Information Bounds and Impossibility Theorems for Simultaneous Discrete Response Models

Shakeeb Khan
Department of Economics, Duke University
Denis Nekipelov
Department of Economics, UC-Berkeley

This version: October 2010

ABSTRACT. This paper considers information bounds for semiparametric models with endogenous regressors, under different support conditions on the outcome variable as well as the endogenous regressor. Our base case is the model where both the outcome variable and the endogenous regressor are binary. Our first result is that under stated conditions, the information for the coefficient on the endogenous variable is 0, implying the impossibility of estimating the coefficient at the parametric rate. Furthermore, we attain a moment condition directly from the likelihood function which equates the parameter of interest to an inverse weighted moment condition. An immediate implication of this result is the need to conduct alternative methods for inference, such as that suggested in Khan and Tamer (2010) or Abrevaya, Hausman, and Khan (2010) Extending from the base model by allowing the support of the endogenous variable to be more rich, we find that the bound can be positive suggesting the existence of regular estimators converging at the parametric rate. Finally, while our first set of results applies to triangular discrete systems, we then consider non-triangular systems, of which a leading example is a simultaneous game (e.g. Tamer (2003) andAradillas-Lopez (2005)). For these models, under stated conditions we attain analogous impossibility theorems and inverse weight moment conditions for parameters of interest, suggesting the need for alternative inference procedures here as well.

JEL Classification: C14, C25, C13.

Key Words: Endogeneity, causal effects, semiparametric estimation, impossibility theorems.

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 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 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. For the endogenous regressors, this restriction is necessary to identify the average structural function (ASF) and its derivatives (i.e., the structural effects).

For a single binary endogenous regressor, Vytlacil and Yildiz (2007) establish conditions under which it is possible to identify the average treatment effect(ATE) in non-linear models. Identification requires variation in exogenous regressors (including the instruments for the binary endogenous regressor) that has the same effect upon the outcome variable as a change in the binary endogenous regressor.

In this paper we also consider identification of parameters of interest in nonlinear models

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

with endogenous variables. Parameters of interest will include the coefficients on endogenous variables, the ATE and the ASF. What we are particularly interested in here is the *information* for such parameters. As is well known in the literature, the information for parameters relates direction to the *regularity* of their identification, as well as the root-n estimability of these parameters. In doing so, we are sometimes able to establish a relationship between the level of information and the richness of support of the endogenous variable. Specifically, we are able to show that positive information is often not possible in nonlinear models when the endogenous variable takes only 2 values, resulting in various impossibility theorems.

However, we show in this paper that positive information in these nonlinear models with endogenous variables can be attained by modifying the problem in a number of ways. One approach is to consider a richer support for the endogenous variable. Another approach is the consider a function of the parameter of interest that reduces the size of the parameter space, implying root-n estimability.

In this paper, we consider necessary and sufficient conditions for *regular* identification of the parameters of interest, specifically, the coefficient on an endogenous variable in a model involving a system of nonlinear equations. As we will show a direct link can be made between the information for the parameter of interest when adopting what we will refer to as the "control function" approach, compared to what we will refer to as an instrumental variable approach. We will also establish that the strength of the link of the two approaches is directly tied to the number of values the endogenous variable takes. At one extreme, where the endogenous variable can only take two values, we establish an equivalence result between the identification for the parameter of interest from the control function approach and the i.v. approach introduced in Lewbel (1998), in the sense that the parameter of interest can be equated to an expectation of a term involving the inverse of a density function.

This has two consequences. One is that both approaches result in a "just identified" GMM setup. Second, since both approaches result in an expectation with inverse weighting, neither results in *regular* identification, precluding the possibility of root-n consistent estimation.

Next, we show that the two approaches begin to differ when the endogenous variable takes more than two values. Specifically, we can show that the control function approach can translate into a set of additional moment conditions, resulting in an overidentified system of equations, thereby achieving efficiency gains. In contrast, the IV approach does not allow us to directly exploit the richer support of the endogenous variable. Therefore, the two approaches represent the classic efficiency robustness tradeoff, but in the context of irregular identification.

Finally, we extend the case where the endogenous variables can take infinitely many values; This will result in infinitely many moment conditions, and we show the control function approach then results in an identification result that is analogous to that found in a *conditional* moment equality model, opening up the possibility of attaining *regular* identification and root-n consistent estimation. This would then be similar to a result found in Blundell and Powell (2004), who were able to attain root-n convergence of their estimator, but only in the case when the endogenous variable was continuously distributed on the real line. Thus, in terms of attaining regular identification, this can only be attained by adopting a control function approach, and only when the support of the endogenous variable is sufficiently rich.

The rest of the paper is organized as follows. The following section introduces the base model we consider- a binary choice with one endogenous variable, for which to complete the specification of the (triangular) model, a reduced-form model is utilized for the endogenous regressor. Focusing first on the the case of a binary endogenous regressor, we first derive the information bound for its coefficient. Section 3-4 Considers 2 modifications of the base model which we show can help attain positive information and root-n estimability of the parameters of interest. The first modification is to increase the richness of the support of the endogenous variable so it is now allowed to take more than two values. The second modification we explore is to take a function of the parameter of interest (i.e. the regression coefficient, the ATE, or ASF) that results in a smaller parameter space. In both settings we are indeed able to attain positive information and thus estimate the (new) parameters at the parametric rate. Sections 4 and 5 carry the same exercise, but for two different models, repectively. The first one with a richer support for the dependent variable, specifically an ordered model with an endogenous regressor. The second is a nontriangular system- specifically a discrete game with complete information. As in the previous sections, we derive the information for parameters of interest in these models under a variety of support and parameter space conditions. Section 6 concludes by summarizing results and proposing suggestions for future research.

