Review of Economic Studies (2003) **70**, 147–165 0034-6527/03/00060147$02.00
© 2003 The Review of Economic Studies Limited

Incomplete Simultaneous Discrete Response Model with Multiple Equilibria

ELIE TAMER
Princeton University

First version received November 1999; *final version accepted August* 2002 (*Eds.*)

A bivariate simultaneous discrete response model which is a stochastic representation of equilibria in a two-person discrete game is studied. The presence of multiple equilibria in the underlying discrete game maps into a region for the exogenous variables where the model predicts a nonunique outcome. This is an example of an incomplete econometric structure. Economists using this model have made simplifying assumptions to avoid multiplicity. I make a distinction between incoherent models and incomplete models, and then analyse the model in the presence of multiple equilibria, showing that the model contains enough information to identify the parameters of interest and to obtain a well defined semiparametric estimator. I also show that the latter is consistent and $\sqrt{n}$ normal. Moreover, by exploiting the presence of multiplicity, one is able to obtain a more efficient estimator than the existing methods.

1. INTRODUCTION

This paper studies a bivariate simultaneous response model which is a stochastic representation of equilibria in a two-person discrete game. Multiplicity of equilibria in the underlying game maps into regions for the exogenous variables where the econometric model predicts more than one outcome. This is an incomplete econometric model.[1] This class of econometric models provides inequality restrictions on moment conditions that are functions of the underlying parameters of interest. To avoid multiplicity, economists studying these models have made simplifying assumptions which either change the outcome space or impose *ad hoc* selection mechanisms in regions of multiplicity. I study this model in the presence of multiplicity, showing that the parameters of interest can be point identified. Using restrictions on the probability of the nonunique outcomes, I also develop a maximum likelihood estimator demonstrating its parameter consistency and $\sqrt{n}$ normality. More importantly, I show that one can exploit multiplicity to garner efficiency gains.

I study the following parametrization of a bivariate discrete game

TABLE 1

Discrete game with stochastic payoff

	$y_2 = 0$	$y_2 = 1$
$y_1 = 0$	$0, 0$	$0, x_2\beta_2 - u_2$
$y_1 = 1$	$x_1\beta_1 - u_1, 0$	$x_1\beta_1 + \Delta_1 - u_1, x_2\beta_2 + \Delta_2 - u_2$

1. For a definition of an incomplete response model, see Section 2 below.

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

which maps directly into the econometric model

$$\begin{aligned} y_1^* &= x_1\beta_1 + y_2\Delta_1 + u_1 \\ y_2^* &= x_2\beta_2 + y_1\Delta_2 + u_2 \end{aligned} \tag{1}$$

$$y_j = \begin{cases} 1 & \text{if } y_j^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{for } j = 1, 2$$

where $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^d$ represents a vector of observed exogenous variables, $\mathbf{u} = (u_1, u_2)$ is a random vector of latent variables with conditional distribution F_u that represents unobserved (to the econometrician) profits, and $\beta = (\beta_1, \beta_2, \Delta_1, \Delta_2)$ are parameters of interest.

This is a well-known econometric model that has been studied widely in econometrics. For example, it is the same model used in Bresnahan and Reiss (1991) and considered in important work by Heckman (1978) on models with structural shift parameters, Blundell and Smith (1993, 1994) on female labour supply and Schmidt (1981) on simultaneity in bivariate econometric models. For a more extensive list of references to model (1) above, see Chapter 5 of Maddala (1983). One can assume that the incremental utility for a player playing one when the other player moves from playing a zero to a one (the Δ's) depends on observable characteristics. For simplicity, the parametrization I use treats the Δ's as fixed[2] parameters to be estimated.[3] Given the response functions in (1) above, we have

$$\begin{aligned} \Pr[(0,0)|x] &= \Pr(u_1 < -x_1\beta_1;\, u_2 < -x_2\beta_2) \\ \Pr[(1,1)|x] &= \Pr(u_1 \geq -x_1\beta_1 - \Delta_1;\, u_2 \geq -x_2\beta_2 - \Delta_2) \\ \Pr[(0,1)|x] &= \Pr(u_1 < -x_1\beta_1 - \Delta_1;\, u_2 \geq -x_2\beta_2) \\ \Pr[(1,0)|x] &= \Pr(u_1 \geq -x_1\beta_1;\, u_2 < -x_2\beta_2 - \Delta_2). \end{aligned}$$

