

Provided the finiteness of the covering integral of the class $\mathcal{F}_{n,\delta}$, we can use the maximum inequality to evaluate

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

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

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

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

This means that

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

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

Q.E.D.

A.3 Proof of Theorem 2.3

First, consider evaluation from the proof of Theorem 2.2

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

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

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

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

We note that evaluation here is different, because, unlike Theorem 2.2 we allow $r_n\beta(c_n) = O(1)$. This allows us to obtain

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

Thus, if $L = 2^M$ then

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

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

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

This means that r_n is the upper rate.

Q.E.D.

A.4 Proof of Theorem 2.4

The log-likelihood function of the model is

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

with

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

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

We note that we use the same distribution of covariates x_1 and x. Then for $c_n \to \infty$ pick

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

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

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

We can then express Therefore

$$\Lambda(\mathcal{P}_1, \mathcal{P}_2) = \exp\left(n\hat{\mathcal{L}}_1(\alpha_1) - n\hat{\mathcal{L}}_2(\alpha_2)\right)$$

$$= \exp\left(n\hat{\ell}(\alpha_1) - n\hat{\ell}(\alpha_2) - n\hat{e}(\alpha_2)\right)$$

$$= \exp\left(n\left[\hat{\ell}(\alpha_1) - \hat{\ell}(\alpha_2) - \ell(\alpha_1) + \ell(\alpha_2)\right] - n\hat{e}(\alpha_2) - n\left(\ell(\alpha_2) - \ell(\alpha_1)\right)\right)$$

Then $\log \Lambda(\mathcal{P}_1, \mathcal{P}_2)$ is bounded from below as n approaches infinity if and only if $n\left(\ell(\alpha_2) - \ell(\alpha_1)\right)$ is bounded. We note that α_1 maximizes $\ell(\alpha_1)$. This means that

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

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

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

This means that $\frac{n\beta(c_n)^2}{\nu(c_n)} = O(1)$ suggests that for large n there exists a lower bound on the likelihood ratio. Invoking Lemma 2.1 we obtain the desired result.

Q.E.D.

A.5 Proof of Theorem 3.1

Denote $\mathbf{P}_{ij}(x_1, x) = P(y_1 = i, y_2 = j \mid x_1, x)$ and assume that $\mathbf{P}_{11}(\cdot, \cdot)$ is the probability satisfying the conditions of the Theorem. Then write this probability

$$\mathbf{P}_{11}(x_1, x) = \int \mathbf{1}\{x_1 - u + \alpha_0\Phi\left(\frac{x - v}{\sigma}\right) > 0\}\Phi\left(\frac{x - v}{\sigma}\right) g(u, v) \, du \, dv.$$

We note that the right-hand side of this expression is a convolution of the error density with the product of indicator and cdf of the noise in treatment. First, find the Fourier transform of conditional choice probability

$$\int e^{-i(t_1 x_1 + t_2 x)} \mathbf{1}\{x_1 + \alpha_0\Phi\left(\frac{x}{\sigma}\right) > 0\}\Phi\left(\frac{x}{\sigma}\right) dx_1 \, dx$$

$$= \frac{\sigma}{2}\left(\delta(t_1) + \frac{1}{it_1}\right) \int e^{i\alpha_0 t_1 \Phi(x) - i\sigma t_2 x}\Phi(x) \, dx.$$

45

Denote $\mathcal{F}_{ij}(t)$ the Fourier transform of the corresponding observable conditional probability. Then, the property of the Fourier transform of the convolution and denoting th cahracteristic funciton of the joint error distribution, we find that

$$\mathcal{F}_{11}(t) - \mathcal{F}_{01}(t) = \frac{\sigma \chi(t)}{it_1} \int e^{i\alpha_0 t_1 \Phi(x) - i\sigma t_2 x} \Phi(x)\, dx$$

Note that given that $\Phi(\cdot)$ is known

$$\zeta_{\alpha_0}(t_1, t_2) = \int e^{i\alpha_0 t_1 \Phi(x) - it_2 x} \Phi(x)\, dx$$

is also known. Therefore, we can express the joint characteristic function as

$$\chi(t) = it_1 \frac{\mathcal{F}_{11}(t) - \mathcal{F}_{01}(t)}{\sigma \zeta(t_1, \sigma t_2)}.$$

This means that if $\chi_v(\cdot)$ is the characteristic function of the marginal distribution of v, it can be expressed as

$$\chi_v(t_2) = \lim_{t_1 \to 0} it_1 \frac{\mathcal{F}_{11}(t) - \mathcal{F}_{01}(t)}{\sigma \zeta(t_1, \sigma t_2)}.$$

Applying the L'Hôpital's rule we find the limit as

$$\chi_v(t_2) = \frac{\mathcal{F}_{11}(0, t_2) - \mathcal{F}_{01}(0, t_2)}{\sigma \alpha_0 \int e^{-i\sigma t_2 x} \Phi(x)^2\, dx}.$$

Now consider the marginal probability

$$\mathbf{P}_1(x) = P(y_2 = 1 \mid x) = \int \Phi\left(\frac{x - v}{\sigma}\right) g(u, v)\, du\, dv,$$

which is a convolution of the marginal density $g_v(\cdot)$ with the known function $\Phi(\cdot)$. Therefore the Fourier transform of the marginal probability can be expressed as

$$\mathcal{F}_1(t_2) = \sigma \int e^{-i\sigma t_2 x} \Phi(x)\, dx \chi_v(t_2).$$

This allows us to express the parameter of interest as

$$\alpha_0 = \frac{\mathcal{F}_{11}(0, t_2) - \mathcal{F}_{01}(0, t_2)}{\mathcal{F}_1(t_2)} \frac{\int e^{-i\sigma t_2 x} \Phi(x)\, dx}{\int e^{-i\sigma t_2 x} \Phi(x)^2\, dx}$$

Q.E.D.

A.6 Proof of Theorem 3.2

To determine the information in this model we proceed as we did before, defining the space of bivariate density functions and paths the same way.

As before, we will focus on the case where $y_2 = y_1 = 1$.