2 The Base 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, is given by

$$y_1 = 1[z_1'\beta_0 + \alpha_0 y_2 - \epsilon > 0]. \tag{2.1}$$

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:

$$y_2 = 1[z'\delta_0 - \eta > 0], \tag{2.2}$$

where $z \equiv (z_1, z_2)$ is the vector of "instruments" and η 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 (ϵ, η) is independent of z. Endogeneity of y_2 in (2.1) arises when ϵ and η 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. 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 the parameter α_0 has 0 information.*

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).

Proof: 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 2.1 *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. $\int\int g(u,v)dudv = 1$. *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. $\int |\partial g(u,v)/\partial u| dudv < \infty$

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

Definition 2.2 Λ *consists of the paths:*

$$\lambda(\delta_1, \delta_2) = g_0[1 + (\delta_1 - \delta_{10})h_1][1 + (\delta_2 - \delta_{20})h_2] \tag{2.3}$$

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

$$\int\int g_0(u,v)h_1(u)h_2(v)dudv = 0 \tag{2.4}$$

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

$$\sum_{i,j=0,1} d_{ij} \log \mathcal{P}_{ij}(\alpha, \delta_1, \delta_2) \tag{2.5}$$

We can then take the derivative of the above term with respect to α, evaluated at $\alpha = \alpha_0$ and $\delta_1 = \delta_{10}, \delta_2 = \delta_{20}$. W.l.o.g. we will only 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

$$d_{11}\mathcal{P}_{11}(\alpha_0, \delta_{10}, \delta_{20})^{-1} \int\int \delta(u - x_1 - \alpha_0)I[v < x]dudv \tag{2.6}$$

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

$$d_{11}\mathcal{P}_{11}(\alpha_0, \delta_{10})^{-1} \int\int I[u < x_1 + \alpha_0]I[v < x]g_0(u,v)h_1(u)dudv \tag{2.7}$$

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

$$\mathcal{P}_{11}(\alpha_0, \delta_0)^{-1}\left(\int\int \delta(u - x_1 - \alpha_0)I[v < x]g_0(u,v)dudv - \int\int I[u < x_1 + \alpha_0]I[v < x]g_0(u,v)h_1(u)dudv\right)^2 \tag{2.8}$$

To show our impossibility result we need to find an $h_1^*(u)$ that sets the above term to 0.

This can be accomplished by setting

$$h_1^*(u) = \delta(u - x_1 - \alpha_0) \tag{2.9}$$

■

Remark 2.1 *It is worth comparing this result in those found in Vytlacil and Yildiz(2007). In their paper, they establish the important result that identify the average effect of the dummy endogenous variable without relying on the use of large support conditions. The use of large support conditions is usually related to the regularity of the identification of the parameter of interest. In their case the large support condition is replaced with a relative index support condition. In the linear index specifications, as considered in this paper, this corresponds to the assumption that* $|\alpha_0| \leq d(z_1'\beta_0)$ *where* $d()$ *denotes the diameter of the interval. This effectively translate into a parameter space constraint if one were to impose a bounded regressor support condition. Not imposing such as assumption results in the* irregularity *of the identification and the impossibility theorem we attain.*

Thus we consider two modifications to see if we can attain positive information. The first is to reduce the parameter space of the parameter of interest by taking some noninvertible function of α_0, such as the sign, or some discrete subset of the parameter space that α_0 lies in. Let $\theta_0 = f(\alpha_0)$, where $f()$ denotes this noninvertible function. Then we have the following theorem:

Theorem 2.2 *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 the parameter θ_0 has positive information when* $f() =$.

A second way to attain positive information, is to y_2 to have richer support, i.e. by taking more values. We would replace the equation for y_2 with

$$y_2 = g(z'\delta_0 + \eta > 0), \tag{2.10}$$

where $g()$ takes a continuum ofvalues; then we have the following theorem:

Theorem 2.3 *Suppose the model is characterized by the two equations (2.10) and (2.1), and suppose that w.l.o.g., z has full support on R^k, then the parameter θ_0 has positive information.*

3 Identification in triangular systems with limited dependent variables

We can provide a constructive identification argument based on the properties of the observable joint distributions of binary outcomes in the triangular system, conditional on the linear index. We assume that our object of interest is the coefficient on the interaction between lower and upper binary outcomes in the triangular system. In the beginning of the analysis we will not focus on the coefficients in the linear index itself. We will treat the entire linear index as a "regressor"and denote $x_1 = z_1'\beta_0$ and $x = z'\delta_0$.

3.1 Systems of binary response models

We begin with considering triangular system

$$\begin{aligned} y_1 &= \mathbf{1}\{x_1 + \alpha_0 y_2 - \epsilon \geq 0\}, \quad (3.1)\\ y_2 &= \mathbf{1}\{x - \eta \geq 0\}. \quad (3.2) \end{aligned}$$

We make the following assumption regarding the structure of the variables.