For example, assuming that the incremental payoffs Δ_1 and Δ_2 are negative, it is easy to see that

$$\Pr[(0,0)|x] + \Pr[(0,1)|x] + \Pr[(1,0)|x] + \Pr[(1,1)|x] > 1.$$

This is an example of an *incoherent* econometric model. Economists using these models have imposed the well-known "coherency" condition $\Delta_1^1 \times \Delta_2^1 = 0$ which is a necessary and sufficient condition for the probability of the four outcomes to sum to one (Heckman, 1978). Unfortunately, imposing this condition eliminates the simultaneity, an essential feature the model is trying to capture. Schmidt (1981) raises the question of whether "economists should even consider these type of systems" to model simultaneity because the coherency conditions render the models recursive. Tracing back the source of the "coherency" problem, I find that the econometric structure above is a particular case of a well-defined *incomplete* discrete econometric model. The above model arises as a result of a particular economic optimization problem involving two decision makers where the presence of multiple equilibria in the underlying game maps into a well-defined incomplete econometric structure which provides *upper and lower* probabilities for the $(0, 1)$ and $(1, 0)$ outcomes, and exact probabilities for the $(1, 1)$ and $(0, 0)$ outcomes (in the case where Δ_1, Δ_2 are negative, the other cases are similar). As in Figure 1, we have the

2. I treat the case where the Δ's are negative for simplicity. The case for when the Δ's are positive is similar. In Section 3, I also discuss the cases where the Δ's have different signs.

3. In Example 1 on page 151 I study the parametrization of a bivariate discrete game that results from a Cournot duopoly where the Δ's become functions of the fundamental parameters of the demand equation in the model. In that case, it would not make sense to talk about identifying and estimating the Δ's. One can focus on estimating the β's which could represent, for example, the impact of an exogenous change in firm profit on the likelihood of the firm entering a market (see Example 1 for more detail).

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

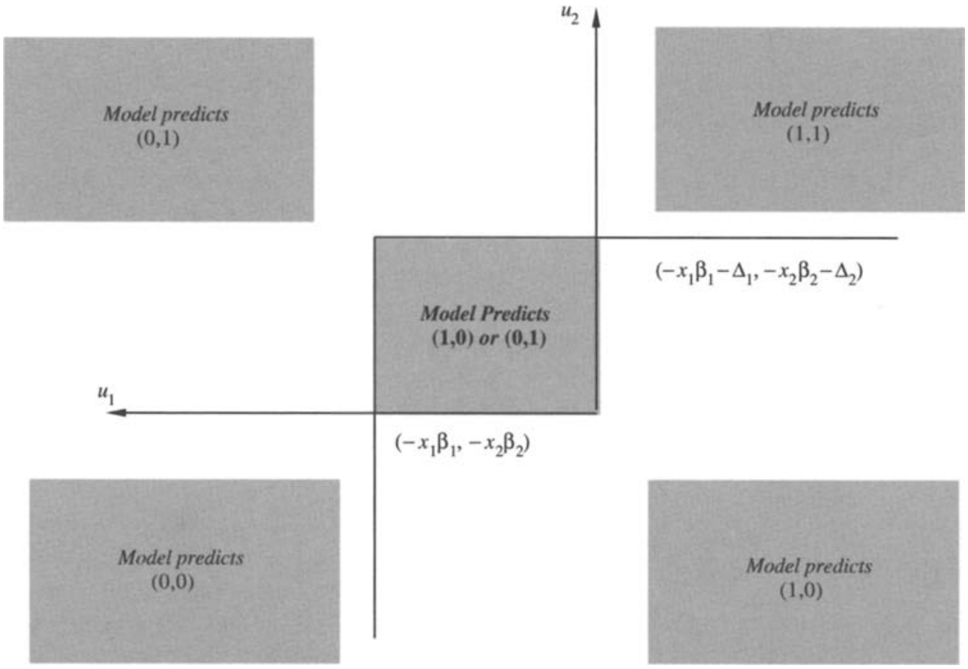

FIGURE 1

Incomplete model with multiple equilibria

following restrictions on the $(0, 1)$ and $(1, 0)$ outcomes:

$$P_3(x, \beta) \le \Pr[(0, 1)|x] \le P_4(x, \beta)$$

where

$$\begin{aligned} P_3(x, \beta) &= \Pr(u_1 < -x_1\beta_1 - \Delta_1; u_2 > -x_2\beta_2 - \Delta_2) \\ &\quad + \Pr(u_1 < -x_1\beta_1; -x_2\beta_2 < u_2 < -x_2\beta_2 - \Delta_2) \\ P_4(x, \beta) &= \Pr(u_1 < -x_1\beta_1 - \Delta_1; u_2 > -x_2\beta_2). \end{aligned} \tag{2}$$

