Quantitative Economics 9 (2018), 995–1017 1759-7331/20180995

Information structure and statistical information in discrete response models

Shakeeb Khan
Department of Economics, Boston College

Denis Nekipelov
Department of Economics, University of Virginia

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

Keywords. Endogenous discrete response, treatment effects, static game, strategic interaction.

JEL classification. C13, C14, C25, C35.

Shakeeb Khan: shakeeb.khan@bc.edu
Denis Nekipelov: denis@virginia.edu
Support from the NSF is gratefully acknowledged. This paper was previously circulated under the title "Information bounds and impossibility theorems for simultaneous discrete response models," and has benefited from helpful comments from conference participants at Cowles, summer 2010, Midwest Econometrics Study Group, fall 2011, and seminar participants at LSE, Penn State, Georgetown, Princeton, UPenn, Stanford, and UC Berkeley. We also greatly benefited from useful comments and insights from S. Athey, H. Hong, M. Jansson, A. Lewbel, A. Mahajan, and J. Powell.

Available at http://qeconomics.org. https://doi.org/10.3982/QE288

1. Introduction

Endogenous regressors are frequently encountered in econometric models, and failure to correct for endogeneity can result in incorrect inference. However, correcting for endogeneity can be particularly difficult in nonlinear models. Recent important work in econometrics has studied the identification and estimation of some nonlinear models with endogenous regressors by adopting what is referred to as a *control function* approach. Seminal papers include Newey, Powell, and Vella (1999), Blundell and Powell (2004), and Imbens and Newey (2009). Their control function approach, however, requires that the system of equations be "triangular" (thereby restricting the type of simultaneity allowed for) and the endogenous regressors be continuously distributed.

In this paper we consider simultaneous discrete response models with discrete endogenous variables.[1] This class includes many important special cases that have received a great deal of attention in both theoretical and empirical work. Examples include strategic compliance models, models of social interactions, and the simultaneous move discrete game model. For this class of models, we are specifically interested in the identifiability of the coefficients of the discrete endogenous variable(s).

The aim of this paper is to relate the ability to recover the parameters of interest to the observable information available to the economic agents in the model. We illustrate this relationship by analyzing four different models, from both a behavioral/economics perspective and from a statistical perspective. In the economic/behavioral analysis of the model, we detail who the economics agents are and which of the variables in the model they observe. In the statistical analysis of the model, we describe what the econometrician observes, what assumptions are made, and what the parameters of interest are and how well they can be estimated. We are primarily interested in quantifying and comparing the *identifiability* of these parameters of interest, by which we mean whether they can be recovered from the data. Our approach is to do so by analyzing what is referred to in classical statistics as the *Fisher information* of these parameters. The Fisher information can be considered an important indicator of the "strength" of the identification, by which we mean how easily the parameters can be recovered from data. The relationship between the two concepts of identification and information dates back to the seminal paper in Rothenberg (1971), who considered identification in parametric models.

We first look at the case of the triangular system of binary equations where the binary outcome in one equation is an explanatory variable in the other equation. The parameter of interest is the coefficient on that binary endogenous variable and is directly related to the treatment effect in that literature. In the other class of models we study—a system

[1]Identification and inference in these models are more complicated than in the continuous endogenous variable case, as first pointed out in the seminal work in Heckman (1978). More recent examples include Chesher (2005), who considers partial identification of general classes of nonlinear, nonseparable models and derives sharp identification regions for parameters of interest. See also Klein, Shan, and Vella (2015) for estimation in this area.

of simultaneous discrete equations with feedback effects—each of the two binary outcomes in the two equations is an explanatory variable in the other equation, such as those that appear in social interaction and static game models, and the parameter of interest characterizes the degree of strategic interaction.

For both the triangular system and the simultaneous system, we consider first what we call a complete information model where the agents have perfect knowledge regarding all the observed variables or, more generally speaking, the information sets of participating economic agents coincide. Then we consider an incomplete information model, where each agent's own actions are only privately observable.

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

The rest of the paper is organized as follows. In the following section we introduce what we call a triangular system of binary equations with complete information. This is a basic binary choice model with a binary endogenous variable determined by a reduced-form model. This model has often been seen in the treatment effects literature. We describe this model first from the behavioral perspective and then from the statistical perspective.

In Section 3 we consider the triangular system with incomplete information, where the agent has uncertainty about the value of the endogenous variable affecting his/her decision. We first detail the information available to the agents in this model, followed by a statistical analysis.

In Section 4 we explore models of simultaneous systems without a triangular structure, first using the example of a two-player simultaneous move game of complete information. Here our economic/behavioral model follows the standard assumptions used to analyze a game theory model, and in the statistical model, the parameters of interest are the two interaction parameters in this game, whose informational content is determined by evaluating their Fisher information.

We then consider a game of incomplete information in which each player has uncertainty about the binary decision of the other player, making this endogenous variable noisily observed. Our behavioral/economic model again follows the standard assumptions in the game theory literature. In the statistical model, of interest here is the identifying power of the uncertainty on the parameter of interest, which, in this case, is the coefficient on that incompletely observed endogenous variable.

Finally, Section 5 concludes the paper by summarizing and suggesting areas for future research. The Appendix collects the proofs of the main theorems. The Supplemental Material, available in a supplementary file on the journal website, http://qeconomics.org/supp/288A/supplement.pdf, provides additional results for the games of complete and incomplete information as well as expanded proof arguments for the proofs in the paper.

2. Triangular discrete response model

In this section we analyze what we refer to as the interaction parameter in a triangular system of binary equations. The model can be described from two perspectives: one behavioral or economic and the other statistical. In the former, we describe what the agent(s) observe, and in the latter, what the econometrician observes from a random sample of realizations of random variables.

2.1 *Economic/behavioral model*

Let Y_1 denote the binary dependent variable of interest, which is assumed to depend on a vector of covariates Z_1 and a single endogenous binary variable Y_2.

We refer to this system as triangular because while the variable Y_2 may have a causal effect on Y_1, in this model we assume a priori that Y_1 does not have a causal effect on Y_2. A leading example of this system could be a single agent model, where Y_2 denotes whether or not the agent receives some sort of treatment, for example, job training (an example in labor economics) or some form of medication (an example in medicine). The variable Y_1 is a binary outcome variable possibly effected by Y_2, for example employment status in labor economics or a binary health status indicator, such as reduction in blood pressure, in the medical example.

In our triangular system of *complete information*, one can model this with two economic agents: one who assigns treatment, and the other who is made aware of treatment assignment and, if assigned treatment, perfectly complies. Here, each agent perfectly observes Y_1 and Y_2 as well as other variables Z_1 and Z_2 that affect each of them. The binary outcomes Y_1 and Y_2 are each determined by a deterministic component—regression and interaction coefficients—and random components, denoted by U and V.

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

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

where $\mathbf{1}$ denotes the indicator function that takes the value 1 if its argument is true and 0 otherwise.