Assumption 1 *The joint distribution of (ϵ, η) has absolutely continuous density function and is assumed distributed independent of x_1, x. The distribution of (x_1, x) is non-degenerate and has continuous support.*

The likelihood function of the model incorporates the probabilities of occurrence of outcomes y_1 and y_2 such that

$$\ell(y_1, y_2, x_1, x_2) = \mathcal{P}(y_1, y_2 \mid x, x_1)\phi(x, x_1), \tag{3.3}$$

where $\phi(\cdot)$ is the joint distribution of linear indices. The conditional likelihood can be fully characterized by three conditional probabilities

$$\begin{aligned}
\mathcal{P}(1,1|x_1,x) &= \int \mathbf{1}\{\epsilon \leq x_1 + \alpha_0\}\, \mathbf{1}\{\eta \leq x\}\, f_{\epsilon\eta} d\epsilon d\eta, && (3.4)\\
\mathcal{P}(1,0 \mid x_1,x) &= \int \mathbf{1}\{\epsilon \leq x_1\}\, \mathbf{1}\{\eta > x\}\, f_{\epsilon\eta} d\epsilon d\eta, && (3.5)\\
\mathcal{P}(0,0 \mid x_1,x) &= \int \mathbf{1}\{\epsilon > x_1\}\, \mathbf{1}\{\eta > x\}\, f_{\epsilon\eta} d\epsilon d\eta. && (3.6)\\
& && (3.7)
\end{aligned}$$

We can make the observation that with sufficient support of indices x and x_1 the conditional probability of either outcome $y_1 = 1$, $y_2 = 0$ or $y_1 = 0$, $y_2 = 0$ identifies the joint distribution of the unobserved shocks (ϵ, η). This result is summarized in the following theorem. The proof of the theorem is constructive and we will base our further discussion on it. Therefore, we outline its elements below.

Theorem 3.1 *Suppose that supp$(\epsilon, \eta) \subseteq$ supp$(X) \times$ supp(X_1). Then the density $f_{\epsilon\eta}(\cdot)$ of unobserved shocks (η, ϵ) in the system (3.1) is identified by the probability of either outcome $y_1 = 1$, $y_2 = 0$ or $y_1 = 0$, $y_2 = 0$ conditional on x_1 and x.*

Proof: Consider Fourier transforms of each of the probabilities with respect to the linear indices. Then for the first probability we can write

$$\begin{aligned}
\mu_{11}(t_1,t_2) &= \int \mathcal{P}(1,1 \mid x_1,x) e^{-it_1x_1 - it_2x} dx_1\, dx && (3.8)\\
&= \int\int e^{-it_1x_1 - it_2x} \mathbf{1}\{\epsilon \leq x_1 + \alpha_0\}\, \mathbf{1}\{\eta \leq x\}\, f_{\epsilon\eta} d\epsilon d\eta dx_1 dx. && (3.9)
\end{aligned}$$

We make a change of variables $u = \epsilon - x_1$ and $v = \eta - x$ and notice that

$$\begin{aligned}
&\int\int e^{-it_1x_1 - it_2x} \mathbf{1}\{\epsilon \leq x_1 + \alpha_0\}\, \mathbf{1}\{\eta \leq x\}\, f_{\epsilon\eta} d\epsilon d\eta dx_1 dx && (3.10)\\
= &\int \mathbf{1}\{u \leq \alpha_0\}\, e^{iut_1} du \int \mathbf{1}\{v \leq 0\}\, e^{ivt_2} dv \int e^{-it_1\epsilon - it_2\eta} f_{\epsilon\eta} d\epsilon d\eta && (3.11)
\end{aligned}$$

Letting $\delta(\cdot)$ denote the delta function, we find that

$$\int \mathbf{1}\left\{u \leq \alpha_0\right\} e^{iut_1} du = e^{it_1\alpha_0}(\pi\delta(-t_1) + \frac{1}{it_1}), \tag{3.12}$$

and we can use the notation $\chi(\cdot)$ for the characteristic function corresponding to the unknown distribution $f_{\epsilon\eta}$. We also denote $\Gamma(t) = \pi\delta(-t) + \frac{1}{it}$ and $\bar{\Gamma}(t) = \pi\delta(-t) - \frac{1}{it}$. Then Fourier transformations of three choice probabilities can be expressed as

$$\mu_{11}(t_1, t_2) = e^{\alpha_0 it_1}\Gamma(t_1)\Gamma(t_2)\chi(t_1, t_2), \tag{3.13}$$

$$\mu_{10}(t_1, t_2) = \Gamma(t_1)\bar{\Gamma}(t_2)\chi(t_1, t_2), \tag{3.14}$$

$$\mu_{00}(t_1, t_2) = \bar{\Gamma}(t_1)\bar{\Gamma}(t_2)\chi(t_1, t_2), \tag{3.15}$$

$$\tag{3.16}$$

Note that neither of the last two equations of system (3.13) contains the unknown interaction parameter α_0. We can for instance use the last equation to recover the density function of the unobserved shocks. We multiply both sides by $t_1 t_2$ and make the inverse Fourier transform:

$$- \frac{1}{(2\pi)^2}\int e^{it_1\epsilon + it_2\eta} t_1 t_2 \mu_{00}(t_1, t_2) dt_1 dt_2 \tag{3.17}$$

$$= \frac{1}{(2\pi)^2}\int e^{it_1\epsilon + it_2\eta}(1 - i\pi t_1\delta(-t_1))(1 - i\pi t_2\delta(-t_2))\chi(t_1, t_2) dt_1 dt_2 \tag{3.18}$$

$$= \frac{1}{(2\pi)^2}\int e^{it_1\epsilon + it_2\eta}\chi(t_1, t_2) dt_1 dt_2 = f_{\epsilon\eta}(\epsilon, \eta), \tag{3.19}$$

where we used the fact that $\int f(t)\delta(t)\, dt = f(0)$ by definition of the delta-function. This means that we can identify the density of interest from the Fourier transformation of the conditional probability of realization $y_1 = y_2 = 0$.

Q.E.D.