Traditionally, economists using simultaneous discrete response models make simplifying assumptions to respond to the nonuniqueness problem without invoking the coherency condition. For example, Bjorn and Vuong (1985) and Kooreman (1994) assume a unique outcome is chosen with known probability λ in the region of incompleteness. This type of assumption changes the model in an *ad hoc* way that might lead to inconsistent estimates. Another practice is to treat the multiple outcomes as one event and hence changing the model into one that predicts the joint equilibria (Bresnahan and Reiss, 1990, 1991). Transforming the model into one that predicts the joint outcome $[(0, 1)$, *or*, $(1, 0)]$ clearly involves a loss of information, narrowing down the class of models that could be examined and limiting the ability of the researcher to predict (Bresnahan and Reiss, 1990). This paper contributes to the literature on inference in nonlinear simultaneous equations model and to a growing literature on inference in models with social interactions by studying a model with two decision makers. To that extent, the paper makes several contributions. First, I make a distinction in Section 2 below between incoherent and incomplete models by showing that an otherwise incoherent model contains information on the parameters of interest. This clarifies the misconception that incoherent

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

models are not useful for inference. Second, using a parametrization of the game above that corresponds to the well-known simultaneous equations model with discrete outcomes, I show in Section 3 how this model can be used to study simultaneity. In particular, I improve on the existing empirical games literature by showing in Section 4 that one can exploit the inequality restrictions in the model to garner *efficiency gains* using a maximum likelihood based estimator. This framework also allows for sharper prediction by providing tighter upper and lower probabilities on the outcomes (see page 159 for more details on this point). Fourth I show how parameters of interest can still be identified in cases where no existing methods can be used. The estimator introduced in Manski and Tamer (2002) can then be used to estimate the parameters.

2. COHERENCY AND IDENTIFICATION

In the econometrics literature, a model is known to be coherent if it admits "a well defined reduced form" (see Gourieroux, Lafont and Monfort, 1981). This is equivalent to the model predicting a unique value for the dependent variables (or endogenous variables) given values of the exogenous variables both observed and unobserved. Coherency conditions appear under the names of logical consistency conditions (Maddala, 1983), internal consistency conditions (Schmidt, 1981), principal conditions (Heckman (1978), Gourieroux *et al.* (1981)) among others and are conditions that guarantee a well-defined likelihood for the endogenous variables given the exogenous ones. On the other hand, in this paper, I make a distinction between *incoherent economic* models and *incomplete econometric* models. When faced with an "incoherent" model, one often needs to *trace back the source* of the incoherency. One type of incoherency for example originates in the underlying economic model where the agent's optimal decision rule is not well-defined, or is "logically inconsistent". In particular, coherency rules in this case represent conditions for a well-defined optimization problem where the existence of an equilibrium is guaranteed. In demand systems with rationing, concavity of the expenditure function maps directly into the coherency conditions which are similar in effect to the Slutzky conditions (see for example, Wales and Woodland (1983), Van Soest and Kooreman (1990)). When mapped into an econometric framework, the economic coherency conditions often represent *sufficient* conditions for the statistical model to have a well-defined likelihood. These are the coherency conditions. On the other hand, as in Manski (1988), a "complete" econometric response model "asserts that a random variable y is a function of a random pair (x, u) where x is observable and u is not". An "incomplete" econometric model is one where the relationship from (x, u) to y is a correspondence and not a function. For example, a selection problem in an underlying well-defined economic model, multiple equilibria, or censoring of regressors or outcomes map into an econometric structure with a non-unique predicted outcome distribution. This incompleteness is often the result of the unwillingness to impose strong (and sometimes untestable) assumptions. Moreover, incomplete models can often be written in terms of inequality restrictions on regressions.

With minimal assumptions, incomplete models in general contain information about the parameters of interest. Heuristically, in a model defined in terms of inequality restrictions on regressions, the identified features of the model is the set of parameter values that satisfy the inequality restrictions. It is possible that this set of inequality restrictions is satisfied uniquely by the true parameter of interest. In this case we say that the parameter is point identified. Hence, incompleteness and identification are logically distinct features of an econometric model: the former is a property of the model that usually predicts a nonunique conditional distribution for the outcome, while the latter asks what features of the model can be consistently estimated.

Next, I highlight the ideas of incompleteness and identification in two examples where the incompleteness is a result of the multiple equilibria in the underlying game. I then provide conditions for the model in (1) above to be point identified. For other examples of models where

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

TABLE 2

Payoffs in a Cournot model

	No entry	Enter
No entry	$0, 0$	$0, \frac{a^2}{4b} + x_2\beta_2 - u_2$
Enter	$\frac{a^2}{4b} + x_1\beta_1 - u_1, 0$	$\frac{a^2}{9b} + x_1\beta_1 - u_1, \frac{a^2}{9b} + x_2\beta_2 - u_2$

the incompleteness is a result of interval observations on a regressor or outcome (see Manski and Tamer, 2002).