Turning to the model for the endogenous regressor, the binary endogenous variable Y_2 is assumed to be determined by the model

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

where $Z \equiv (Z_1, Z_2)$ is the vector of "instruments" and (U, V) is a pair of random shocks. The subcomponent Z_2 provides the exclusion restrictions in the model and is required to be nondegenerate conditional on $Z_1'\gamma_0$.

We characterize what we call the economic model by which of the variables in a system of equations are observed by each of the two economic agents. In the above system, the set of variables is Y_1, Y_2, Z_1, Z_2, U, V. In the treatment effect model with complete information, we assume each agent observes each of these variables. The binary outcomes Y_1 and Y_2 correspond directly to the sign of the latent variables $Z_1'\gamma_0 + \alpha_0 Y_2 - U$ and $Z'\delta_0 - V$, respectively. As we see later on, this is in contrast to the model with incomplete information, where the agent potentially receiving treatment does not perfectly observe the variable Y_2.

2.2 *Econometric model*

We characterize the econometric model by the set of variables the econometrician observes from a random sample of observations. For the triangular system with complete information, the econometrician observes a random sample of draws from the variables Y_1, Y_2, Z_1, and Z_2, but not U and V. Not observing U and V is precisely what distinguished the econometric model from the economic model where we assumed each of the two agents observed these variables. The aim of the econometrician is to conduct statistical inference on the parameters δ_0, γ_0, and α_0 based on the sample of observed variables and a set of assumptions.

For this model, these assumptions are that the unobserved variables U and V are jointly independent of the observed variables Z_1 and Z_2. The endogeneity of Y_2 in (2.1) arises when U and V are not independent.

This type of econometric/statistical model fits into the class of models considered in Chesher (2005), Vytlacil and Yildiz (2007), and Klein, Shan, and Vella (2015). Here the econometrician is particularly interested in the parameter α_0, which is related to a treatment effect. To simplify exposition, we assume that the parameters δ_0 and γ_0 are known. Thus we introduce single indices $X_1 = Z_1'\gamma_0$ and $X = Z'\delta_0$.[2] The discrete response econometric model can then be written as[3]

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

To give a full characterization of the class of distributions of unobservables (to the econometrician) (U, V) and observables (X_1, X) that we consider, we make the following assumption.

Assumption 1. (i) *The single indices X_1 and X have a joint distribution with full support on $\mathbb{R}^2$, which is not contained in any proper one-dimensional subspace. The parameter of interest is in the interior of a convex compact set $\mathcal{A}$.*

[2] Although additional complications may arise if covariates Z are not truly exogenous, under our assumptions one can regularly identify parameters γ_0 and δ_0. The issues of inference for these parameters are discussed, for instance, in Abrevaya, Hausman, and Khan (2010).

[3] Expressed in this way, we note that the first equation is a binary choice model with one exogenous variable and an endogenous variable. As stated in Assumption 1 below, the exogenous variable is continuously distributed with support on the real line, resembling the "special regressor" introduced in Lewbel (2000). Identification of coefficients on endogenous variables such as Y_2 was also established in Lewbel (2000), but for a model where there was no second equation modeling Y_2.

(ii) *The variables* (U, V) *are independent of* X_1 *and* X, *and have an absolutely continuous density with full support on* $\mathbb{R}^2$ *with an absolutely integrable envelope and joint cumulative distribution function (c.d.f.)* $G(\cdot, \cdot)$. *The partial derivative* $\frac{\partial G(u,v)}{\partial u}$ *exists and is strictly positive on* $\mathbb{R}^2$.

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

Here we focus on the information for α_0 in the statistical model (see, e.g., Ibragimov and Has'minskii (1981), Chamberlain (1986), and Newey (1990) for the relevant definitions). We formally state our first result in the following theorem.[4]

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

We find that under our conditions the parameter α_0 cannot be estimated at the parametric rate. An analogous "impossibility" result[5] for α_0 was found in Khan and Tamer (2010) in a setting where Y_2 was not modeled. Thus Theorem 2.1 shows that the additional structure imposed by modeling Y_2 does not suffice to overturn this result.

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

Furthermore, it is important to emphasize that the conclusion of the theorem does *not* imply that the interaction parameter α_0 is not identified. In fact, as shown in our Technical Report, available in a supplementary file on the journal website, http://qeconomics.org/supp/288B/supplement.pdf, the parameter is point identified and can be consistently estimated, just not at the parametric rate. In that sense the interaction parameter relates to other parameters in the econometrics literature that are "identified at infinity." A notable example of such a parameter is the intercept in a sample selection model; see Andrews and Schafgans (1998) and Heckman (1990). Interestingly, estimating the parameter of interest in this paper—the interaction parameter—shares many features with estimating the intercept term in the selection model. Notably, the rate of convergence depends on *relative tail conditions* on distributions of observable and unobservable variables, loosely analogous to results in Andrews and Schafgans (1998) and Khan and Tamer (2010). For illustrative purposes, the results in the Technical Report demonstrate this point by considering normal and logistic distributions.

[4] The proof of this and all subsequent theorems is provided in the Appendix.

[5] Other impossibility theorems for different models can be found in Chamberlain (1986), Chen and Khan (2003), Khan (2013), and Chen, Khan, and Tang (2016). In all cases they implied that the parameters of interest could generally not be estimated at the parametric rate.

As we see in the next section, this zero information result can be overturned by introducing a treatment effect model with what we refer to as incomplete information on the part of the economic agent.

3. Triangular model with incomplete information

In the previous section, we considered a classical triangular discrete response model and demonstrated that, in general, that model has zero Fisher information for the interaction parameter α_0. Our results suggested that, as is the case with other statistical models with zero information, the optimal convergence rate for the estimator of the interaction parameter is subparametric.

We characterized the model as one of complete information because we assumed the economic agent observed all the variables in the system and knew the coefficients, whereas the econometrician did not observe all the variables and did not know the deterministic parameters. To describe the model of incomplete information and distinguish it from the model of complete information in the previous section, it again proves helpful to describe both the behavioral and statistical models separately.

3.1 *Economic/behavioral model*

In this section, we consider a new model that can be arbitrarily "close" to the triangular model with complete information. We construct this model by adding additional noise to the second (treatment) equation in the triangular system.

Again, we assume two economic agents: one assigning treatment and the other potentially receiving treatment. Consider the model where the endogenous variable Y_2 is now defined as

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

Now we can characterize the behavioral model with incomplete information as follows. The second economic agent (the one potentially receiving treatment) observes X and V as before, and knows the parameter σ. However, this economic agent does not observe η, and knows only that it is drawn from a known distribution. Thus unlike before, this agent does not know the value the variable $X - V - \sigma\eta$. So now this economic agent *cannot* perfectly infer Y_2, but can perfectly infer the probability that $Y_2 = 1$. The first agent (the one assigning treatment) observes all the variables: X_1, X, V, U, and η.