Note that the argument in the previous theorem exploits the tail behavior of the conditional probability. The idea behind the analysis is to find sets of values of indices x_1 and x for which both binary variables are equal to zero. These events most likely occur when the indices are close to the extremes in the support of the unobservable errors. We can provide further steps for identification of the interaction parameter α_0. We use the steps in the proof of Theorem 3.1 to identify this parameter. As before, we use $\mu_{y_1 y_2}(\cdot)$ for the Fourier transformation of the probability for pair (y_1, y_2) conditional on the index variables. Then we use the result of Theorem 3.1 and identify the joint density of ϵ and η from the last equation of (3.13). Then using the first equation we can write

$$e^{-\alpha_0 it_1} = \frac{\Gamma(t_1)\Gamma(t_2)\chi(t_1, t_2)}{\mu_{11}(t_1, t_2)}. \tag{3.20}$$

We can directly extract α_0 from this expression as

$$\alpha_0 = -t_1^{-1}\,\mathrm{Arg}(\frac{\Gamma(t_1)\Gamma(t_2)\chi(t_1,t_2)}{\mu_{11}(t_1,t_2)}). \tag{3.21}$$

where Arg denotes the phase of the complex number.

It is clear that the complex number $e^{-\alpha_0 i t_1}$ determining the interaction coefficient is defined by an inverse weight function $\mu_{11}(t_1,t_2)$. Given that this is a non-parametric object, finding coefficient α_0 is similar to using inverse density weighting. In fact, note that the inverse Fourier transformation leads to

$$\delta(x_1-\alpha_0)\delta(x) = \frac{1}{(2\pi)^2}\int e^{it_1x_1+it_2x}\frac{\Gamma(t_1)\Gamma(t_2)\chi(t_1,t_2)}{\mu_{11}(t_1,t_2)}\,dt_1dt_2. \tag{3.22}$$

Multiplication by x_1 and taking expectation of both sides leads to the following expression for the interaction coefficient

$$\alpha_0 = -\frac{1}{(2\pi)^2}E\left[x_1\int e^{it_1x_1+it_2x}\frac{\Gamma(t_1)\Gamma(t_2)\chi(t_1,t_2)}{\mu_{11}(t_1,t_2)}\,dt_1dt_2\right]. \tag{3.23}$$

This structure of the coefficient shows the direct relationship between this model and the literature on inverse density weighting. One can completely skip the step of estimating the characteristic function of the unobservable shocks if the distribution of the unobserved shocks is symmetric.

Corollary 3.1 *Suppose that the distribution of the unobserved shocks is symmetric, i.e.* $f_{\epsilon\eta}(\epsilon,\eta) = f_{\epsilon\eta}(-\epsilon,-\eta)$. *Then the interaction coefficient can be identified solely from the Fourier transforms of probabilities* $y_1=y_2=1$ *and* $y_1=y_2=0$

Proof: Note that if the distribution is symmetric, then the characteristic function will be symmetric as well. Provided the Hermitian property and symmetry of the characteristic function, we can take the complex conjugate of (3.13) to obtain:

$$\bar{\mu}_{00}(t_1,t_2) = \Gamma(t_1)\Gamma(t_2)\chi(t_1,t_2). \tag{3.24}$$

This means that

$$e^{-\alpha_0 i t_1} = \frac{\bar{\mu}_{00}(t_1,t_2)}{\mu_{11}(t_1,t_2)}. \tag{3.25}$$

As a result, we can write

$$\alpha_0 = -\frac{1}{(2\pi)^2} E\left[x_1 \int e^{it_1x_1+it_2x} \frac{\bar{\mu}_{00}(t_1,t_2)}{\mu_{11}(t_1,t_2)}\, dt_1dt_2\right]. \tag{3.26}$$

Thus, we can obtain the interaction coefficient of interest from estimating the conditional probabilities and averaging the ratio of their Fourier transformation over the values of the linear index.
Q.E.D.

4 Systems with ordered endogenous variables

We start with a simple system with the ordered endogenous regressor.

$$y_1 = \mathbf{1}\left\{x_1 + \alpha_0 y_2 - \epsilon \geq 0\right\}, \tag{4.1}$$

$$y_2 = \mathbf{1}\left\{x - \eta \geq 0\right\} - \mathbf{1}\left\{x - \eta < -1\right\}. \tag{4.2}$$

In this case the discrete endogenous variable takes values -1,0 and 1, depending on whether $x-\eta$ crosses thresholds -1 and 0. Then the conditional likelihood of the model is characterized by the following probabilities

$$\mathcal{P}\ \ (1,1 \mid x_1,x) = \int \mathbf{1}\left\{\epsilon \leq x_1+\alpha_0\right\}\mathbf{1}\left\{\eta \leq x\right\} f_{\epsilon\eta} d\epsilon d\eta, \tag{4.3}$$

$$\mathcal{P}\ \ (1,0 \mid x_1,x) = \int \left[\mathbf{1}\left\{\epsilon \leq x_1\right\}\mathbf{1}\left\{\eta \leq x+1\right\} - \mathbf{1}\left\{\epsilon \leq x_1\right\}\mathbf{1}\left\{\eta \leq x\right\}\right] f_{\epsilon\eta} d\epsilon d\eta, \tag{4.4}$$

$$\mathcal{P}\ \ (1,-1 \mid x_1,x) = \int \mathbf{1}\left\{\epsilon \leq x_1-\alpha_0\right\}\mathbf{1}\left\{\eta > x+1\right\} f_{\epsilon\eta} d\epsilon d\eta, \tag{4.5}$$

$$\mathcal{P}\ \ (0,1 \mid x_1,x) = \int \mathbf{1}\left\{\epsilon > x_1+\alpha_0\right\}\mathbf{1}\left\{\eta \leq x\right\}\ f_{\epsilon\eta} d\epsilon d\eta, \tag{4.6}$$

$$\mathcal{P}\ \ (0,0 \mid x_1,x) = \int \left[\mathbf{1}\left\{\epsilon > x_1\right\}\mathbf{1}\left\{\eta \leq x+1\right\} - \mathbf{1}\left\{\epsilon > x_1\right\}\mathbf{1}\left\{\eta \leq x\right\}\right] f_{\epsilon\eta} d\epsilon d\eta, \tag{4.7}$$

$$\tag{4.8}$$

We can use the same approach that we applied to the binary model. We make Fourier transforms of both sides for all probabilities. We can again use notation

$$\mu_{y_1y_2}(t_1,t_2) = \int e^{-it_1x_1-it_2x}\mathcal{P}(y_1,y_2 \mid x_1,x)\, dx_1\, dx. \tag{4.9}$$