2.1. *Example* 1

In this example, I consider a simple firm entry problem studied by Bresnahan and Reiss (1990, 1991). In a Cournot model,[4] consider the linear demand $p = a - bq$ while firms 1 and 2 have stochastic payoffs $\pi_i = pq_i + x_i\beta_i - u_i$ for $i = 1, 2$, where x_i represent exogenous determinants of firm profits that are observed and u_i represent exogenous mean zero unobserved (by the econometrician) random variables. Moreover assume that firms earn zero profits for producing no output and that the joint distribution of (u_1, u_2) is known up to a finite-dimensional parameter. Firm payoffs are given in Table 2. One important purpose of the exercise is to predict the effect of an exogenous increase/decrease in x_1 or x_2 on the likelihood for firm 1 or 2 to exit/enter the market given that the response functions in the payoff table above are invariant to the exogenous policy change in the x's. To estimate this effect, one needs an estimate of the fundamental parameters β_1 and β_2. The above table maps directly into the parametrization in (1) above with $\Delta_1 = \Delta_2 = -\frac{5a^2}{36b}$. Given downward sloping demand ($b > 0$), the game admits (*Enter*, *No entry*) and (*No entry*, *Enter*) as the multiple equilibria of the game if the u's have rich enough support. By looking at markets with no entrants and at duopoly markets, it is easy to see that the model above point identifies $(\frac{a^2}{b}, \beta_1, \beta_2)$ given enough variation in the x's (more on identification in Section 3 below). Even though one can estimate the identified features (including the β's) of the above model using maximum likelihood[5] by looking at data from duopoly or no entrant markets, the framework presented in this paper is able to exploit the presence of multiple equilibria to estimate (β_1, β_2) more efficiently and hence obtain sharper estimates for the given exogenous impacts.[6] Moreover, we can obtain sharper upper and lower probabilities for the events[7] Pr(*Firm* 1 *is a monopolist*$|x_1, x_2$) and Pr(*Firm* 2 *is a duopolist*$|x_1, x_2$).

2.2. *Example* 2

This case draws on an example by Jovanovic (1989). Let the payoff function of two players be defined in the following way:

$$\Pi_1(y_1, y_2, u_1) = \begin{cases} \theta_1 y_2 - u_1 & \text{if } y_1 = 1 \\ 0 & \text{if } y_1 = 0 \end{cases}$$

4. I am very grateful to a reviewer whose comments clarified important features of this example which helped improve the paper.

5. The parameters a^2 and b are not separately point identified in this model; however, we know that $b > 0$ and that $a^2 = K \times b$ where K is a constant that can be consistently estimated.

6. One can add to the above game observed exogenous determinants of market demand z (similar to the x's). This essentially makes the parameter a a function of the observed variables z. In this case, we can still point identify (up to scale) coefficients that represent exogenous determinants of firm profit that are excluded from the z's.

7. This is in contrast to the existing methods in estimating entry models where one is only able to obtain the probability that one firm is serving a given market which would imply the trivial bound on the outcome of interest.

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

$$\Pi_2(y_1, y_2, u_2) = \begin{cases} \theta_2 y_1 - u_2 & \text{if } y_2 = 1 \\ 0 & \text{if } y_2 = 0. \end{cases}$$

This is the payoff function for a two-player game where $y_i \in \{0, 1\}$ is player i's action set, and $u_i \sim U(0, 1)$ is the player's type. I assume that players' types are common knowledge between the two players, but the econometrician knows only that they are uniformly distributed. One can think of the above as a technology adoption game for complementary products. Firm 1 would provide technology 1 (software) only if firm 2 provides technology 2 (hardware) where u_i represents investment by firm i. To map this into our framework, notice that the only two Nash-equilibria are $(1, 1)$ and $(0, 0)$ (either hardware and software are provided or none is):

$$(y_1, y_2)_{NE} = \begin{cases} (1, 1) & \text{if } u_1 < \theta_1 \text{ and } u_2 < \theta_2 \\ (0, 0) & \text{all } u_i \in [0, 1] \end{cases}$$

where $(\theta_1, \theta_2) \in \mathbf{\Theta} \subset R^{2+}$. The players' strategies ($y_i \in \{0, 1\}, i = 1, 2$) are perfectly correlated and are denoted by $y = y_1 = y_2$. This means that the only possible equilibria of the game are that both firms provide the complementary goods or neither does. Given the parameter space Θ and the distribution F_U of $U = (u_1, u_2)$, the stochastic game admits multiple equilibria. The incompleteness in this model is a direct result of the presence of multiple equilibria in the game. For example, if $(1, 1) \leq (\theta_1, \theta_2)$, the model cannot predict a unique outcome for any value of $(u_1, u_2) \in (0, 1) \times (0, 1)$. This implies that

$$0 \leq \Pr(y = 1) \leq 1.$$

As a result, for $(\theta_1, \theta_2) \geq (1, 1)$, the model will not provide any information about the parameters (θ_1, θ_2).