Note that this is in contrast to the complete information model where the second economic agent observed X and V, and could thus perfectly infer the value of Y_2. To complete the description of the behavioral model, the variable Y_1 reflects the response of the second economic agent who does not observe the realization of noise η but observes V and X, and knows the distribution of η and constants $\tilde{\alpha}_0$ and σ. As a result, the response in the first equation can be determined by the agent as

$$Y_1 = \mathbf{1}\big\{X_1 + \tilde{\alpha}_0 E_\eta[Y_2|X, V] - U > 0\big\},$$

where $E_\eta[Y_2|X, V]$ denotes the conditional expectation formed by the agent of receiving treatment Y_2 based on the distribution of η.

3.2 *Statistical model*

In the statistical model, the econometrician observes the draws of the random variables Y_1, Y_2, X_1, and X from a random sample, but does not observe U, V, and η. The constant parameters $\tilde{\alpha}_0$ and σ are unknown, and the aim here is to conduct inference on them given a set of assumptions.

In this model we express our assumption regarding the additional noise component η formally.

Assumption 2. *Suppose that $\eta \perp (U, V)$ and $\eta \perp (X_1, X)$. The distribution of η has a differentiable density with full support on $\mathbb{R}$ and a c.d.f. $\Phi(\cdot)$ that is known by the economic agent and the econometrician.*[6] *In addition, we assume that density $\phi(\cdot)$ is strictly log-concave.*

The variable Y_1 reflects the response of the second agent who does not observe the realization of noise η but observes the error term in her own outcome equation V. As a result, the response in the first equation can be characterized as

$$Y_1 = \mathbf{1}\big\{X_1 + \tilde{\alpha}_0 E_\eta[Y_2|X, V] - U > 0\big\},$$

where the parameter of interest is $\tilde{\alpha}_0$ for which we wish to derive the information.[7] We can express the conditional expectation in the above term as

$$E_\eta[Y_2|x, v] = \Phi\big((x - v)/\sigma\big)$$

and note how this differs from the definition in the economic model, where the term was an indicator function. The constructed discrete response model can then be written as

$$\begin{aligned} Y_1 &= \mathbf{1}\big\{X_1 - U + \tilde{\alpha}_0 E[Y_2|X, V] > 0\big\}, \\ Y_2 &= \mathbf{1}\{X - V + \sigma\eta > 0\}. \end{aligned} \tag{3.1}$$

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

---

[6]We show later in the paper that identification of the interaction parameter does not depend on the knowledge of this distribution (or a specific assumption regarding its structure other than that it has full support). However, the identification of the joint distribution of unobservable variables (U, V) requires this distribution to be known.

[7]We note that here $\tilde{\alpha}_0$ generally denotes a different treatment parameter than before. Specifically, it now measures the response to probability of treatment, as opposed to treatment itself. We argue that it is still a useful parameter to conduct inference on for two reasons: First, as the amount of noise becomes arbitrarily small, the probability of treatment becomes arbitrarily close to the standard treatment status indicator, and the new parameter approximates the standard parameter (the remainder of this section elaborates on this argument with more precision). Second, even if the amount of noise (quantified by σ) is not small, the new parameter will have the same sign as the old one.

As we show, this model has features of the continuous treatment model considered in the literature; examples include Hirano and Imbens (2004), Florens, Heckman, Meghir, and Vytlacil (2008), and Imbens and Wooldridge (2009). While in these cases the economic agent responds to an intrinsically continuous quantity (such as dosage), in our case the continuity of treatment is associated with uncertainty of the agent regarding the treatment. Outside of the treatment effect setting, binary choice models with a continuous endogenous variable are also studied in Blundell and Powell (2004), who demonstrate the attainability of positive information for the coefficient on the endogenous variable.

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

The incomplete information triangular model presented here also places the standard triangular model considered in the previous section in the context of the models with strategic compliance of the treated subjects, as in Chassang and Snowberg (2010). The complete information triangular model characterizes the compliance behavior in the local average treatment effect (LATE) model of Angrist and Imbens (1995), Abadie, Angrist, and Imbens (2002), and Imbens (2009) as a special case: the orthogonality assumption of LATE is satisfied if the error terms U and V, in our terminology, are independent. The variable Y_2 corresponds to the "treatment assignment" (i.e., the binary instrument of the LATE model) and the variable Y_1 corresponds to the compliance decision. The complete information model represents the case where the treated subject knows all of the inputs into the treatment decision. As a result, the compliance decision will be correlated with the treatment decision unless the unobservables in the two decisions are orthogonal. Once the treatment decision contains noise, which may come from the deliberate treatment randomization (e.g., through a placebo) or can suffer from measurement error, the treated subject may only react to the expected treatment. This setting motivates the triangular model with treatment uncertainty.

We can illustrate the structure of the model using Figure 1. Panel (a) in Figure 1 corresponds to the classical binary triangular system and panels (b)–(d) correspond to the triangular system with incomplete information. The panels show the areas of joint support of U and V that correspond 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 0 or 1. Figure 1(b)–(d) shows the area where, for a given quantile q, the probability of Y_2 equal 0 or 1 exceeds $1-q$. The noise in the second equa-

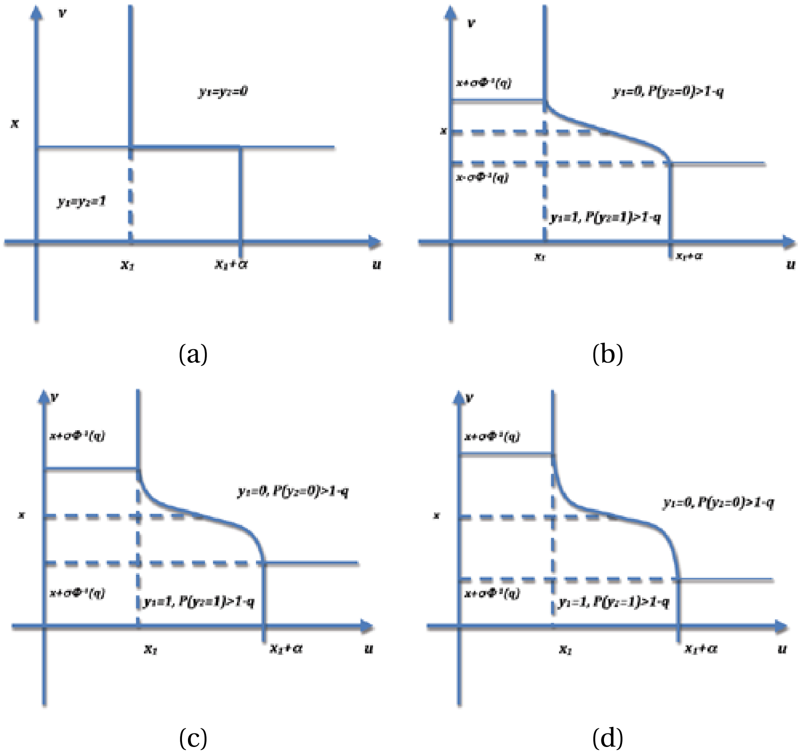

FIGURE 1. Incomplete information triangular model.

tion decreases from panel (d) to panel (b), which in the limit approaches the figure on panel (a).