The conditional probability of this outcome is of the form:

$$\int \Phi((x-v)/\sigma) F_{u|v}(x_1 + \alpha_0 \Phi(x-v)) f_v(v) dv \equiv \mathcal{P}_{11} \tag{A.11}$$

so the derivative with respect to α at $\alpha = \alpha_0$ is of the form:

$$\int \Phi((x-v)/\sigma) f_{u|v}(x_1 + \alpha_0 \Phi((x-v)/\sigma)) \Phi((x-v)/\sigma) f_v(v) dv \equiv s_{\alpha,11} \tag{A.12}$$

Similarly, the derivative with respect to δ_2 at $\delta_2 = \delta_{20}$ is

$$\int \Phi((x-v)\sigma) F_{u|v}(x_1 + \alpha_0 \Phi((x-v)\sigma)) h_2(v) f_v(v) dv \equiv s_{\delta_2,11} \tag{A.13}$$

Thus now we can set $h_2^*(v)$ to be

$$h_2^*(x, x_1, v) = \frac{f_{u|v}(x_1 + \alpha_0 \Phi((x-v))/\sigma) \Phi((x-v)\sigma)}{F_{u|v}(x_1 + \alpha_0 \Phi((x-v)\sigma))} \tag{A.14}$$

However, to show positive information, we can turn to the outcome where $y_2 = 1, y_1 = 0$. Carrying through with the same arguments as above, we get

$$h_2^*(x, x_1, v) = \frac{S'_{u|v}(x_1 + \alpha_0 \Phi((x-v)\sigma)) \Phi((x-v))/\sigma}{S_{u|v}(x_1 + \alpha_0 \Phi((x-v)\sigma))} \tag{A.15}$$

where $S_{u|v}(\cdot)$ denotes the conditional survival function of u given v. This is distinct from the previous value of $h_1^*(x, x_1, v)$ in the $y_1 = y_2 = 1$ case; therefore we cannot choose a function $h_2(v)$ to set the information for α_0 equal to 0, as we did in the complete information model.

Given that we have proven that there is positive information for this incomplete information model, the next step is to derive what the level of the information is. For this we need to turn attention to the choice probabilities of all the outcomes. When considering the problem at hand, this involves choosing functions h_1, h_2 to minimize:

$$
\begin{aligned}
E_{x,x_1} \Big[& \mathcal{P}_{11}^{-1} \left((s_{\alpha,11} - s_{\delta_1,11} h_1 - s_{\delta_2,11} h_2)^2 \right) \\
& + \mathcal{P}_{00}^{-1} \left((s_{\alpha,00} - s_{\delta_1,00} h_1 - s_{\delta_2,00} h_2)^2 \right) \\
& + \mathcal{P}_{01}^{-1} \left((s_{\alpha,01} - s_{\delta_1,01} h_1 - s_{\delta_2,01} h_2)^2 \right) \\
& + \mathcal{P}_{10}^{-1} \left((s_{\alpha,10} - s_{\delta_1,10} h_1 - s_{\delta_2,10} h_2)^2 \right) \Big]
\end{aligned}
\tag{A.16}
$$

where here $s_{\alpha,11}$ denotes the derivative of probability $y_1 = y_2 = 1$, with respect to α, evaluated at $\alpha = \alpha_0, \delta_1 = \delta_{10}, \delta_2 = \delta_{20}$. $s_{\delta_1,11}, s_{\delta_2,11}$ denote the derivatives with with respect to δ_1, δ_2, respectively, evaluated at $\alpha = \alpha_0, \delta_1 = \delta_{10}, \delta_2 = \delta_{20}$. Note we evaluated these functions in the $y_1 = y_2 = 1$ previously in our proof of positive information. Note also the above objective function depends on σ. Interestingly, we can take the limit of the above objective function as $\sigma \to 0$; we can see that h_1, h_2 can now be chosen to set the objective function to be arbitrarily small, bringing us back to the results obtained with the complete information setting. Note the objective function in (A.16) involves integrals with respect to u, v, and can be viewed as a weighted least squares objective function, with weights corresponding to the inverse of the outcome probabilities. and dependent variable values corresponding to $s_{\alpha 11}, s_{\alpha 01}, s_{\alpha 00}, s_{\alpha 10}$.

Therefore, the optimal functions $\mathbf{h}^* = (h_1^*, h_2^*)'$ are

$$\mathbf{h}^* = \left(\sum_{i=1}^4 w_i^2 \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{i=1}^4 w_i^2 \mathbf{x}_i \jmath_i \tag{A.17}$$

where the summations are across the 4 "observations" corresponding to the outcomes $y_1 = y_2 = 1, y_1 = y_2 = 0, y_1 = 0, y_2 = 1$; w_i denotes the weights, $\mathbf{x}_i$ denotes the two dimensional vector of "regressors", corresponding to scores with respect to the two perturbation parameters, and $\jmath_i$ denotes the "dependent variables" in the regression, in this case the scores with respect to the parameter of interest.

Specifically, we find that $w_1 = \mathcal{P}_{11}^{-1}$, $w_2 = \mathcal{P}_{00}^{-1}$, $w_3 = \mathcal{P}_{01}^{-1}$, $w_4 = \mathcal{P}_{10}^{-1}$, $\mathbf{x}_1 = s_{\delta_1,11}, s_{\delta_2,11}$, $\mathbf{x}_2 = s_{\delta_1,00}, s_{\delta_2,00}$, $\mathbf{x}_3 = s_{\delta_1,01}, s_{\delta_2,01}$, $\mathbf{x}_4 = s_{\delta_1,10}, s_{\delta_2,10}$, $\jmath_2 = 0$, $\jmath_4 = 0$, and

$$\jmath_1 = \int \Phi(x - v) f_{u|v}(x_1 + \alpha_0 \Phi(x - v)) \Phi(x - v) f_v(v) dv$$

$$\jmath_3 = \int \Phi(x - v) f_{u|v}(x_1 + \alpha_0 S(x - v)) \Phi(x - v) f_v(v) dv,$$

where $S(\cdot)$ above denotes $(1 - \Phi(\cdot))$.

Decompose $\mathbf{x}_i = (x_{1i}, x_{2i})$, then

$$h_1^* = \frac{\sum w_i^2 \jmath_i x_{1i} \sum w_i^2 x_{2i}^2 - \sum w_i^2 \jmath_i x_{2i} \sum w_i^2 x_{1i} x_{2i}}{\sum w_i^2 x_{1i}^2 \sum w_i^2 x_{2i}^2 - \left(\sum w_i^2 x_{1i} x_{2i} \right)^2}$$

$$h_2^* = \frac{\sum w_i^2 \jmath_i x_{2i} \sum w_i^2 x_{1i}^2 - \sum w_i^2 \jmath_i x_{1i} \sum w_i^2 x_{1i} x_{2i}}{\sum w_i^2 x_{1i}^2 \sum w_i^2 x_{2i}^2 - \left(\sum w_i^2 x_{1i} x_{2i} \right)^2}$$

Finally, the values of $\mathbf{h}^*$ can be used in (A.16) to obtain the information for α_0 in this model. *Q.E.D.*

A.7 Proof of Theorem 3.3

Consider the vector $\zeta(\cdot)$ from the proof of Theorem 3.4. This vector depends on $a(t) = \int e^{i\sigma tx}\Phi(x)\,dx / \int e^{i\sigma tx}\Phi(x)^2$
We note that given that both integrals do not exist at $\sigma = 0$ provided that $\Phi(\cdot)$ has support on the real line. We note that given that $\Phi(x) = \int_\infty^x \phi(t)\,dt$, then if the characteristic function of $\phi(\cdot)$ is $\chi_\phi(\cdot)$, then

$$\int e^{i\sigma tx}\Phi(x)\,dx = -\frac{1}{2}\chi_\phi(-\sigma t)(-\delta(-\sigma t) + \frac{i}{\pi\sigma t})$$