We use Fubini theorem and the property of the Fourier transform for step functions. This leads to the expression

$$\begin{aligned}
\mu_{11}(t_1,t_2) &= e^{\alpha_0 it_1}\Gamma(t_1)\Gamma(t_2)\chi(t_1,t_2), && (4.10)\\
\mu_{10}(t_1,t_2) &= (e^{it_2}-1)\Gamma(t_1)\bar{\Gamma}(t_2)\chi(t_1,t_2), && (4.11)\\
\mu_{1,-1}(t_1,t_2) &= e^{-\alpha_0 it_1+it_2}\Gamma(t_1)\bar{\Gamma}(t_2)\chi(t_1,t_2), && (4.12)\\
\mu_{01}(t_1,t_2) &= e^{\alpha_0 it_1}\bar{\Gamma}(t_1)\Gamma(t_2)\chi(t_1,t_2), && (4.13)\\
\mu_{00}(t_1,t_2) &= (e^{it_2}-1)\bar{\Gamma}(t_1)\Gamma(t_2)\chi(t_1,t_2). && (4.14)\\
& && (4.15)
\end{aligned}$$

As one can see, this system becomes over-identified for α_0. However, given that each equation contains the unknown characteristic function, the solution for $e^{\alpha_0 it_1}$ will always contain the inverse of this non-parametric object. In fact, we can see that one can identify the joint distribution of the errors, for instance, from the last equation. For this expression we can rearrange the terms, multiply both sides by t_1t_2 and perform an inverse Fourier transformation of both sides

$$\begin{aligned}
& \frac{1}{(2\pi)^2}\int e^{it_1\epsilon+it_2\eta}t_1t_2\frac{\mu_{00}(t_1,t_2)}{e^{it_2}-1}dt_1dt_2 && (4.16)\\
= & \frac{1}{(2\pi)^2}\int e^{it_1\epsilon+it_2\eta}(1-i\pi t_1\delta(-t_1))(1+i\pi t_2\delta(-t_2))\chi(t_1,t_2)\,dt_1dt_2 && (4.17)\\
= & \frac{1}{(2\pi)^2}\int e^{it_1\epsilon+it_2\eta}\chi(t_1,t_2)\,dt_1dt_2 = f_{\epsilon\eta}(\epsilon,\eta). && (4.18)
\end{aligned}$$

Then, for instance, using the first equation, and using subsequent averaging over x_1 and x one can express

$$\alpha_0 = -\frac{1}{(2\pi)^2}E[x_1\int e^{it_1x_1+it_2x}\frac{\Gamma(t_1)\Gamma(t_2)\chi(t_1,t_2)}{\mu_{11}(t_1,t_2)}\,dt_1dt_2]. \quad (4.19)$$

Similarly, one can obtain the coefficient from the third and the fourth equations leading to

$$\alpha_0 = \frac{1}{(2\pi)^2}E\left[x_1\int e^{it_1x_1+it_2(x+1)}\frac{\Gamma(t_1)\bar{\Gamma}(t_2)\chi(t_1,t_2)}{\mu_{1,-1}(t_1,t_2)}\,dt_1dt_2\right], \quad (4.20)$$

and

$$\alpha_0 = -\frac{1}{(2\pi)^2}E\left[x_1\int e^{it_1x_1+it_2x}\frac{\bar{\Gamma}(t_1)\Gamma(t_2)\chi(t_1,t_2)}{\mu_{01}(t_1,t_2)}\,dt_1dt_2\right]. \quad (4.21)$$

Provided that these equations over-identify the coefficient of interest, one can use a weighted sum of these expressions to improve efficiency. However, each of the expressions still contains a non-parametric weighting function $\mu_{y_1y_2}(\cdot)$ in the denominator.

An obvious generalization of the considered example is the ordered response model with a countable number of thresholds that cover the entire support. In fact, suppose that y_2 takes integer values from $-\infty$ to $+\infty$ and the triangular system takes the form

$$
\begin{aligned}
y_1 &= \mathbf{1}\left\{x_1 + \alpha_0 y_2 - \epsilon \geq 0\right\}, && (4.22)\\
y_2 &= k \text{ if } x - \eta \in (\lambda_{k-1},\, \lambda_k]. && (4.23)\\
& && (4.24)
\end{aligned}
$$

We start with the case where thresholds λ_k are known, but in practice they can be unknown as well. Then the likelihood of the model is characterized by the probabilities

$$
\begin{aligned}
\mathcal{P} \quad (1, k \mid x_1, x) &= \int \mathbf{1}\left\{\epsilon \leq x_1 + \alpha_0 k\right\}\left[\mathbf{1}\left\{\eta \leq x + \lambda_k\right\} - \mathbf{1}\left\{\eta \leq x + \lambda_{k-1}\right\}\right] f_{\epsilon\eta} d\epsilon d\eta && (4.25)\\
\mathcal{P} \quad (0, k \mid x_1, x) &= \int \mathbf{1}\left\{\epsilon > x_1 + \alpha_0 k\right\}\left[\mathbf{1}\left\{\eta \leq x + \lambda_k\right\} - \mathbf{1}\left\{\eta \leq x + \lambda_{k-1}\right\}\right] f_{\epsilon\eta} d\epsilon d\eta && (4.26)\\
& && (4.27)
\end{aligned}
$$