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

Given that this is a new statistical model, we need to establish first that the model is identified from the data. The following theorem considers the identification of the interaction parameter $\tilde{\alpha}_0$.

THEOREM 3.1. *Under Assumptions* 1 *and* 2*, the interaction parameter* $\tilde{\alpha}_0$ *in model* (3.1) *is identified.*

The proof of identification for parameter $\tilde{\alpha}_0$ is based on the idea of relating the expectations of Y_1 to the expectation of Y_2. First of all note that

$$\int \frac{\partial}{\partial x} E[Y_1 | X_1 = x_1, X = x] \, dx_1 = \tilde{\alpha}_0 \int \int g\left(x_1 + \tilde{\alpha}_0 \Phi\left(\frac{x-v}{\sigma}\right), v\right) \frac{1}{\sigma} \phi\left(\frac{x-v}{\sigma}\right) dv \, dx_1$$
$$= \tilde{\alpha}_0 \int g_v(v) \frac{1}{\sigma} \phi\left(\frac{x-v}{\sigma}\right) dv.$$

Recall that $E[Y_2 | X = x] = \int \Phi(\frac{x-v}{\sigma}) g_v(v) \, dv$, meaning that

$$\tilde{\alpha}_0 = \int \frac{\frac{\partial}{\partial x} E[Y_1 | X_1 = x_1, X = x]}{\frac{\partial}{\partial x} E[Y_2 | X = x]} \, dx_1,$$

which identifies the interaction parameter. Then, for sufficiently small variance of the noise σ^2, we note that $E[Y_1 Y_2 | X_1 = x_1, X = x] \approx G(x_1 + \tilde{\alpha}_0, x)$, which allows us to identify the joint distribution of the error terms (U, V).

Having established the identification of the parameter of interest, we analyze its Fisher information. We find that for any finite variance σ^2 (which can be arbitrarily small), the information for $\tilde{\alpha}_0$ in the incomplete information triangular model is strictly positive. Moreover, the Fisher information of the strategic interaction parameter $\tilde{\alpha}_0$ approaches 0 when the variance of noise shrinks to 0. In other words, *the smaller is the informational asymmetry between agents, the smaller is the Fisher information of the interaction parameter* $\tilde{\alpha}_0$.

We state this in the following theorem.

Theorem 3.2. *Suppose that Assumptions* 1 *and* 2 *are satisfied.*

(i) *For any* $\sigma > 0$, *the information for* $\tilde{\alpha}_0$ *in the triangular model of incomplete information* (3.1) *is strictly positive.*

(ii) *As* $\sigma \to 0$, *the information for* $\tilde{\alpha}_0$ *in the triangular model of incomplete information* (3.1) *converges to* 0.

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

3.2.1 *Convergence rate for the interaction parameter* $\tilde{\alpha}_0$ The previous subsection proved that the triangular model with incomplete information has positive Fisher information for any amount of noise added to the second equation. Our results, therefore, guarantee that the semiparametric efficiency bound is finite. We note that the analyzed model has two unknown nonparametric components: the distribution of covariates and the distribution of unobserved heterogeneity. Due to the independence of the unobserved heterogeneity and the observed covariates, and the fact that the distribution of covariates does not depend on the parameter $\tilde{\alpha}_0$, this parameter is fully characterized by the expectations of and the covariance between the observed binary variables Y_1 and Y_2 conditional on covariates. In other words, the parameter of interest is characterized by a system of conditional moment equations. The efficiency bound for parameter$\tilde{\alpha}_0$ is computed explicitly in the Technical Report and our results are its derivation based on the result for the semiparametric efficiency bound in conditional moment systems provided in Ai and Chen (2003), which also suggests an efficient method of moments estimator. Our efficiency result provides the semiparametric efficiency bound for the new discrete response model.

As discussed in detail in the Supplemental Material, the optimal rate of convergence is parametric ($\sqrt{n}$) and the minimum variance of the estimator converging at a parametric rate corresponds to the semiparametric efficiency bound.

4. Nontriangular systems

In this section, we consider two simultaneous systems of equations without the triangular structure of the previous sections. Examples of these models are frequently encountered in game theory as in empirical industrial organization. These models aim to study games of complete and incomplete information. Since the results here are analogous to those attained for the models studied in the previous sections, we describe them less formally and leave many of the technical details of the main results to the Supplemental Material.

Games of complete information

A leading example of a nontriangular system is a two-player discrete game with complete information (e.g., Bjorn and Vuong (1985) and Tamer (2003)). Following the pattern we used in the previous sections, we distinguish the behavioral models from the statistical one.

Economic model A binary game of complete information is characterized by the players' deterministic payoffs, strategic interaction coefficients, and random payoff components U and V. There are two players $i = 1, 2$ and the action space of each player consists of two points $A_i = \{0, 1\}$ with the actions denoted $Y_i \in A_i$. The payoff to player 1 from choosing action $Y_1 = 1$ can be characterized as a function of observed covariates and player 2's action,

$$Y_1^* = Z_1'\gamma_0 + \alpha_{10} Y_2 - U,$$

where Z_1 denotes a vector of covariates and the payoff of player 2 from choosing action $Y_2 = 1$ is characterized as

$$Y_2^* = Z_2'\delta_0 + \alpha_{20} Y_1 - V,$$

where Z_2 denotes a vector of covariates. The variables $(\gamma_0, \delta_0, \alpha_{10}, \alpha_{20})$ denote coefficients and, analogous to before, the econometrician is primarily interested in the parameters α_{10} and α_{20}, which often are referred to as the interaction parameters in the empirical industrial organization literature. Because of this, for convenience of both notation and analysis, we assume the parameters γ_0 and δ_0 are known, and we change change notation to $X_1 = Z_1'\gamma_0$ and $X_2 = X_2'\delta_0$. We normalize the payoff from action $Y_i = 0$ to 0, and we assume that realizations of covariates X_1 and X_2 are commonly observed by the players along with realizations of the variables U and V.

Here 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 constitutes a *Nash equilibrium* if they reflect the best responses to the rival's equilibrium actions. This is the equilibrium concept we are assuming players use in our behavioral model.