Then provided that

$$\int e^{-itx}\Phi(x)^2\,dx = \int_{-\infty}^{+\infty} \frac{1}{4}\chi_\phi(t-u)\chi_\phi(u)(\delta(u) + \frac{i}{\pi u})(\delta(t-u) + \frac{i}{\pi(t-u)})\,du$$

$$= \frac{2i}{\pi t}\chi_\phi(t) - \frac{1}{\pi^2}\int \frac{\chi_\phi(u)\chi_\phi(t-u)}{u(t-u)}\,du.$$

Then

$$\int e^{\sigma itx}\Phi(x)^2\,dx = -\frac{2i}{\pi\sigma t}\chi_\phi(t) + \frac{1}{\sigma^2\pi^2}\int \frac{\chi_\phi(u)\chi_\phi(-t-u)}{u(t+u)}\,du.$$

As a result, we can see that $\lim_{\sigma\to 0} a(t) = 0$. This means that $\zeta(\cdot) \to 0$ pointwise. By dominated convergence theorem, we conclude that $\Sigma^{-1} \to 0$, which means that the information of the model converges to zero.

Q.E.D.

A.8 Proof of Theorem 3.4

Our model contains the unknown interaction parameter and the unknown distributions. Consider a parametric sub-model parametrized by scalar parameter θ. Note that the likelihood of the model can be written as

$$f_\theta(y_1, y_2, x_1, x) = \prod_{ij=0}^{1} \mathbf{P}_{ij,\theta}(x_1, x)^{\mathbf{1}\{y_1=i,y_2=j\}} f_\theta(x_1, x)$$

Then we can express the score of the model as

$$S_\theta(y_1, y_2, x_1, x_2) = \frac{\mathbf{P}_{00}y_1 y_2 - \mathbf{P}_{11}(1-y_1)(1-y_2)}{\mathbf{P}_{00}\mathbf{P}_{11}}\dot{\mathbf{P}}_{11}$$
$$+ \frac{(\mathbf{P}_{00}y_1 - \mathbf{P}_{10}(1-y_1))(1-y_2)}{\mathbf{P}_{00}\mathbf{P}_{10}}\dot{\mathbf{P}}_{10}$$
$$\frac{(\mathbf{P}_{00}y_2 - \mathbf{P}_{01}(1-y_2))(1-y_1)}{\mathbf{P}_{00}\mathbf{P}_{01}}\dot{\mathbf{P}}_{01} + s_\theta(x_1, x)$$

where s_θ is the score of the distribution of covariates. Then we can express the tangent set of the model as

$$\mathcal{T} = \Big\{ \xi_1(x_1, x)(\mathbf{P}_{00}y_1 y_2 - \mathbf{P}_{11}(1-y_1)(1-y_2))$$
$$+\xi_2(x_1, x)(\mathbf{P}_{00}y_1 - \mathbf{P}_{10}(1-y_1))(1-y_2)$$
$$+\xi_3(x_1, x)(\mathbf{P}_{00}y_2 - \mathbf{P}_{01}(1-y_2))(1-y_1) + t(x_1, x)\Big\},$$

49

where $\xi_i(\cdot)$ for $i = 1, 2, 3$ are square integrable functions, and $E[t(x_1, x)] = 0$. Denote $\alpha(\theta) = (\alpha_1(\theta), \alpha_2(\theta))'$ and find the derivative of this vector along the parametrization path. To do that we use the moment system and the system of equilibrium probabilities. Consider equation (3.3) and denote

$$a(t_2) = \frac{\int e^{-i\sigma t_2 x} \Phi(x) \, dx}{\int e^{-i\sigma t_2 x} \Phi(x)^2 \, dx}.$$

Given that $\mathbf{P}_1(x) = \mathbf{P}_{11}(x_1, x) + \mathbf{P}_{01}(x_1, x)$, we can write the expression for the interaction parameter

$$\alpha_0 = \mathbf{P}_1(x)^{-1} \int e^{it_2 x} a(t_2) \left(\mathcal{F}_{11}(0, t_2) - \mathcal{F}_{01}(0, t_2) \right) dt_2.$$

Suppose that at least one of the Fourier transforms has a non-vanishing support and for simplicity assume that the support is on the entire real line. Let $\mu(\cdot)$ be a non-negative weight function that integrates to 1. Then we can express parameter α_0 as

$$E[\mu(X)]\alpha_0 = E\left[\frac{\mu(X)}{\mathbf{P}_1(X)} \int e^{it_2 X} a(t_2) \left(\mathcal{F}_{11}(0, t_2) - \mathcal{F}_{01}(0, t_2) \right) dt_2 \right].$$

Then we compute the directional derivative of both sides of this equation with respect to the parametrization path θ. Combining the terms, we can represent the derivative as

$$\frac{\partial \alpha(\theta)}{\partial \theta} = E[\mu(X)]^{-1} \int \left(\dot{\mathbf{P}}_{11} - \dot{\mathbf{P}}_{10} \right) E\left[\frac{\mu(X)}{\mathbf{P}_1(X)} \int e^{it_2(X-x)} a(t_2) \, dt_2 \right] dx_1 \, dx$$

$$- E\left[\frac{\mu(X) \left(\dot{\mathbf{P}}_{11} + \dot{\mathbf{P}}_{10} \right)}{\mathbf{P}_1^2(X)} \int e^{it_2 x} a(t_2) \left(\mathcal{F}_{11}(0, t_2) - \mathcal{F}_{01}(0, t_2) \right) dt_2 \right]$$

Next we find function $\Psi(y_1, y_2, x_1, x)$ from the tangent set such that

$$\frac{\partial \alpha(\theta)}{\partial \theta} = E\left[\Psi \, S_\theta \right].$$

We can express such function as

$$\Psi(y_1, y_2, x_1, x) = a(x_1, x) \left(y_1 y_2 - \mathbf{P}_{11}(x_1, x) \right) + b(x_1, x) \left((1 - y_1) y_2 - \mathbf{P}_{01}(x_1, x) \right),$$

where

$$a(x_1, x) = f(x_1, x)^{-1} E\left[\frac{\mu(X)}{\mathbf{P}_1(X)} \int e^{it_2(X-x)} a(t_2) \, dt_2 \right]$$

$$- \frac{\mu(x)}{\mathbf{P}_1(x)^2} \int e^{it_2 x} a(t_2) \left(\mathcal{F}_{11}(0, t_2) - \mathcal{F}_{01}(0, t_2) \right) dt_2,$$

and

$$b(x_1, x) = - f(x_1, x)^{-1} E\left[\frac{\mu(X)}{\mathbf{P}_1(X)} \int e^{it_2(X-x)} a(t_2) \, dt_2 \right]$$

$$- \frac{\mu(x)}{\mathbf{P}_1(x)^2} \int e^{it_2 x} a(t_2) \left(\mathcal{F}_{11}(0, t_2) - \mathcal{F}_{01}(0, t_2) \right) dt_2,$$