Then application of the Fourier transform leads to

$$
\begin{aligned}
\mu_{1,k}(t_1, t_2) &= e^{i\alpha_0 k t_1}(e^{i\lambda_k t_2} - e^{i\lambda_{k-1} t_2})\Gamma(t_1)\Gamma(t_2)\chi(t_1, t_2), && (4.28)\\
\mu_{0,k}(t_1, t_2) &= e^{i\alpha_0 k t_1}(e^{i\lambda_k t_2} - e^{i\lambda_{k-1} t_2})\bar{\Gamma}(t_1)\Gamma(t_2)\chi(t_1, t_2). && (4.29)
\end{aligned}
$$

Then we can identify the joint distribution of the errors from the combination $y_1 = y_2 = 0$, leading to the identifying expression

$$
f_{\epsilon\eta}(\epsilon, \eta) = \frac{1}{(2\pi)^2} \int e^{it_1\epsilon + it_2\eta} t_1 t_2 \frac{\mu_{00}(t_1, t_2)}{e^{i\lambda_0 t_2} - e^{i\lambda_{-1} t_2}} dt_1 dt_2. \qquad (4.30)
$$

Then to identify α_0 we can use all $k \neq 0$:

$$
e^{-i\alpha_0 k t_1} = \frac{(e^{i\lambda_k t_2} - e^{i\lambda_{k-1} t_2})\Gamma(t_1)\Gamma(t_2)\chi(t_1, t_2)}{\mu_{1,k}(t_1, t_2)}. \qquad (4.31)
$$

We can perform the inverse Fourier transform and average over the single indices. This leads to the representation for each k

$$
\alpha_0 = -\frac{1}{(2\pi)^2 k} E\left[x_1 \int e^{it_1 x_1 + it_2 x} \frac{(e^{i\lambda_k t_2} - e^{i\lambda_{k-1} t_2})\Gamma(t_1)\Gamma(t_2)\chi(t_1, t_2)}{\mu_{1,k}(t_1, t_2)}\, dt_1 dt_2\right]. \qquad (4.32)
$$

Each of these estimates will be consistent for a given k. For the sake of efficiency one can investigate the use of a linear combination of the estimates obtained for different k. Then we can use weights $\omega_k > 0$ that sum to 1 to construct an combination of available estimates for α_0.

$$\alpha_0 = -\frac{1}{(2\pi)^2} \sum_{\substack{k=-\infty \\ k\neq 0}}^{+\infty} \frac{2\omega_k}{k} E\left[x_1 \int e^{it_1x_1+it_2x} \frac{(e^{i\lambda_k t_2} - e^{i\lambda_{k-1}t_2})\Gamma(t_1)\Gamma(t_2)\chi(t_1,t_2)}{\mu_{1,k}(t_1,t_2)}\, dt_1 dt_2\right] \tag{4.33}$$

Note that in this case the properties of the estimate will depend on the limit of the functional series:

$$\zeta(t_1,t_2) = \sum \lim_{\substack{k=-\infty \\ k\neq 0}}^{+\infty} \frac{2\omega_k}{k} \frac{(e^{i\lambda_k t_2} - e^{i\lambda_{k-1}t_2})\Gamma(t_1)\Gamma(t_2)\chi(t_1,t_2)}{\mu_{1,k}(t_1,t_2)}. \tag{4.34}$$

Note that this limit may have a more regular behavior than the inverses $\mu_{1,k}(t_1,t_2)^{-1}$.

5 Systems with continuous endogenous variables

In the previous subsection the endogenous variable took integer values. We can consider the case where the endogenous variable can take a continuum of values. We analyze the model where the error term η is additively separable with the linear index and the model takes the form

$$y_1 = \mathbf{1}\left\{x_1 + \alpha_0 y_2 - \epsilon \geq 0\right\}, \tag{5.1}$$

$$y_2 = \varphi(x-\eta). \tag{5.2}$$

We assume that function $\varphi(\cdot)$ is strictly monotone. This model can be interpreted as a continuous generalization of the discrete threshold model considered in the previous subsection. The likelihood function of the model can be derived from the joint probability

$$\mathcal{P}\quad (y_1 = 1, y_2 \geq y \mid x_1, x) = \int \mathbf{1}\left\{\epsilon \leq x_1 + \alpha_0 y\right\} \mathbf{1}\left\{\eta \leq x - \varphi^{-1}(y)\right\} f_{\epsilon\eta} d\epsilon d\eta. \tag{5.3}$$

We make the Fourier transform of both sides of this expression by integrating over x_1 and x. For a fixed y we can represent this Fourier transform as

$$\mu_{1,y}(t_1,t_2) = e^{i\alpha_0 y t_1 - i\varphi^{-1}(y)t_2}\Gamma(t_1)\Gamma(t_2)\chi(t_1,t_2). \tag{5.4}$$

Note that the structural function $\varphi(\cdot)$ is identifiable from the mean normalization on the errors and the independence of the errors from the covariates, implying that

$$E[\varphi^{-1}(y_2) \mid x_1, x] = x. \tag{5.5}$$

Then we can identify the joint distribution of the errors using the point $y_1 = y_2 = 0$, for which

$$\mu_{00}(t_1, t_2) \;\; = \;\; e^{-i\varphi^{-1}(0)t_2}\Gamma(t_1)\Gamma(t_2)\chi(t_1, t_2). \tag{5.6}$$

Then we can recover the density of interest as

$$f_{\epsilon\eta}(\epsilon, \eta) \;\; = \;\; \frac{1}{(2\pi)^2}\int e^{it_1\epsilon + it_2(\eta + \varphi^{-1}(0))}t_1 t_2 \mu_{00}(t_1, t_2)\, dt_1\, dt_2. \tag{5.7}$$