Statistical model In the statistical model, the econometrician observes a random sample of equilibrium outcomes and covariates, but not the realizations of the random variables U and V. The observed equilibrium actions are described by a pair of binary equations in the statistical model,

$$\begin{aligned} Y_1 &= \mathbf{1}\{X_1 + \alpha_{10} Y_2 - U > 0\}, \\ Y_2 &= \mathbf{1}\{X_2 + \alpha_{20} Y_1 - V > 0\}, \end{aligned} \tag{4.1}$$

where the unobserved variables U and V have an unknown joint distribution and the econometrician is interested in conducting statistical inference on the parameters α_{10} and α_{20}.

As noted in Tamer (2003), the system of simultaneous discrete response equations (4.1) has a fundamental problem of *indeterminacy*. To resolve this, we impose an *equilibrium selection* mechanism, based on randomization, similar to that imposed in Bjorn and Vuong (1985). The specifics of this mechanism are described in detail in the Supplemental Material. As explained there, the selection mechanism addresses both the *incoherency* and *incompleteness* that may arise in these models.[8] This mechanism is admittedly a strong condition that we deliberately impose to demonstrate how difficult it is to identify the interaction parameters in this model. Specifically, while the assumption eliminates the difficulties that arise from incompleteness and incoherency, we will show that it does not suffice to attain positive Fisher information for the interaction parameters. Our formal analysis is based on regularity conditions on observed and unobserved random variables that are standard and yield two main results: (i) point identification for the interaction parameters and (ii) zero Fisher information for these same parameters.

These main results fully illustrate why the zero Fisher information of the interaction parameters is not related to the lack of their point identification or the multiplicity of equilibria that arise in these models. Thus we conclude that the estimation and inference for the interaction parameters are nonstandard even in a model simplified with a strong equilibrium selection rule.

[8]Following the terminology introduced in Tamer (2003), incoherency refers to the nonexistence of an equilibrium and incompleteness refers to multiplicity of equilibria.

Games with incomplete information

Here we modify the simultaneous equation model to allow for incomplete information, first describing both the economic and statistical models.

Economic model This model is based on standard two-player game theory models with incomplete information. Game theoretical results have demonstrated that the introduction of what is referred to in that literature as payoff perturbations leads to a reduction in the number of equilibria.[9]

In the two-player game with incomplete information, we again interpret the binary variables Y_1 and Y_2 as the actions of player 1 and player 2. Each player is characterized by the deterministic payoff (corresponding to linear indices X_1 and X_2), an interaction parameter, unobserved heterogeneity terms U and V, and what we refer to here as payoff perturbations, denoted by η_1 and η_2. The payoff of player 1 from action $Y_1 = 1$ can be represented as

$$Y_1^* = X_1 + \tilde{\alpha}_{10} Y_2 - U - \sigma\eta_1,$$

while the payoff from action $Y_1 = 0$ is normalized to 0. We impose the following informational assumptions, which are similar to those imposed in the incomplete information triangular system: Specifically, we assume that η_1 and η_2 are privately observed by the two players, meaning player 1 observes η_1 but not η_2, and analogously for player 2. We assume $\eta_1 \perp \eta_2$ and both satisfy Assumption 2.

Thus in this model, player 1 observes X_1, X_2, U, V, and η_1 but does not observed η_2, and player 2 observes X_1, X_2, U, V, and η_2 but does not observe η_1. We note that the strategy of player i is a mapping from the observable (to the agents) variables into actions:

$$(X_1, X_2, U, V, \eta_i) \mapsto \{0, 1\}.$$

Also, player i forms beliefs regarding the action of the rival. Provided that η_1 and η_2 are independent, the beliefs will be functions only of U, V, and linear indices. Thus, if $P_i(X_1, X_2, U, V)$ are players' beliefs regarding actions of opponent players, then the strategy, for instance, of player 1, can be characterized as a random variable

$$\begin{aligned} Y_1 &= \mathbf{1}\{E[Y_1^* \mid X_1, X_2, U, V, \eta_1] > 0\} \\ &= \mathbf{1}\{X_1 - U + \tilde{\alpha}_{10} P_2(X_1, X_2, U, V) - \sigma\eta_1 > 0\}. \end{aligned} \tag{4.2}$$

Similarly, the strategy of player 2 can be written as

$$Y_2 = \mathbf{1}\{X_2 - V + \tilde{\alpha}_{20} P_1(X_1, X_2, U, V) - \sigma\eta_2 > 0\}. \tag{4.3}$$

We note that when σ approaches 0, the payoffs in the incomplete information model are identical to those in the complete information model and are observable by both players.

[9]See the seminal work of Harsanyi (1995). Multiplicity of equilibria can still be an important issue in games of incomplete information as noted in Sweeting (2009), Aradillas-Lopez (2010), and de Paula and Tang (2012).

To characterize the Bayes–Nash equilibrium[10] in the simultaneous move game of incomplete information, we consider a pair of strategies defined by (4.2) and (4.3). Moreover, the beliefs of players are consistent with their action probabilities conditional on the information set of the rival. Taking into consideration the independence of player types η_i and the fact that their c.d.f. is known, we can characterize the pair of equilibrium beliefs as a solution to a system of nonlinear equations. This is explained in detail in the Supplemental Material.

This system of nonlinear equations can have multiple solutions.[11] To resolve the uncertainty over equilibria, we again assume an equilibrium selection rule based on randomization as described in the Supplemental Material.

Statistical model Here the econometrician observes a random sample of equilibrium outcomes and covariates, but does not observe realizations of U, V, η_1, and η_2, and knows the distributions of η_1 and η_2, but not of U, V.

In the Supplemental Material, we provide conditions under which two main results are established: (i) the parameters $\tilde{\alpha}_{10}$ and $\tilde{\alpha}_{20}$ are each identified, and (ii) each has positive Fisher information. Regarding the second result, we find that for any finite value σ, the Fisher information in the model of the incomplete information game is strictly positive, implying estimability at the parametric rate. As in the incomplete information triangular model, the Fisher information of the interaction parameters approaches 0 whenever σ does.

5. Conclusions

This paper considers identification and inference in simultaneous equation models with discrete endogenous variables. We analyzed both triangular systems, where the parameter of interest is the coefficient of a discrete endogenous variable, and nontriangular systems, focusing on simultaneous discrete games, where we are interested in the strategic interaction parameters.

Our main findings are that the complete information models have zero Fisher information under our conditions, whereas the incomplete information models can have positive information, which facilitates the conducting of inference on these parameters. In the nontriangular case, our zero Fisher information result implies that the difficulty in identification of the strategic interaction parameters is not due to multiplicity of equilibria, as we obtain this result even after introducing an equilibrium selection rule.

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

[10] This is the equilibrium concept often used in the incomplete information game literature; see, for example, Sweeting (2009), de Paula and Tang (2012), Wan and Xu (2014), and Xu (2014).