Then the variance of the influence function can be expressed as

$$\text{Var}(\Psi) = E\left[\mathbf{P}_{11}(1-\mathbf{P}_{11})a^2(X_1, X) + \mathbf{P}_{10}(1-\mathbf{P}_{10})b^2(X_1, X)\right]$$

Then the efficiency bound can be expressed as a minimum

$$\Sigma = \min_{\mu(\cdot)} E[\mu(X)]^{-1} E\left[\mathbf{P}_{11}(1-\mathbf{P}_{11})\left(f(x_1,x)^{-1}E\left[\frac{\mu(X)}{\mathbf{P}_1(X)}\int e^{it_2(X-x)}a(t_2)\,dt_2\right]\right.\right.$$
$$\left.- \frac{\mu(x)}{\mathbf{P}_1(x)^2}\int e^{it_2x}a(t_2)\left(\mathcal{F}_{11}(0,t_2) - \mathcal{F}_{01}(0,t_2)\right)\,dt_2\right)^2$$
$$+ \mathbf{P}_{10}(1-\mathbf{P}_{10})\left(f(x_1,x)^{-1}E\left[\frac{\mu(X)}{\mathbf{P}_1(X)}\int e^{it_2(X-x)}a(t_2)\,dt_2\right]\right.$$
$$\left.\left.+ \frac{\mu(x)}{\mathbf{P}_1(x)^2}\int e^{it_2x}a(t_2)\left(\mathcal{F}_{11}(0,t_2) - \mathcal{F}_{01}(0,t_2)\right)\,dt_2\right)^2\right]E[\mu(X)]^{-1\prime}$$

which exists with a minimum attained in the basis of Hermite polynomials, which are eigenfunctions of the Fourier transform. We can represent $\mu(x) = \mathbf{P}_1(x)^2\rho(x)$, denote

$$\Omega(x_1, x) = \text{diag}\left(\mathbf{P}_{11}(1-\mathbf{P}_{11}),\ \mathbf{P}_{10}(1-\mathbf{P}_{10})\right),$$

and

$$T = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

Finally, let

$$\zeta(x_1, x) = \left(f(x_1,x)^{-1}E\left[\mathbf{P}_1(X)\rho(X)\int e^{it_2(X-x)}a(t_2)\,dt_2\right],\right.$$
$$\left.\rho(x)\int e^{it_2x}a(t_2)\left(\mathcal{F}_{11}(0,t_2) - \mathcal{F}_{01}(0,t_2)\right)\,dt_2\right).$$

Then we follow the argument in (Ai and Chen 2003), and find $\rho^*(\cdot)$ solving

$$\min_{\rho\in\mathcal{H}} E\left(\zeta(x_1,x)'T^{-1}\Omega(x_1,x)^{-1}T^{-1}\zeta(x_1,x)\right)$$

Then the semiparametric efficiency bound can be expressed as

$$\Sigma = E\left(\zeta(x_1,x)^{*\prime}T^{-1}\Omega(x_1,x)^{-1}T^{-1}\zeta(x_1,x)^*\right)^{-1}$$

Q.E.D.

A.9 Proof of Theorem 4.2

Thus we can see, that under our conditions the parameter α_0 cannot be estimated at the parametric rate. This result is analogous to impossibility theorems in Chamberlain (1986).

To establish the information of the parameters α_1, α_2, we will proceed as before; in this case the likelihood function conditional on the indexes is of the form

$$P(y_1 = y_2 = 1 | x_1, x_2) = P(y_1 = y_2 = 1 | x_1, x_2, (\epsilon, \eta) \in S_1)$$
$$(P((\epsilon, \eta) \in S_1 | x_1, x_2) + P(y_1 = y_2 = 1 | x_1, x_2, (\epsilon, \eta) \notin S_1)$$
$$(P((\epsilon, \eta) \notin S_1 | x_1, x_2)$$
$$= \frac{1}{2}(P((\epsilon, \eta) \in S_1 | x_1, x_2) + P(y_1 = y_2 = 1 | x_1, x_2, (\epsilon, \eta) \notin S_1)$$
$$(P((\epsilon, \eta) \notin S_1 | x_1, x_2)$$

Thus, we can express this conditional probability as

$$\frac{1}{2} \int I[\alpha_1 + x_1 < u < x_1] I[x_2 < v < \alpha_2 + x_2] g_0(u, v) du dv \tag{A.18}$$

$$+ \int (I[x_1 < u < x_1 + \alpha_1] I[v < x_2] + I[u < x_1] I[v < x_2 + \alpha_2]) g_0(u, v) du dv \tag{A.19}$$

As before, we will only focus on the probability of this outcome, for similar reasons. We will show here that there is 0 information for α_1, noting that by the symmetry of the problem, identical arguments can be used to show zero information for α_2.

The method to show this will be similar- we take the derivative of the above probability with respect to α_1 (or, more precisely, take the derivative with respect to α evaluated at $\alpha = \alpha_1$. This derivative is of the form

$$\frac{1}{2} \int -\delta[\alpha_1 + x_1 - u] I[x_2 < v < \alpha_2 + x_2] g_0(u, v) du dv + \int \delta[-u + x_1 + \alpha_1] I[v < x_2] g_0(u, v) du dv \tag{A.20}$$

where again δ denotes the dirac function.

Also we can take the derivative with respect to to the parameter perturbing the joint density of the disturbances u, v.

This is of the form:

$$\frac{1}{2} \int I[\alpha_1 + x_1 < u < x_1] I[x_2 < v < \alpha_2 + x_2] g_0(u, v) h_1(u) du dv \tag{A.21}$$

$$+ \int (I[x_1 < u < x_1 + \alpha_1] I[v < x_2] + I[u < x_1] I[v < x_2 + \alpha_2]) g_0(u, v) h_1(u) du dv \tag{A.22}$$

As before we need to find a function $h_1(u)$ that makes these two integrals as close as possible. For the problem at hand, we can set

$$h_1^*(u) = -\delta[\alpha_0 - x_1 - u] \qquad (A.23)$$

Q.E.D.