Then, provided that the joint distribution of the errors is identified, we can identify the interaction term. Making the inverse Fourier transform and taking the expectation of the obtained expression over the covariates we can obtain

$$\alpha_0 y \;\; = \;\; -\frac{1}{(2\pi)^2}E[x_1\int e^{it_1x_1 + i(x - \varphi^{-1}(y))t_2}\frac{\Gamma(t_1)\Gamma(t_2)\chi(t_1, t_2)}{\mu_{1,y}(t_1, t_2)}dt_1 dt_2]. \tag{5.8}$$

Then we can choose a weighting function such that $\int \omega(y)y dy \neq$ and combine the information for different values of y, leading to

$$\alpha_0 \;\; = \;\; -\frac{1}{(2\pi)^2}E\left[x_1\int e^{it_1x_1 + ixt_2}\Gamma(t_1)\Gamma(t_2)\chi(t_1, t_2)\left\{\int \frac{e^{-i\varphi^{-1}(y)t_2}}{y\mu_{1,y}(t_1, t_2)}dy\right\}dt_1 dt_2\right]. \tag{5.9}$$

Note that the inverse weighting problem will be alleviated if the integral $\int \frac{e^{-i\varphi^{-1}(y)t_2}}{y\mu_{1,y}(t_1,t_2)}dy$ converges to a finite function of t_1 and t_2. This can occur even though $\mu_{1,y}(t_1, t_2)$ can vanish at some parts of its support.

6 A Nontriangular System

Here consider information of parameters of interest in a simultaneous discrete system of equations. In this section we no longer impose the triangular structure os the previous section. A leading example of this type os system is a 2-player discrete game with complete information. See, e.g. Tamer (2003)

We characterize this system as follows:

$$y_1 = 1[z'\beta_0 + \alpha_0 y_2 + \epsilon > 0]. \tag{6.1}$$

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

$$y_2 = 1[z'\delta_0 + \alpha_1 y_1 + \eta > 0], \tag{6.2}$$

where $z \equiv (z_1, z_2)$ is the vector of "instruments" and η is an error disturbance. Of particular interest is the identification and information of the parameter α_0, α_1 assuming that the errors η, ϵ are correlated with each other with unknown distribution. In particular, we are interested in determining under which conditions those two parameters can be estimated at the parametric rate, and in situations where they cannot be, which functions of the parameters can be.

6.1 Identification in systems with endogenous dummy variables

As in the previous section we use the notation $x_1 = z_1'\beta_0$ and $x = z'\delta_0$. We consider system of binary outcome variables

$$y_1 = \mathbf{1}\left\{x_1 + \alpha_{02} y_2 - \epsilon \geq 0\right\}, \tag{6.3}$$

$$y_2 = \mathbf{1}\left\{x + \alpha_{01} y_1 - \eta \geq 0\right\}. \tag{6.4}$$

Unlike the triangular system, this model displays a simultaneous effects of both dummy variables on each other. We make the same assumption regarding the errors as in the triangular system. The likelihood function of the model incorporates the probabilities of occurrence of outcomes y_1 and y_2 with the general structure similar to that in the triangular system. The conditional likelihood can be fully characterized by three conditional probabilities

$$\mathcal{P}\ \ (0,1 \mid x_1, x) = \int \mathbf{1}\left\{\epsilon > x_1 + \alpha_{02}\right\} \mathbf{1}\left\{\eta \leq x\right\} f_{\epsilon\eta} d\epsilon d\eta, \tag{6.5}$$

$$\mathcal{P}\ \ (1,0 \mid x_1, x) = \int \mathbf{1}\left\{\epsilon \leq x_1\right\} \mathbf{1}\left\{\eta > x + \alpha_{01}\right\} f_{\epsilon\eta} d\epsilon d\eta, \tag{6.6}$$

$$\mathcal{P}\ \ (0,0 \mid x_1, x) = \int \mathbf{1}\left\{\epsilon > x_1\right\} \mathbf{1}\left\{\eta > x\right\} f_{\epsilon\eta} d\epsilon d\eta. \tag{6.7}$$

$$\tag{6.8}$$

Note that the unknown interaction coefficients α_{01} and α_{02} are not present in the last expression. Therefore, one can use the same Fourier transformation technique that we used for the case of the triangular system to analyze this case. We consider the system of Fourier images of the conditional probabilities when we integrate over single indices x_1 and x. Then for the first probability we can write

$$\begin{aligned}
\mu_{01}(t_1,t_2) &= \int \mathcal{P}(1,1 \mid x_1,x) e^{-it_1x_1-it_2x} dx_1\, dx && (6.9)\\
&= \int\int e^{-it_1x_1-it_2x} \mathbf{1}\{\epsilon > x_1+\alpha_{02}\}\, \mathbf{1}\{\eta \le x\}\, f_{\epsilon\eta} d\epsilon d\eta dx_1 dx. && (6.10)
\end{aligned}$$

Making the change of variables $u = \epsilon - x_1 - \alpha_{02}$ and $v = \eta - x$ we notice that

$$\begin{aligned}
&\int\int e^{-it_1x_1-it_2x} \mathbf{1}\{\epsilon > x_1+\alpha_{02}\}\, \mathbf{1}\{\eta \le x\}\, f_{\epsilon\eta} d\epsilon d\eta dx_1 dx && (6.11)\\
= \; & e^{it_1\alpha_{02}} \int \mathbf{1}\{u \le 0\}\, e^{iut_1} du \int \mathbf{1}\{v \le 0\}\, e^{ivt_2} dv \int e^{-it_1\epsilon - it_2\eta} f_{\epsilon\eta} d\epsilon d\eta && (6.12)
\end{aligned}$$