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

Appendix: Proofs

A.1 *Proof of Theorem* 2.1

To simplify our argument, we assume that coefficients β_0 and δ_0 are known. We thus refer to the indices in each equation as X_1 and X, respectively, and denote realized values by x_1 and x, respectively. To derive the information of the model, we follow the approach in Chamberlain (1986) by demonstrating that for each triangular model generated by a distribution satisfying the conditions of Theorem 2.1, we can construct a parametric submodel passing through that model for which the information for the parameter α is equal to 0. Suppose that Γ contains all distributions of errors that satisfy the conditions of Theorem 2.1 along with all distributions of indices $x_1 = \beta_0 z_1$ and $x = \delta_0 z$ for which $E[q(X_1, X)^2] < \infty$ for $q(\cdot, \cdot)$ defined in the statement of the theorem such that x_1 and x have a continuous joint distribution with full support on $\mathbb{R}^2$. We first construct the likelihood function of the model and introduce the notation

$$P^{11}(t_1, t) = \Pr(U \le t_1, V \le t) = G(t_1, t),$$

$$P^{01}(t_1, t) = \Pr(U > t_1, V \le t),$$

$$P^{10}(t_1, t) = \Pr(U \le t_1, V > t),$$

and

$$P^{00}(t_1, t) = \Pr(U > t_1, V > t).$$

The likelihood function is determined by the density

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

with respect to the measure μ defined on $\Omega = \{0, 1\}^2 \times \mathbb{R}^2$ such that for any Borel set A in $\mathbb{R}^2$,

$$\mu(\{1, 1\} \times A) = \mu(\{1, 0\} \times A) = \mu(\{0, 1\} \times A) = \mu(\{0, 0\} \times A) = \nu(A),$$

where $P((X_1, X) \in A) = \int_A d\nu$. Let $h : \mathbb{R}^2 \mapsto \mathbb{R}$ be a continuously differentiable function supported on a given compact set S with its derivative being continuous in the interior of that compact set such that $\frac{\partial h(u,v)}{\partial u} \ge B$ for some constant B on that compact set. We define $\tilde{\Lambda}$ as the collection of paths through the original model that we design as

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

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

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

and

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

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

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

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

with

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

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

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

and the derivative with respect to λ can be expressed as

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

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

$$\begin{aligned}I_{\lambda,\alpha} &\leq 4\int (\psi_\alpha - \psi_\lambda)^2 d\mu \\ &= \int \frac{G_v(x)}{G(x_1 + \alpha_0, x)\big(G_v(x) - G(x_1 + \alpha_0, x)\big)} \\ &\quad \times \left(\frac{\partial G(x_1 + \alpha_0, x)}{\partial u}\right)^2 h^2(x_1 + \alpha_0, x)\, d\nu(x_1, x).\end{aligned}$$

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

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

Following Chamberlain (1986), we let $L_2(\pi)$ denote the space of measurable functions $q : R^m \to R$ such $\int q^2\, d\pi < \infty$, allowing us to conclude that

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

Chamberlain (1986) demonstrates that the space of differentiable functions with compact support is dense in $L_2(\pi)$. Moreover, we require the derivative of h to be continuous in the interior of its support. Let S be the support of h. We take $\varepsilon^* > 0$ and construct the set S_{ε^*} to be a compact subset of S such that the Euclidean distance of the boundary of S from the boundary of S_{ε^*} is at least ε^*, where ε^* is selected such that $\pi(S \setminus S_{\varepsilon^*}) < \sqrt{\varepsilon}$. Since the set of differentiable functions is dense in $L_2(\pi)$, for any $\varepsilon > 0$ we can find $a \in C_c^2(\mathbb{R}^2)$ (where $C_c^2(R^2)$ denotes the set of real-valued functions on R^2 that have compact support and continuous partial derivatives of order 2) such that $\|a\|_{L_2(\pi)} < \sqrt{\varepsilon}$. The derivative $\frac{\partial a(u,v)}{\partial u}$ is continuous in the interior of S. Provided that $S_{\varepsilon^*} \subset S$, this derivative is continuous on the entire set S_{ε^*} and, due to its compactness, it is uniformly continuous there. As a result, there exists $M = \sup_{S_{\varepsilon^*}} |\frac{\partial a(u,v)}{\partial u}|$. There also exists $M' = \sup_S |a|$. Then we pick the direction h^* as a function with support on S such that $h^* = \frac{B}{2}(a/M)$ in S_{ε^*}. Then we note that

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

As a result, $I_{\lambda,\alpha} \leq \frac{B^2(M'+1)^2}{M^2}\varepsilon$. As the choice of ε was arbitrary, this proves that $\inf_{\lambda\in\tilde{\Lambda}} I_{\lambda,\alpha} = 0$. □

A.2 *Proof of Theorem* 3.1

Our model is generated by two binary variables, Y_1 and Y_2. As a result, its parametric components are fully characterized by conditional probabilities $E[Y_1|x_1, x]$, $E[Y_2|x_1, x]$, and $E[Y_1 Y_2|x_1, x]$. Here we provide a simple argument that demonstrates the identification of the parameter $\tilde{\alpha}_0$. Take points $x > x'$ in the support of X and consider

$$\begin{aligned} &E[Y_2|X_1 = x_1, X = x] - E\big[Y_2|X_1 = x_1, X = x'\big] \\ &\quad = \int \mathbf{1}\left\{u - \tilde{\alpha}_0\Phi\left(\frac{x-v}{\sigma}\right) \leq x_1 \leq u - \tilde{\alpha}_0\Phi\left(\frac{x'-v}{\sigma}\right)\right\} g(u, v)\, du\, dv. \end{aligned} \tag{A.1}$$

For each fixed pair of $x > x'$, this function is absolutely integrable inside any interval $[-c, c]$. In fact, the set of points (u, v),

$$S_c = \left\{(u, v) : x_1 + \tilde{\alpha}_0\Phi\left(\frac{x'-v}{\sigma}\right) \leq u \leq x_1 + \tilde{\alpha}_0\Phi\left(\frac{x-v}{\sigma}\right), -c \leq x_1 \leq c\right\},$$

is a closed connected subset of $\mathbb{R}^2$ and thus has finite measure with respect to $g(\cdot, \cdot)$. Moreover, for any sequence $c \to \infty$, the limit of measures of sets is well defined and is bounded by 1. Therefore, the improper integral of (A.1) on $\mathbb{R}$ over x_1 is well defined. Taking this integral, we note that

$$\begin{aligned} &\int\int \mathbf{1}\left\{u - \tilde{\alpha}_0\Phi\left(\frac{x-v}{\sigma}\right) \leq x_1 \leq u - \tilde{\alpha}_0\Phi\left(\frac{x'-v}{\sigma}\right)\right\} g(u, v)\, du\, dv\, dx_1 \\ &\quad = \tilde{\alpha}_0 \int \left(\Phi\left(\frac{x-v}{\sigma}\right) - \Phi\left(\frac{x'-v}{\sigma}\right)\right) g(u, v)\, du\, dv. \end{aligned}$$