A.10 Proof of Theorem 5.1

To prove the theorem we stablish the fact that the rank and order conditions of the system of conditional moments are satisfied for some subset of the covariate values. We compute the Jacobi matrix of system (??) with respect to α and show that it is non-zero at least for some subset of covariate values. Define $P(x_1 - u, x_2 - v) = (P(y_1 = i, y_2 = j | x_1 - u, x_2 - v), \ i, j \in 0, 1, \ i \neq j)'$, and $\mathbf{P}(x_1, x_2) = (\mathbf{P}_{ij}(x_1, x_2), \ i, j \in 0, 1, \ i \neq j)'$. Then system (5.4) can be written in a simplified form

$$\mathbf{P}(x_1, x_2) - \int P(x_1 - u, x_2 - v)\bar{g}(u, v) \, du \, dv = 0.$$

We now compute the Jacobi matrix of this system which will correspond to the derivative of the second term with respect to vector $\alpha = (\alpha_1, \alpha_2)$. For convenience we also pick a specific path θ and take the derivative of parameters along this path. Differentiating the system defining the Bayes-Nash equilibrium we can express

$$\sigma \frac{\phi(P_1)}{\Phi(P_1)} \dot{P}_{1\theta} = \frac{\partial \alpha_1(\theta)}{\partial \theta} P_2 + \alpha_1 \dot{P}_{2\theta},$$
$$\sigma \frac{\phi(P_2)}{\Phi(P_2)} \dot{P}_{2\theta} = \frac{\partial \alpha_2(\theta)}{\partial \theta} P_1 + \alpha_2 \dot{P}_{1\theta}.$$

Therefore, we can express

$$\begin{pmatrix} \dot{P}_{1\theta} \\ \dot{P}_{2\theta} \end{pmatrix} = \begin{pmatrix} -\alpha_2 & \sigma \frac{\phi(P_2)}{\Phi(P_2)} \\ \sigma \frac{\phi(P_1)}{\Phi(P_1)} & -\alpha_1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & P_1 \\ P_2 & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial \alpha_1(\theta)}{\partial \theta} \\ \frac{\partial \alpha_1(\theta)}{\partial \theta} \end{pmatrix}$$

Denote

$$\Omega(x_1, x_2) = \left(\alpha_1 \alpha_2 - \sigma^2 \frac{\phi(P_1)\phi(P_2)}{\Phi(P_1)\Phi(P_2)} \right)^{-1} \begin{pmatrix} \sigma \frac{\phi(P_2)}{\Phi(P_2)} P_2 & -\alpha_1 P_1 \\ -\alpha_2 P_2 & \sigma \frac{\phi(P_1)}{\Phi(P_1)} P_1 \end{pmatrix}$$

Next, we compute the directional derivative of the vector

$$\dot{P}(x_1, x_2) = \begin{pmatrix} 1 - P_2 & -P_1 \\ -P_2 & 1 - P_1 \end{pmatrix} \begin{pmatrix} \dot{P}_{1\theta} \\ \dot{P}_{2\theta} \end{pmatrix} = M(x_1, x_2) \begin{pmatrix} \dot{P}_{1\theta} \\ \dot{P}_{2\theta} \end{pmatrix}.$$

Therefore, we can express

$$\dot{P}(x_1, x_2) = M(x_1, x_2)\Omega(x_1, x_2)\frac{\partial \alpha(\theta)}{\partial \theta}.$$

Then we can differentiate $\bar{g}(\cdot,\cdot)$ along the parametrization path. Consider the vector

$$\varphi_{11}(t) = \int\int e^{-i(t_1 q_1 + t_2 q_2)} \left(P_2(q_1,q_2),\, P_1(q_1,q_2)\right) \Omega(q_1,q_2)\, dq_1\, dq_2$$

$$\frac{\partial \bar{g}(u,v)}{\partial \theta} = \int\int E\left[e^{i(t_1(u-x_1)+t_2(v-x_2))} \frac{\dot{\mathbf{P}}_{11}(X_1,X_2)}{f(X_1,X_2)\mathcal{F}_{11}(t_1,t_2)}\right] dt_1\, dt_2$$
$$+ \left(\int\int E\left[e^{i(t_1(u-X_1)+t_2(v-X_2))}\frac{\mathbf{P}_{11}(X_1,X_2)\varphi_{11}(t_1,t_2)}{f(X_1,X_2)\mathcal{F}_{11}^2(t_1,t_2)}\right] dt_1\, dt_2\right)\frac{\partial \alpha}{\partial \theta}$$

Then the Jacobi matrix can be expressed as the matrix pre-multiplying the derivative $\frac{\partial \alpha(\theta)}{\partial \theta}$ in

$$\int \frac{\partial}{\partial \theta}P(x_1 - u, x_2 - v)\bar{g}(u,v)\, du\, dv + \int P(x_1 - u, x_2 - v)\frac{\partial}{\partial \theta}\bar{g}(u,v)\, du\, dv$$

We start the analysis with the second element which can be written as

$$\int P(x_1 - u, x_2 - v)\left(\int\int E\left[e^{i(t_1(u-X_1)+t_2(v-X_2))}\frac{\mathbf{P}_{11}(X_1,X_2)\varphi_{11}(t)}{f(X_1,X_2)\mathcal{F}_{11}^2(t)}\right] dt_1\, dt_2\right) du\, dv.$$

Provided that Fubbini theorem applies, we can transform this expression using the change of variables into

$$E\left[\frac{\mathbf{P}_{11}(X_1,X_2)}{f(X_1,X_2)}\int\left(\int e^{-i(q_1 t_1 + q_2 t_2)}P(q_1,q_2)\, dq_1\, dq_2\right)\frac{e^{i(t_1(x_1-X_1)+t_2(x_2-X_2))}\varphi_{11}(t)}{\mathcal{F}_{11}^2(t)}\right]$$

We then note that

$$\int e^{-i(q_1 t_1 + q_2 t_2)}P(q_1,q_2)\, dq_1\, dq_2 = \begin{pmatrix} \mathcal{F}_1(t) - \mathcal{F}_{11}(t) \\ \mathcal{F}_2(t) - \mathcal{F}_{11}(t) \end{pmatrix}.$$

Therefore, we can define the 2×2 matrix of inverse Fourier transforms

$$F(x_1 - X_1, x_2 - X_2) = \int e^{i(t_1(x_1-X_1)+t_2(x_2-X_2))}\begin{pmatrix} \frac{(\mathcal{F}_1(t)-\mathcal{F}_{11}(t))\varphi_{11}(t)}{\mathcal{F}_{11}^2(t)} \\ \frac{(\mathcal{F}_2(t)-\mathcal{F}_{11}(t))\varphi_{11}(t)}{\mathcal{F}_{11}^2(t)} \end{pmatrix} dt_1\, dt_2,$$

which is not equal to zero at some compact subset of x_1 and x_2. Then we can express the object under consideration as

$$E\left[\frac{\mathbf{P}_{11}(X_1,X_2)}{f(X_1,X_2)}F\left(x_1 - X_1, x_2 - X_2\right)\right].$$