If on the other hand $\theta_1 = \theta_2 = \theta < 1$, then for $(u_1, u_2) \in (0, \theta) \times (0, \theta)$, the model predicts $y = 1$ or $y = 0$. This implies that

$$0 \leq \Pr(y = 1) \leq \theta^2.$$

Using this restriction, we are not able to point identify this model since the inequality above holds for all θ' such that $\theta^2 \leq \theta'^2 \leq 1$. However, if I make a random sample assumption I could use the empirical analogue $\widehat{\Pr}(y = 1)$ of $\Pr(y = 1)$ to get that

$$\widehat{\Pr}(y = 1) \leq \theta^2 \leq 1.$$

The above is a nontrivial bound.

3. BIVARIATE DISCRETE RESPONSE MODEL: IDENTIFICATION

In this section I shall examine identification of the bivariate simultaneous model defined in (1) above. To recap, the model in which we are interested is

$$\begin{aligned} y_1^* &= x_1\beta_1 + y_2\Delta_1 + u_1 \\ y_2^* &= x_2\beta_2 + y_1\Delta_2 + u_2 \end{aligned} \tag{3}$$

$$y_j = \begin{cases} 1 & \text{if } y_j^* > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{for } j = 1, 2.$$

The set of endogenous variables Y consists of $Y = \{(0, 0), (1, 1), (1, 0), (0, 1)\}$.

It is easy to see that if the u's in (3) have large enough support, the theoretical game in Table 1 above always admits multiple equilibria. For example, if the Δ's are both negative, and for $-x_i\beta_i \leq u_i \leq -x_i\beta_i - \Delta_i$ $(i = 1, 2)$ the econometric model predicts either $(0, 1)$ or $(1, 0)$ which are the multiple equilibria of the underlying game.

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

In the case above where the Δ's are negative the model provides the following *inequality restrictions* on conditional regressions:

$$P_1(x,\beta) = \Pr[(0,0)|x] = \Pr(u_1 < -x_1\beta_1; u_2 < -x_2\beta_2)$$
$$P_2(x,\beta) = \Pr[(1,1)|x] = \Pr(u_1 \geq -x_1\beta_1 - \Delta_1; u_2 \geq -x_2\beta_2 - \Delta_2)$$
$$P_3(x,\beta) \leq \Pr[(0,1)|x] \leq P_4(x,\beta) \quad (4)$$

where P_3 and P_4 are defined in (2) above. The objective now becomes what identified features of the econometric model above defined in terms of inequality (and equality) restrictions can be consistently estimated.

3.1. *Identification when* $\Delta_1 \times \Delta_2 > 0$

In this section, I study the case where the shift parameters are negative. The case when the shift parameters are both positive is symmetric. I begin with three assumptions. The first is the sampling assumption, the second assumption requires that the joint distribution of the unobserved terms is known up to a finite-dimensional parameter, and the third assumption requires that the shift parameters are negative.

Assumption 1. *We have an iid sample* $\{(y_{1i}, y_{2i}), x_{1i}, x_{2i}\}$ *such that* $0 < \Pr[(y_1, y_2)|(x_1, x_2)] < 1$ *for all* $(y, x_1, x_2) \in Y \times \mathrm{R}^{d_1} \times \mathrm{R}^{d_2}$ *where* $x = (x_1, x_2) \in \mathrm{R}^d$ *and* $Y = \{(0,0), (1,1), (0,1), (0,1)\}$.

Assumption 2. *Let* $U = (u_1, u_2)$ *be a random vector independent of* x *with a known joint conditional distribution* F_u *that is absolutely continuous with mean* 0 *and unknown covariance matrix* Ω.