Recall that

$$E[Y_1|X = x] = \int \Phi\left(\frac{x-v}{\sigma}\right) g(u, v)\, du\, dv.$$

Therefore, we can write the final expression that identifies the parameter $\tilde{\alpha}_0$ as

$$\tilde{\alpha}_0 = \int \big(E[Y_2|X_1 = x_1, X = x] - E\big[Y_2|X_1 = x_1, X = x'\big]\big) \\ /\big(E[Y_1|X = x] - E\big[Y_1|X = x'\big]\big)\, dx_1.$$

Therefore, the parameter $\tilde{\alpha}_0$ is identified. □

A.3 *Proof of Theorem* 3.2

A.3.1 *Proof of part (i)* In the proof of Theorem 3.1, we presented an explicit expression for the parameter of interest. To compute the information corresponding to the parameter of interest, we construct the log-likelihood of the model by explicitly expressing the probabilities

$$P_{11}(x_1, x; \tilde{\alpha}_0, g) = \int \mathbf{1}\left\{x_1 - u + \tilde{\alpha}_0 \Phi\left(\frac{x-v}{\sigma}\right) > 0\right\} \Phi\left(\frac{x-v}{\sigma}\right) g(u, v)\, du\, dv,$$
$$P(x_1, x; \tilde{\alpha}_0, g) = \int \Phi\left(\frac{x-v}{\sigma}\right) g_v(v)\, dv,$$

and

$$Q(x_1, x; \alpha, g) = \int \mathbf{1}\left\{x_1 - u + \tilde{\alpha}_0 \Phi\left(\frac{x-v}{\sigma}\right) > 0\right\} g(u, v)\, du\, dv.$$

We can then express all probabilities of interest as

$$P_{01}(x_1, x; \tilde{\alpha}_0, g) = P(x_1, x; \tilde{\alpha}_0, g) - P_{11}(x_1, x; \tilde{\alpha}_0, g),$$
$$P_{10}(x_1, x; \tilde{\alpha}_0, g) = Q(x_1, x; \tilde{\alpha}_0, g) - P_{11}(x_1, x; \tilde{\alpha}_0, g),$$

and

$$P_{00}(x_1, x; \tilde{\alpha}_0, g) = 1 - Q(x_1, x; \tilde{\alpha}_0, g) - P(x_1, x; \tilde{\alpha}_0, g) + P_{11}(x_1, x; \tilde{\alpha}_0, g),$$

and express the derivatives of the probabilities of interest as

$$\frac{\partial P_{11}(x_1, x; \alpha, g)}{\partial \alpha} = \int \Phi\left(\frac{x-v}{\sigma}\right) g\left(x_1 + \alpha \Phi\left(\frac{x-v}{\sigma}\right), v\right) dv \equiv D_1(x_1, x; \alpha, g),$$
$$\frac{\partial Q(x_1, x; \alpha, g)}{\partial \alpha} = \int g\left(x_1 + \alpha \Phi\left(\frac{x-v}{\sigma}\right), v\right) dv \equiv D_2(x_1, x; \alpha, g).$$

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

$$r(y_1, y_2, x_1, x; \tilde{\alpha}_0, g)^{1/2} = y_1 y_2 P_{11}(x_1, x; \tilde{\alpha}_0, g)^{1/2} + y_1(1 - y_1) P_{10}(x_1, x; \tilde{\alpha}_0, g)^{1/2} \\ + (1 - y_1) y_2 P_{01}(x_1, x; \tilde{\alpha}_0, g)^{1/2} \\ + (1 - y_1)(1 - y_1) P_{00}(x_1, x; \tilde{\alpha}_0, g)^{1/2}.$$

We can express the mean-square derivative with respect to $\tilde{\alpha}_0$ as

$$
\begin{aligned}
&\psi_{\tilde{\alpha}}(y_1, y_2, x_1, x) \\
&\quad = \frac{1}{2}\big[y_1 y_2 P_{11}(x_1, x; \tilde{\alpha}_0, g)^{-1/2} - (1 - y_1) y_2 P_{01}(x_1, x; \tilde{\alpha}_0, g)^{-1/2}\big] \\
&\qquad \times D_1(x_1, x; \tilde{\alpha}_0, g) \\
&\qquad + \frac{1}{2}\big[(1 - y_1)(1 - y_1) P_{00}(x_1, x; \tilde{\alpha}_0, g)^{1/2} - y_1(1 - y_1) P_{10}(x_1, x; \tilde{\alpha}_0, g)^{1/2}\big] \\
&\qquad \times \big(D_1(x_1, x; \tilde{\alpha}_0, g) - D_2(x_1, x; \tilde{\alpha}_0, g)\big).
\end{aligned}
$$

Thus, we can express the information for the parameter $\tilde{\alpha}_0$ as

$$
I_{\tilde{\alpha}_0} = 4 \int (\psi_{\tilde{\alpha}_0})^2 \, d\mu.
$$

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

$$
\pi_1(A) = \int_A \frac{P_1(x_1, x; \tilde{\alpha}_0, g)}{P_{11}(x_1, x; \tilde{\alpha}_0, g)\big(P_1(x_1, x; \tilde{\alpha}_0, g) - P_{11}(x_1, x; \tilde{\alpha}_0, g)\big)} d\nu(x_1, x)
$$

and

$$
\pi_2(A) = \int_A \frac{1 - P_1(x_1, x; \tilde{\alpha}_0, g)}{P_{00}(x_1, x; \tilde{\alpha}_0, g)\big(1 - P_1(x_1, x; \tilde{\alpha}_0, g) - P_{00}(x_1, x; \tilde{\alpha}_0, g)\big)} d\nu(x_1, x).
$$

We can then express the information of the model as

$$
I_{\tilde{\alpha}} = \big\| D_1(x_1, x; \tilde{\alpha}_0, g) \big\|^2_{L_2(\pi_1)} + \big\| D_1(x_1, x; \tilde{\alpha}_0, g) - D_2(x_1, x; \tilde{\alpha}_0, g) \big\|^2_{L_2(\pi_2)}. \tag{A.2}
$$

We construct the measure π^* that minorizes the Radon–Nikodym density of measures π_1 and π_2, meaning that $\frac{d\pi^*}{d\nu} = \min\{\frac{d\pi_1}{d\nu}, \frac{d\pi_2}{d\nu}\}$. Based on this structure of the measure, we can write

$$
I_{\tilde{\alpha}} \geq \big\| D_1(x_1, x; \tilde{\alpha}_0, g) \big\|^2_{L_2(\pi^*)} + \big\| D_2(x_1, x; \tilde{\alpha}_0, g) - D_1(x_1, x; \tilde{\alpha}_0, g) \big\|^2_{L_2(\pi^*)}.
$$