Going back to the expression for the Jacobi matrix, we can express its first term as

$$\int M(x_1 - u, x_2 - v)\Omega(x_1 - u, x_2 - v)\bar{g}(u,v)\, du\, dv.$$

Denoting

$$Q(t) = \int e^{-i(t_1 q_1 + t_2 q_2)}M(q_1,q_2)\Omega(q_1,q_2)\, dq_1\, dq_2,$$

we can define the inverse Fourier transformation

$$R(x_1 - X_1, x_2 - X_2) = \int e^{i(t_1(x_1-X_1)+t_2(x_2-X_2))} \mathcal{F}_{11}(t)^{-1} Q(t) \, dt_1 \, dt_2.$$

Then the Jacobi matrix can be written as

$$J(x_1, x_2) = E\left[\frac{\mathbf{P}_{11}(X_1, X_2)}{f(X_1, X_2)} \left[F(x_1 - X_1, x_2 - X_2) + R(x_1 - X_1, x_2 - X_2) \right] \right].$$

We note that

$$F(x_1 - X_1, x_2 - X_2) + R(x_1 - X_1, x_2 - X_2)$$
$$= \int e^{i(t_1(x_1-X_1)+t_2(x_2-X_2))} \mathcal{F}_{11}(t)^{-1}$$

$$\times \int e^{-i(q_1 t_1 + q_2 t_2)} \left(\begin{pmatrix} \frac{(\mathcal{F}_1(t) - \mathcal{F}_{11}(t))}{\mathcal{F}_{11}(t)} \\ \frac{(\mathcal{F}_2(t) - \mathcal{F}_{11}(t))}{\mathcal{F}_{11}(t)} \end{pmatrix} (P_2(q_1), \; P_1(q_1)) + M(q_1, q_2) \right) \Omega(q_1, q_2) \, dq_1 \, dq_2$$

Given that $|\det(\Omega(q_1, q_2))| = 1$, we conclude that the Jacobi matrix is non-singular where the Fourier transform of the probability $\mathbf{P}_{11}(x_1, x_2)$ is non-vanishing.
Q.E.D.

A.11   Proof of Theorem 5.2

Nonetheless, we express the likelihood function as follows; w.l.o.g., we start with the conditional probability that $y_1 = y_2 = 1$, where at first we condition on u, v, x_1, x_2; note that given our assumption of η_1, η_2 being mutually independent, this joint conditional probability is the product of marginal conditional probabilities; so we have

$$P(y_1 = y_2 = 1 | x_1, x_2, u, v) = P(y_1 = 1 | x_1, x_2, u, v) P(y_2 = 1 | x_1, x_2, u, v) \tag{A.24}$$
$$= \Phi\left(\frac{x_1 + \alpha_1 P_2 - u}{\sigma} \right) \Phi\left(\frac{x_2 + \alpha_2 P_1 - v}{\sigma} \right) \tag{A.25}$$

Thus we can attain the conditional probability of this outcome, conditioning now only on x_1, x_2 by integrating this term with respect to the joint density of u, v. To ease our notation, let $P_1(u, v, \alpha_1, \alpha_2)$ denote the above conditional probability of $y = 1$, where have suppressed the dependence on x_1, x_2, and adopt the analogous notation for $y_2 = 1$; then the probability of the $y_1 = y_2 = 1$ outcome is

$$\int P_1(u, v, \alpha_1, \alpha_2) P_2(u, v, \alpha_1, \alpha_2) g(u, v) du dv \tag{A.26}$$

Now we can proceed as before; we take the partial derivative of the above probability with respect to α_1;

this is of the form:

$$\int \frac{\partial P_1(u,v,\alpha_1,\alpha_2)}{\partial \alpha_1} P_2(u,v,\alpha_1,\alpha_2) + P_1(u,v,\alpha_1,\alpha_2) \frac{\partial P_2(u,v,\alpha_1,\alpha_2)}{\partial \alpha_1} g(u,v) du dv \qquad \text{(A.27)}$$

Which to ease notation we will express as

$$\int (P_1 P_2)'(u,v) g(u,v) du du v \qquad \text{(A.28)}$$

We can express the above integral with with respect to the marginal density of u:

$$\int \mu(u) g(u) du \qquad \text{(A.29)}$$

where $\mu(u)$ denotes the conditional expectation $E[(P_1 P_2)'(u,v)|u]$ and $g(u)$ denotes the marginal density of u..

Now we turn attention to the score with respect to density perturbation parameter:

$$\int P_1(u,v) P_2(u,v) g(u,v) h_1(u) du dv \qquad \text{(A.30)}$$

which we can express as

$$\int \mu_2(u) g(u) h_1(u) du \qquad \text{(A.31)}$$

Thus we can set

$h_1^*(u) = \mu(u)/\mu_2(u)$ to make the integrals equal to each other; Thus to show information if positive. we can apply the same $h_1^*(u)$ in the outcome where $y_1 = y_2 = 0$.

In this case the score with respect to α_1 can be expressed as

$$\int \left(-E[P_1'|u] - E[P_2'|u] + \mu(u) \right) g(u) du \qquad \text{(A.32)}$$

where P_1' denotes $\frac{\partial P_1(u,v)}{\partial \alpha_1}$, P_2' denotes $\frac{\partial P_2(u,v)}{\partial \alpha_1}$.

The score with respect to δ_1 is of the form

$$1 - \int \left(E[P_1|u] - E[P_2|u] \right) h_1(u) g(u) du + \int \mu_2(u) h_1(u) g(u) du \qquad \text{(A.33)}$$

where P_1 above denotes $P_1(u,v)$;

For our choice of $h_1^*(u)$, the last term above equates to the last term of the score with respect to α_1. But for $h_1^*(u)$, it will generally not be the case that:

$$\int \left(-E[P_1'|u] - E[P_2'|u] \right) g(u) du = 1 - \int \left(E[P_1|u] - E[P_2|u] \right) h_1^*(u) g(u) du \tag{A.34}$$

; Therefore, the information cannot be 0 for this model.

Q.E.D.

A.12 Proof of Theorem 5.3

Consider the Jacobi matrix. Consider the term

$$F(x_1 - X_1, x_2 - X_2) + R(x_1 - X_1, x_2 - X_2)$$
$$= \int e^{i(t_1(x_1-X_1)+t_2(x_2-X_2))} \mathcal{F}_{11}(t)^{-1}$$

$$\times \int e^{-i(q_1 t_1 + q_2 t_2)} \left(\begin{pmatrix} \frac{(\mathcal{F}_1(t) - \mathcal{F}_{11}(t))}{\mathcal{F}_{11}(t)} \\ \frac{(\mathcal{F}_2(t) - \mathcal{F}_{11}(t))}{\mathcal{F}_{11}(t)} \end{pmatrix} (P_2(q_1),\ P_1(q_1)) + M(q_1, q_2) \right) \Omega(q_1, q_2)\, dq_1\, dq_2$$

from the proof of Theorem 5.1. Then noting that in the limit

$$\Omega(x_1, x_2) = \begin{pmatrix} 0 & -\frac{P_1}{\alpha_2} \\ -\frac{P_2}{\alpha_1} & 0 \end{pmatrix},$$

we notice that $J(x_1, x_2) \to 0$ with a proper limit equal to zero. As a result, we conclude that $\Sigma^{-1} \to 0$ due to the dominated convergence theorem.