Assumption 3. Δ_1 *and* Δ_2 *are negative.*

This is an important assumption that will determine which of the four outcomes are observationally equivalent in a region of the exogenous variables. If Δ_1 and Δ_2 are both negative, the model provides exact probabilities for the $(0,0)$ and $(1,1)$ outcomes and upper/lower probabilities for the $(1,0)$ and $(0,1)$ outcomes. On the other hand if Δ_1 and Δ_2 are both positive, the model provides exact probabilities for the $(0,1)$ and $(1,0)$ outcomes and upper/lower probabilities for the $(1,1)$ and $(0,0)$ outcomes. For the cases where $\Delta_1 \times \Delta_2 < 0$ the model provides upper/lower probabilities for every outcome. This case is examined in the next section. From the model above at the true parameter value β, one obtains the restrictions on the conditional distribution of the outcomes given the observables described in (4) above. This is an incomplete discrete model. The next theorem shows that this incompleteness will not present any problems for identification. Since this is a threshold crossing model, we normalize the variances in Ω of u_1 and u_2 to be one. We focus on the identification of the parameter β in the next theorem that presents *sufficient* point identification conditions.

Theorem 1. *Let Assumptions* 1–3 *hold. Moreover, assume that for* $i = 1$ *or* $i = 2$, *there exists a regressor* x_{ik} *with* $\beta_{ik} \neq 0$ *such that* $x_{ik} \notin x_{3-i}$ *and such that the distribution of* $x_{ik} \mid x_{-ik}$ *has an everywhere positive Lebesgue density where* $x_{-ik} = (x_{i1}, \ldots, x_{ik-1}, x_{ik+1}, \ldots, x_{id_i})$. *Then the parameter vector* $\beta = (\beta_1, \beta_2, \Delta_1, \Delta_2)$ *is identified if the matrices* x_1 *and* x_2 *have full column rank.*

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

Proof of Theorem 1. Without loss of generality let $x_{ik} = x_{1k}$ and $\beta_{1k} > 0$. Let (b_1, b_2) be such that $(b_1, b_2) \neq (\beta_1, \beta_2)$. Let $b_{1k} > 0$. By the support condition on x_{1k}, as x_{1k} increases to minus infinity given x_{-1k}, we get that both $x_1\beta_1$ and $x_1 b_1$ go to minus infinity. Let x_2^* be such that

$$x_2^*\beta_2 \neq x_2^* b_2.$$

The existence of such an x_2^* is guaranteed by the full rank condition on x_2. Hence by Assumption 2

$$\Pr\left[(0,0)|x_1, x_2^*\right] = \Pr\left[u_1 \leq -x_1\beta_1, u_2 \leq -x_2^*\beta_2\right] \neq \Pr\left[u_1 \leq -x_1 b_1, u_2 \leq -x_2^* b_2\right]$$

since

$$\begin{aligned}\Pr\left[u_1 \leq -x_1\beta_1, u_2 \leq -x_2^*\beta_2\right] &\simeq \Pr\left[u_2 \leq -x_2^*\beta_2\right] \neq \Pr\left[u_2 \leq -x_2^* b_2\right] \\ &\simeq \Pr\left[u_1 \leq -x_1 b_1, u_2 \leq -x_2^* b_2\right]\end{aligned}$$

for $-x_1\beta_1, -x_1 b_1 \gg 0$. This implies that β_2 is identified. As for β_1, let x_1 be such that $x_1\beta_1 \neq x_1 b_1$. The existence of such an x_1 is guaranteed by the full rank condition on x_1. Hence we also have

$$\Pr[(0,0)|x_1, x_2] = \Pr[u_1 \leq -x_1\beta_1, u_2 \leq -x_2\beta_2] \neq \Pr[u_1 \leq -x_1 b_1, u_2 \leq -x_2\beta_2].$$

As for the case when $b_{1k} < 0$ (we still have $\beta_{1k} > 0$), it is easy to see that given x_{-1k}, as x_{1k} increases, $x_1\beta_1$ increases while $x_1 b_1$ decreases which implies that b_{1k} is identified relative to β_{1k} using arguments similar to the above. This implies that β_1 and β_2 are identified. In particular the constant terms are also identified. Repeating the above analysis while replacing (0, 0) with (1, 1) and $x_i\beta_i$ with $x_i\beta_i + \Delta_i$ will yield identification of the constant term which implies that Δ_i is identified. ||

Comments on Theorem 1:

- I require that a continuous regressor be included in either x_1 or x_2. This is a sufficient condition for point identification. All one needs is the support of x to be rich enough to identify the parameters.
- In Assumption 2, I require that the joint distribution F_u be known. This is not needed for identification.[8] Moreover, it is possible to weaken the independence assumption between the error vector u and the x at the expense of strong assumptions on the conditional distribution of u given x.
- The parameter vector β is identified up to scale given the threshold crossing nature of the model. One can then assume that u_1 and u_2 have unit variances. As far as identification of the rest of the parameters present in the matrix Ω, those will depend on the specific bivariate distribution of (u_1, u_2). The identification strategy of these parameters is similar to that used for identifying β mainly that these parameters should be point identified with rich enough support on the x's. In addition to using the outcome probabilities, one can also use the conditional distribution of y_1 given y_2 (and vice versa) as a function of x and (β, Ω) to identify the parameters.
- Without imposing the rich support conditions above, the model still provides information about the parameters of interest. The identified feature of the model, which would be a set in general (this set would shrink to a point under the conditions in the theorem) can be estimated using the modified minimum distance estimator introduced

8. I thank a referee for this insight.

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

in Manski and Tamer (2002). This estimator was proved to consistently set-estimate the identified feature of an incomplete model that is based on inequality restrictions on regressions.