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

$$
D_1(x_1, x; \tilde{\alpha}_0, g) = \sigma \int w(t) g\big(x_1 + \tilde{\alpha}_0 w(t), x - \sigma t\big) \, dt
$$

and

$$
D_2(x_1, x; \tilde{\alpha}_0, g) - D_1(x_1, x; \tilde{\alpha}_0, g) = \sigma \int \big(1 - w(t)\big) g\big(x_1 + \tilde{\alpha}_0 w(t), x - \sigma t\big) \, dt.
$$

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

$$I_{\tilde{\alpha}_0} \geq \|D_2(x_1, x; \alpha_0, g)\|^2_{L_2(\pi^*)} \geq \|D_2(x_1, x; \tilde{\alpha}_0, g)\mathbf{1}_S\|^2_{L_2(\pi^*)}$$
$$\geq C\sigma^2 \left\| \int_{\mathbb{R}} M(t)\,dt \right\|^2 \geq C\sigma \left\| \int_{-B}^{B} M(t)\,dt \right\|^2 \geq \frac{1}{2} C R^2 \varepsilon \sigma^2 > 0.$$

Therefore, the information corresponding to the parameter $\tilde{\alpha}_0$ is strictly positive whenever $\sigma > 0$. □

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. [1003]

Abrevaya, J., J. Hausman, and S. Khan (2010), "Testing for causal effects in a generalized regression model with endogenous regressors." *Econometrica*, 78, 2043–2061. [999]

Ahn, H. and C. F. Manski (1993), "Distribution theory for the analysis of binary choice under uncertainty with nonparametric estimation of expectations." *Journal of Econometrics*, 56 (3), 291–321. [1002]

Ai, C. and X. Chen (2003), "Efficient estimation of models with conditional moment restrictions containing unknown functions." *Econometrica*, 71 (6), 1795–1843. [1006]

Andrews, D. W. K. and M. Schafgans (1998), "Semiparametric estimation of the intercept of a sample selection model." *Review of Economic Studies*, 65, 497–517. [1000]

Angrist, J. D. and G. W. Imbens (1995), "Two-stage least squares estimation of average causal effects in models with variable treatment intensity." *Journal of the American Statistical Association*, 90 (430), 431–442. [1003]

Aradillas-Lopez, A. (2010), "Semiparametric estimation of a simultaneous game with incomplete information." *Journal of Econometrics*, 157, 409–431. [1008]

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010), "Computing equilibria in static games of incomplete information using the all-solution homotopy." Working paper. [1009]

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. [1006, 1007]

Blundell, R. and J. Powell (2004), "Endogeneity in semiparametric binary response models." *Review of Economic Studies*, 71 (3), 655–679. [996, 1003]

Chamberlain, G. (1986), "Asymptotic efficiency in semi-parametric models with censoring." *Journal of Econometrics*, 32 (2), 189–218. [1000, 1010, 1011]

Chassang, S. and E. Snowberg (2010), "Selective trials: A principal-agent approach to randomized controlled experiments." Discussion paper, National Bureau of Economic Research. [1003]

Chen, S. and S. Khan (2003), "Rates of convergence for estimating regression coefficients in heteroskedastic discrete response models." *Journal of Econometrics*, 117, 245–278. [1000]

Chen, S., S. Khan, and X. Tang (2016), "Informational content of special regressors in heteroskedastic discrete response models." *Journal of Econometrics*, 193, 162–182. [1000]

Chesher, A. (2005), "Nonparametric identification under discrete variation." *Econometrica*, 73, 1525–1550. [996, 999]

de Paula, A. and X. Tang (2012), "Inference of signs of interaction effects in simultaneous games with incomplete information." *Econometrica*, 80, 143–172. [1008, 1009]

Florens, J. P., J. Heckman, C. H. D. 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. [1003]

Harsanyi, J. C. (1995), "A new theory of equilibrium selection for games with complete information." *Games and Economic Behavior*, 8 (1), 91–122. [1008]

Heckman, J. (1978), "Dummy endogenous variables in a simultaneous equation system." *Econometrica*, 46, 931–959. [996]

Heckman, J. (1990), "Varieties of selection bias." *American Economic Review*, 80 (2), 313–318. [1000]

Hirano, K. and G. W. Imbens (2004), "The propensity score with continuous treatments." In *Applied Bayesian Modeling and Causal Inference From Incomplete-Data Perspectives*, 73–84, Wiley, New York. [1003]

Ibragimov, I. and R. Has'minskii (1981), *Statistical Estimation Asymptotic Theory*. Springer, New York. [1000]

Imbens, G. and W. Newey (2009), "Identification and estimation of triangular simultaneous equations models without additivity." *Econometrica*, 77 (5), 1481–1512. [996]

Imbens, G. W. (2009), "Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009)." Discussion paper, National Bureau of Economic Research. [1003]

Imbens, G. W. and J. M. Wooldridge (2009), "Recent developments in the econometrics of program evaluation." *Journal of Economic Literature*, 47 (1), 5–86. [1003]

Khan, S. (2013), "Distribution free estimation of heteroskedastic binary response models using probit criterion functions." *Journal of Econometrics*, 172, 168–182. [1000]

Khan, S. and E. Tamer (2010), "Irregular identification, support conditions and inverse weight estimation." *Econometrica*, 78, 2021–2042. [1000]

Klein, R., C. Shan, and F. Vella (2015), "Estimation of marginal effects in semiparametric selection models with binary outcomes." *Journal of Econometrics*, 185, 82–94. [996, 999]

Lewbel, A. (2000), "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables." *Journal of Econometrics*, 97, 145–177. [999]

Manski, C. F. (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*, 259–275, Cambridge University Press, Cambridge. [1002]

Newey, W. (1990), "Semiparametric efficiency bounds." *Journal of Applied Econometrics*, 5 (2), 99–135. [1000]

Newey, W., J. Powell, and F. Vella (1999), "Nonparametric estimation of triangular simultaneous equations models." *Econometrica*, 67, 565–603. [996]

Rothenberg, T. (1971), "Identification in parametric models." *Econometrica*, 39, 577–591. [996]

Sweeting, A. (2009), "The strategic timing of radio commercials: An empirical analysis using multiple equilibria." *RAND Journal of Economics*, 40, 710–742. [1008, 1009]

Tamer, E. (2003), "Incomplete bivariate discrete response model with multiple equilibria." *Review of Economic Studies*, 70, 147–165. [1006, 1007]

Vytlacil, E. J. and N. Yildiz (2007), "Dummy endogenous variables in weakly separable models." *Econometrica*, 75, 757–779. [999]

Wan, Y. and H. Xu (2014), "Semiparametric identification of binary decision games of incomplete information with correlated private signals." *Journal of Econometrics*, 182, 235–246. [1009]

Xu, H. (2014), "Estimation of discrete games with correlated types." *Econometrics Journal*, 17, 241–270. [1009]

Co-editors Orazio Attanasio and Rosa L. Matzkin handled this manuscript.

Manuscript received 6 June, 2012; final version accepted 3 October, 2017; available online 5 December, 2017.