Q.E.D.

A.13 Proof of Theorem 5.4

We reduce the dimensionality of the problem and consider a particular parametrization path θ. Then we compute the directional derivative of the moment equation along th eparametrization path. We note that the likelihood of the model can be written as

$$f_\theta(y_1, y_2, x_1, x_2) = \prod_{ij=0}^{1} \mathbf{P}_{ij}(x_1, x_2)^{\mathbf{1}(y_1=i,y_2=j)} \varphi(x_1, x_2)$$

Then we can express the score of the model as

$$S_\theta(y_1, y_2, x_1, x_2) = \frac{\mathbf{P}_{00} y_1 y_2 - \mathbf{P}_{11}(1-y_1)(1-y_2)}{\mathbf{P}_{00}\mathbf{P}_{11}} \dot{\mathbf{P}}_{11}$$
$$+ \frac{(\mathbf{P}_{00} y_1 - \mathbf{P}_{10}(1-y_1))(1-y_2)}{\mathbf{P}_{00}\mathbf{P}_{10}} \dot{\mathbf{P}}_{10}$$
$$\frac{(\mathbf{P}_{00} y_2 - \mathbf{P}_{01}(1-y_2))(1-y_1)}{\mathbf{P}_{00}\mathbf{P}_{01}} \dot{\mathbf{P}}_{01}$$

where $s_{g,\theta}$ is the score of the distribution of errors and s_θ is the score of the distribution of covariates. Then we can express the tangent set of the model as

$$
\begin{aligned}
\mathcal{T} = \Big\{ & \xi_1(x_1, x_2)(\mathbf{P}_{00}y_1y_2 - \mathbf{P}_{11}(1 - y_1)(1 - y_2)) \\
& + \xi_2(x_1, x_2)(\mathbf{P}_{00}y_1 - \mathbf{P}_{10}(1 - y_1))(1 - y_2) \\
& + \xi_3(x_1, x_2)(\mathbf{P}_{00}y_2 - \mathbf{P}_{01}(1 - y_2))(1 - y_1) + t(x_1, x_2) \Big\},
\end{aligned}
$$

where $\xi_i(\cdot)$ for $i = 1, 2, 3$ are square integrable functions, and $E[t(x_1, x_2)] = 0$. Denote $\alpha(\theta) = (\alpha_1(\theta), \alpha_2(\theta))'$ and find the derivative of this vector along the parametrization path. To do that we use the moment system and the system of equilibrium probabilities.

Then we can express the directional derivative of the moment function as

$$
\begin{aligned}
& E\left[A(X_1, X_2)\dot{\mathbf{P}}(X_1, X_2)\right] - E\left[A(X_1, X_2)J(X_1, X_2)\right]\frac{\partial\alpha}{\partial\theta} \\
& - \int A(x_1, x_2)\int P(x_1 - u, x_2 - v) \\
& \quad \times E\left[\frac{\dot{\mathbf{P}}_{11}(X_1, X_2)}{f(X_1, X_2)}\int \frac{e^{i(t_1(u - X_1) + t_2(v - X_2))}}{\mathcal{F}_{11}(t_1, t_2)}dt_1 dt_2\right] f(x_1, x_2)\, du\, dv\, dx_1\, dx_2 = 0.
\end{aligned}
$$

From the proof of Theorem 5.1 we recall that

$$
\int e^{-i(q_1 t_1 + q_2 t_2)}P(q_1, q_2)\, dq_1\, dq_2 = \begin{pmatrix} \mathcal{F}_1(t) - \mathcal{F}_{11}(t) \\ \mathcal{F}_2(t) - \mathcal{F}_{11}(t) \end{pmatrix}.
$$

Thus, application of the Fubbini theorem allows us to represent the last term as

$$
\begin{aligned}
& \int A(x_1, x_2)E\left[\frac{\dot{\mathbf{P}}_{11}(X_1, X_2)}{f(X_1, X_2)}\right. \\
& \qquad \left. \int e^{i(t_1(x_1 - X_1) + t_2(x_2 - X_2))}\begin{pmatrix} \frac{\mathcal{F}_1(t)}{\mathcal{F}_{11}(t_1, t_2)} - 1 \\ \frac{\mathcal{F}_2(t)}{\mathcal{F}_{11}(t_1, t_2)} - 1 \end{pmatrix} dt_1 dt_2\right] f(x_1, x_2)\, dx_1\, dx_2 \\
& = E\left[\frac{\dot{\mathbf{P}}_{11}(X_1, X_2)}{f(X_1, X_2)}\int e^{-i(t_1 X_1 + t_2 X_2)}\mathcal{A}(-t)\begin{pmatrix} \frac{\mathcal{F}_1(t) - \mathcal{F}_{11}(t_1, t_2)}{\mathcal{F}_{11}(t_1, t_2)} \\ \frac{\mathcal{F}_2(t) - \mathcal{F}_{11}(t_1, t_2)}{\mathcal{F}_{11}(t_1, t_2)} \end{pmatrix}' dt_1 dt_2\right],
\end{aligned}
$$

where

$$
\mathcal{A}(t) = E\left[e^{-i(t_1 X_1 + t_2 X_2)}A(X_1, X_2)\right].
$$

Next, considering

$$
K(X_1, X_2) = \int e^{-i(t_1 X_1 + t_2 X_2)}\mathcal{A}(-t)\begin{pmatrix} \frac{\mathcal{F}_1(t) - \mathcal{F}_{11}(t_1, t_2)}{\mathcal{F}_{11}(t_1, t_2)} \\ \frac{\mathcal{F}_2(t) - \mathcal{F}_{11}(t_1, t_2)}{\mathcal{F}_{11}(t_1, t_2)} \end{pmatrix}' dt_1 dt_2,
$$

We can equivalently express

$$K(x_1, x_2) = E\left[A(X_1, X_2) \int e^{i(t_1(x_1 - X_1) + t_2(x_2 - X_2))} \left(\frac{\mathcal{F}_1(t) - \mathcal{F}_{11}(t_1, t_2)}{\mathcal{F}_{11}(t_1, t_2)}, , \frac{\mathcal{F}_2(t) - \mathcal{F}_{11}(t_1, t_2)}{\mathcal{F}_{11}(t_1, t_2)}\right) dt_1 dt_2\right],$$

we can express the directional derivative of the parameter vector as

$$E\left[A(X_1, X_2) J(X_1, X_2)\right] \frac{\partial \alpha}{\partial \theta} = E\left[A(X_1, X_2) \dot{\mathbf{P}}(X_1, X_2) - K(X_1, X_2) \frac{\dot{\mathbf{P}}_{11}(X_1, X_2)}{f(X_1, X_2)}\right]$$