Moreover, the above model identifies upper and lower probabilities on the outcomes $\Pr[(0,1)|x_1,x_2]$ and $\Pr[(1,0)|x_1,x_2]$. Given that β can be consistently estimated and the incomplete model defined in (4) above, we have

$$P_3(x,\beta) \le \Pr[(0,1)|x_1,x_2] \le P_4(x,\beta).$$

The upper and lower probabilities on the $(1,0)$ outcome are similar. The bound on the conditional probabilities provided by the incomplete model are usually much tighter than the ones obtained from models that treat the $(0,1)$ and $(1,0)$ outcomes as one event. The bound provided by these models on the $(0,1)$ outcome is

$$0 \le \Pr[(0,1)|x] \le 1 - P_1(x,\beta) - P_2(x,\beta).$$

A note on mixed strategy equilibria. In the above, we have implicitly assumed away mixed strategy equilibria in the underlying game. For a payoff structure such that $-x_i\beta_i \le u_i \le -x_i\beta_i - \Delta_i$ $(i = 1,2)$, $(0,1)$ and $(1,0)$ are the multiple equilibria of the game. If we allow mixed strategies then there is a positive probability that the other two outcomes $(1,1)$ and $(0,0)$ will appear (part of a coordination failure for example by the two players). In this case[9] the restrictions provided by the model will be

$$\begin{aligned} P_1(x,\beta) &\le \Pr[(1,1) \mid x] \le P_1(x,\beta) + P_{\text{square}}(x,\beta) \\ P_2(x,\beta) &\le \Pr[(0,0) \mid x] \le P_2(x,\beta) + P_{\text{square}}(x,\beta) \\ P_3(x,\beta) &\le \Pr[(0,1)|x] \le P_4(x,\beta) \end{aligned}$$

where

$$P_{\text{square}}(x,\beta) = \Pr[-x_1\beta_1 \le u_1 \le -x_1\beta_1 - \Delta_1; -x_2\beta_2 \le u_2 \le -x_2\beta_2 - \Delta_2|x].$$

Identification in this model is similar to identification in the next section.

A note on identification when $\Delta_1 \times \Delta_2 < 0$. I consider the case when $\Delta_1 > 0$ and $\Delta_2 < 0$, the other case is symmetric.

Assumption 4. $\Delta_1 > 0$ *and* $\Delta_2 < 0$.

It is easy to see that for some values of the exogenous variables, either of the four outcomes is likely. This is illustrated in Figure 2.

In the underlying game, for example, there is no equilibrium in pure strategies in the region of multiplicity: each player is indifferent between choosing 1 or 0 given that the other is randomizing. This maps into the model having the following restrictions:

$$\begin{aligned} P_1(x,\beta) &\le \Pr[(1,1)|x] \le P_1(x,\beta) + P_{\text{square}}(x,\beta) \\ P_2(x,\beta) &\le \Pr[(0,0)|x] \le P_2(x,\beta) + P_{\text{square}}(x,\beta) \\ P_3'(x,\beta) &\le \Pr[(1,0)|x] \le P_3'(x,\beta) + P_{\text{square}}(x,\beta) \end{aligned} \tag{5}$$

9. There are other mixed strategy equilibria that can appear in the set of games above. Those equilibria depend on a particular value for (u_1,u_2) and hence have probability zero given Assumption 2.