We keep using notation $\chi(\cdot)$ for the characteristic function corresponding to the unknown distribution $f_{\epsilon\eta}$, $\Gamma(t) = \pi\delta(-t) + \frac{1}{it}$, and $\bar{\Gamma}(t) = \pi\delta(-t) - \frac{1}{it}$. Then Fourier transformations of three choice probabilities can be expressed as

$$\begin{aligned}
\mu_{01}(t_1,t_2) &= e^{it_1\alpha_{02}} \bar{\Gamma}(t_1)\Gamma(t_2)\chi(t_1,t_2), && (6.13)\\
\mu_{10}(t_1,t_2) &= e^{it_2\alpha_{01}} \Gamma(t_1)\bar{\Gamma}(t_2)\chi(t_1,t_2), && (6.14)\\
\mu_{00}(t_1,t_2) &= \bar{\Gamma}(t_1)\bar{\Gamma}(t_2)\chi(t_1,t_2), && (6.15)\\
& && (6.16)
\end{aligned}$$

In this case only the last equation does not contain the unknown interaction parameters. As a result, we can identify the density of the unobserved shocks from the last equation. To do that we multiply both sides of the last equation of (3.13) by $-t_1t_2$ and make an inverse Fourier transform. This leads to

$$\begin{aligned}
&-\frac{1}{(2\pi)^2} \int e^{it_1\epsilon + it_2\eta} t_1t_2\mu_{00}(t_1,t_2) dt_1 dt_2 && (6.17)\\
= \; & \frac{1}{(2\pi)^2} \int e^{it_1\epsilon + it_2\eta} (1 - i\pi t_1\delta(-t_1))(1 - i\pi t_2\delta(-t_2))\chi(t_1,t_2) dt_1 dt_2 && (6.18)\\
= \; & \frac{1}{(2\pi)^2} \int e^{it_1\epsilon + it_2\eta} \chi(t_1,t_2) dt_1 dt_2 = f_{\epsilon\eta}(\epsilon,\eta), && (6.19)
\end{aligned}$$

which is the same expression as in the case of the triangular system. This result is intuitive because the triangular binary system is a particular case of the simultaneous system.

Then the remaining two equations allow us to identify the interaction coefficients. In particular using the first equation we can write

$$e^{-\alpha_{02} i t_1} = \frac{\bar{\Gamma}(t_1)\Gamma(t_2)\chi(t_1,t_2)}{\mu_{01}(t_1,t_2)}, \tag{6.20}$$

and from the second equation we can find that

$$e^{-\alpha_{01} i t_2} = \frac{\Gamma(t_1)\bar{\Gamma}(t_2)\chi(t_1,t_2)}{\mu_{10}(t_1,t_2)}. \tag{6.21}$$

Thus, we can extract these coefficients directly as

$$\alpha_{02} = -t_1^{-1}\,\mathrm{Arg}\Big(\frac{\bar{\Gamma}(t_1)\Gamma(t_2)\chi(t_1,t_2)}{\mu_{01}(t_1,t_2)}\Big), \tag{6.22}$$

and

$$\alpha_{01} = -t_2^{-1}\,\mathrm{Arg}\Big(\frac{\Gamma(t_1)\bar{\Gamma}(t_2)\chi(t_1,t_2)}{\mu_{10}(t_1,t_2)}\Big). \tag{6.23}$$

We can also express these coefficients in terms of expectations of observable indices. This leads to expressions

$$\alpha_{02} = -\frac{1}{(2\pi)^2} E\left[x_1 \int e^{it_1x_1+it_2x}\frac{\bar{\Gamma}(t_1)\Gamma(t_2)\chi(t_1,t_2)}{\mu_{01}(t_1,t_2)}\,dt_1dt_2\right], \tag{6.24}$$

$$\alpha_{01} = -\frac{1}{(2\pi)^2} E\left[x \int e^{it_1x_1+it_2x}\frac{\Gamma(t_1)\bar{\Gamma}(t_2)\chi(t_1,t_2)}{\mu_{10}(t_1,t_2)}\,dt_1dt_2\right]. \tag{6.25}$$

Note that unlike the triangular system the simultaneous system does not contain "overidentifying" equations. There is one single equation for each structural parameter (interaction parameters and the unknown density).

7 Conclusions

This paper considers the *regularity* of the parameters of interest in simultaneous systems of nonlinear equations with dummy endogenous variables. It is found that regularity is directly related to the number of values the discrete endogenous variable can take. Consequently, attaining the parametric rate of convergence is impossible for many parameters of interest.

An important example of this case is the binary choice model with a single dummy endogenous variable, where we show the impossibility of attaining any estimation procedure which can converge at the root-n rate. Another important example is when the endogenous variable can take an infinite number of values in the same binary simultaneous system. As we show, regular identification can be achieved here, implying the possibility of developing an estimation procedure that converges at the parametric rate.

The work here suggests areas for future research. One is to establish methods of inference for the parameters of interest when they cannot be estimated at the standard parametric rate. Related to that issue is to derive optimal rates of convergence for those same parameters, and then hopefully, derive an estimation procedure which will be efficient in the sense that it can converge at the optimal rate.

References

Abrevaya, J., J. Hausman, and S. Khan (2010): "Testing for Causal Effects in a Generalized Regression Model with Endogenous Regressors," *Econometrica*, forthcoming.

Aradillas-Lopez, A. (2005): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information," Working Paper, Princeton University.

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.

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

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

Hong, H., and E. Tamer (2003): "Inference in Censored Models with Endogenous Regressors," *Econometrica*, 71(3), 905–932.

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

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

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

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

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

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.