Next, we search for the function $\Psi(y_1, y_2, x_2, x_2)$ such that

$$E\left[A(X_1, X_2) J(X_1, X_2)\right] \frac{\partial \alpha}{\partial \theta} = E\left[\Psi(Y_1, Y_2, X_2, X_2) S_\theta(Y_1, Y_2, X_2, X_2)\right].$$

Then we search for the influence function in the form

$$\Psi(y_1, y_2, x_2, x_2) = a(x_1, x_2)(\mathbf{P}_{00} y_1 y_2 - \mathbf{P}_{11}(1 - y_1)(1 - y_2))$$
$$+ b(x_1, x_2)(\mathbf{P}_{00} y_1 - \mathbf{P}_{10}(1 - y_1))(1 - y_2)$$
$$+ c(x_1, x_2)(\mathbf{P}_{00} y_2 - \mathbf{P}_{01}(1 - y_2))(1 - y_1),$$

where $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are 2×1 vectors. Then consider projection

$$E\left[\Psi S_\theta\right] = -E\left[\dot{\mathbf{P}}_{10}(a + c) + \dot{\mathbf{P}}_{01}(a + b) + \dot{\mathbf{P}}_{11}(b + c)\right].$$

Equating the corresponding coefficients with the expression for the directional derivative and denoting $A_1(\cdot)$ and $A_2(\cdot)$ the first and the second columns of matrix $A(\cdot)$, we obtain

$$a(x_1, x_2) = -\frac{1}{2}\left[A(x_1, x_2)e - \frac{K(x_1, x_2)}{f(x_1, x_2)}\right],$$
$$b(x_1, x_2) = -\frac{1}{2}\left[A(x_1, x_2)e_1 + \frac{K(x_1, x_2)}{f(x_1, x_2)}\right],$$
$$c(x_1, x_2) = -\frac{1}{2}\left[-A(x_1, x_2)e_1 + \frac{K(x_1, x_2)}{f(x_1, x_2)}\right],$$

where $e = (1, 1)'$ and $e_1 = (1, -1)'$. Combining the terms, we can write the influence function as

$$\Psi(y_1, y_2, x_2, x_2) = \left[A(x_1, x_2), -f(x_1, x_2)^{-1} K(x_1, x_2)\right] \begin{pmatrix} y_1(1 - y_2) - \mathbf{P}_{10}(x_1, x_2) \\ (1 - y_1)y_2 - \mathbf{P}_{01}(x_1, x_2) \\ y_1 y_2 - \mathbf{P}_{11}(x_1, x_2) \end{pmatrix}$$

Next we find the variance of the efficient influence function as

$$\Sigma = J^{-1} E\left[\left[A(X_1, X_2), -\frac{K(X_1, X_2)}{f(X_1, X_2)}\right] \text{diag}\left(\mathbf{P}_{10}(1 - \mathbf{P}_{10}), \mathbf{P}_{01}(1 - \mathbf{P}_{01}), \mathbf{P}_{11}(1 - \mathbf{P}_{11})\right) \right.$$
$$\left. \left[A(X_1, X_2), -\frac{K(X_1, X_2)}{f(X_1, X_2)}\right]'\right] J^{-1'},$$

59

with $J = E\left[A(X_1, X_2)\, J(X_1, X_2)\right]$. Denote

$$\Omega(x_1, x_2) = \operatorname{diag}\left(\mathbf{P}_{10}(1 - \mathbf{P}_{10}),\, \mathbf{P}_{01}(1 - \mathbf{P}_{01}),\, \mathbf{P}_{11}(1 - \mathbf{P}_{11})\right).$$

Introduce a two-dimensional $\rho(x_1, x_2)$ and denote

$$\zeta_\rho(x_1, x_2) = \left(\begin{array}{c} \rho(x_1, x_2) \\ E\left[\rho(X_1, X_2) \int e^{i(t_1(x_1 - X_1) + t_2(x_2 - X_2))} \left(\frac{\mathcal{F}_1(t) - \mathcal{F}_{11}(t_1, t_2)}{\mathcal{F}_{11}(t_1, t_2)},\, ,\, \frac{\mathcal{F}_2(t) - \mathcal{F}_{11}(t_1, t_2)}{\mathcal{F}_{11}(t_1, t_2)}\right) dt_1 dt_2\right] \end{array}\right)$$

Then, following Ai and Chen (2003) we can express the semiparametric efficiency bound as

$$\Sigma = E\left[J(X_1, X_2)' \zeta_{\rho^*}(X_1, X_2)' \Omega(X_1, X_2)^{-1} \zeta_{\rho^*}(X_1, X_2) J(X_1, X_2)\right]^{-1},$$

where

$$\rho(\cdot, \cdot)^* = \operatorname{argmin}_\rho \{E\left[J(X_1, X_2)' \zeta_\rho(X_1, X_2)' \Omega(X_1, X_2)^{-1} \zeta_\rho(X_1, X_2) J(X_1, X_2)\right]\}$$

Q.E.D.

B Estimators with optimal rate

B.0.1 Triangular model: Two-step estimator

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

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

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

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

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

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

and

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

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

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

Then we can express the empirical score as

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

This expression can be rewritten as

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

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

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

B.0.2 Iterative estimator

As in the case of triangular binary system we approximate the error density using the normalized Hermite polynomials. We take the sequence $c_n \to \infty$, function $\omega_n(x) = \mathbf{1}\{|x| \le c_n\}$. We introduce 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 probabilties of indicators taking values $y_1 = y_2 = 0$ as

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

Similarly, we approximate the remaining probabilities

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

and

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

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

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

Then we consider profile sample log-likelihood

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

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

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

C Examples of convergence rates for common classes of distributions

Logistic errors with logistic covariates

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

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

Change of variables $z = e^x$ allows us to re-write the expression as

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

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

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

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

Logistic errors with normal covariates

Using the same notations 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 via evaluation

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

As a result, we can use $\nu(c) \equiv 1$ and the bias will vanish. This 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 integral

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

First of all note, that one can find 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 can note 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 $\nu(c) = e^{c^2}$. Then we evaluate the leading component of the bias term as

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

Table 1: Elements of the optimal rate for various distributions

Error distribution	Covariate distribution	
	Logistic	Normal
Logistic	$\nu(c) = c^2,\ \beta(c) = e^{-c}$	$\nu(c) = 1,\ \beta(c) = e^{-c^2}$
	Rate $\sim \sqrt{n/\log n}$	Rate $\sim \sqrt{n}$
Normal	$\nu(c) = e^{c^2},\ \beta(c) = c^2 e^{-c}$	$\nu(c) = c^4,\ \beta(c) = c e^{-c^2}$
	Rate $\sim n^{1/4}$	Rate $\sim \sqrt{n}/\log n$

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

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

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

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

Then the optimal rate has expression $\sqrt{n}/c_n^2$ with c_n solving $c_n^2 e^{-c_n^2}/n = O(1)$.