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

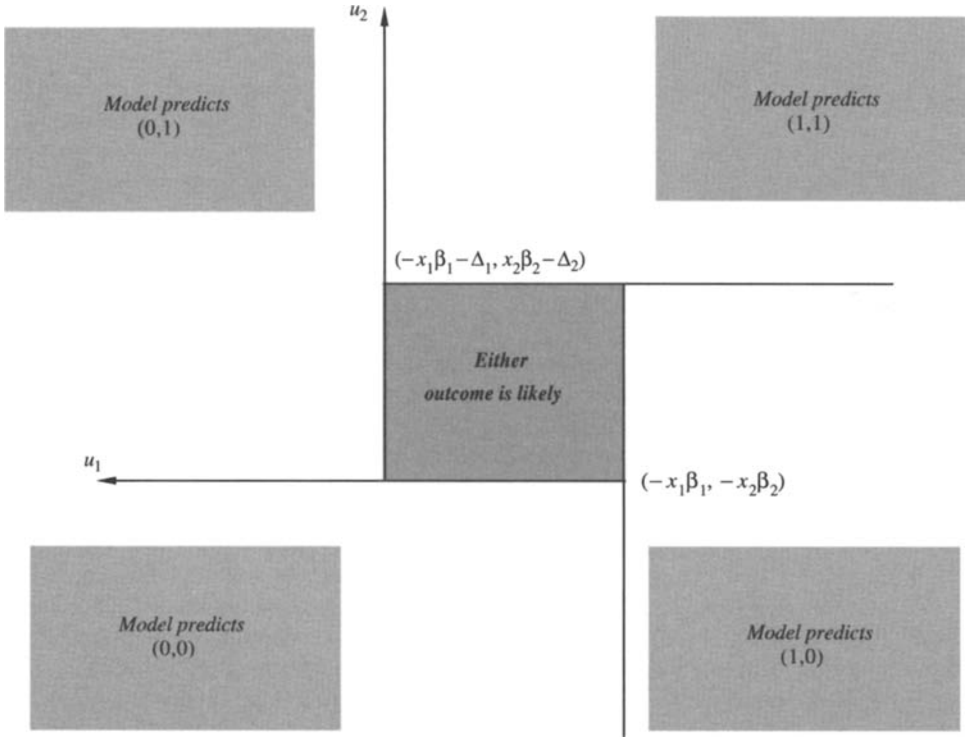

FIGURE 2

Incomplete bivariate model for the case where $\Delta_1 > 0$, $\Delta_2 < 0$

where P_1 and P_2 are the same as above and

$$\begin{aligned} P_3'(x,\beta) &= \Pr[u_1 \geq -x_1\beta_1; u_2 < -x_2\beta_2 - \Delta_2] \\ P_{\text{square}}(x,\beta) &= \Pr[-x_1\beta_1 - \Delta_1 \leq u_1 \leq -x_1\beta_1; -x_2\beta_2 \leq u_2 \leq -x_2\beta_2 - \Delta_2]. \end{aligned} \tag{6}$$

The restrictions provided by the model are in terms of upper and lower probabilities. The next theorem provides sufficient conditions under which the true parameter vector is identified. I require the existence of one continuous regressor with an everywhere positive Lebesgue density (continuous regressor on $\mathbb{R}$). Heuristically what a continuous regressor allows us to do is by looking at the conditional probability of the (0, 0) outcome and for any $b \neq \beta$ we will be able to find a set with positive probability X' such that for all $x' \in X'$ the lower bound $P_2(x',\beta)$ at β is *greater* than the upper bound $P_2(x',b) + P_{\text{square}}(x',b)$ at b.

Theorem 2. *Given Assumptions* 2 *and* 4 *above, and assume that there exists at least one regressor* x_{ik} *where* $k \in \{1,\dots,d_i\}$ $\beta_{ik} \neq 0$, *for* $i = 1, 2$, *such that* $x_{1k} \neq x_{2k}$ *whose conditional distribution given* x_{-ik} *has everywhere positive Lebesgue density where* $x_{-ik} = (x_{i1},\dots,x_{ik-1},x_{ik+1},\dots,x_{id_i})$, *then the parameter vector* β *is identified.*

As we can see here, in the case where the model only delivers upper/lower probabilities for *every* outcome the parameters of interest are still identified.

One way to estimate the identified features of the model defined in terms of the inequality restrictions in (5) above is to use a modified minimum distance estimator similar to the one used

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

in Manski and Tamer (2002). Heuristically, for a given parameter in the parameter space, define a loss function that is positive only if the inequalities in (5) are not satisfied and zero otherwise; this can be made operational if we replace every conditional probability by its empirical counterpart. If the true parameter satisfies the inequalities uniquely for all values of the regressors then with support restrictions on the regressors one can consistently estimate the parameter of interest. For more details about this estimation strategy refer to Manski and Tamer (2002).

4. BIVARIATE DISCRETE RESPONSE MODEL: EFFICIENCY GAINS WITH MULTIPLE EQUILIBRIA

In this section I formulate an estimator that exploits the presence of multiple equilibria in the case where the shift parameters have the same sign. In particular, I study the case where Δ_1 and Δ_2 are negative (the other case is symmetric). More importantly, I show that there are efficiency gains in exploiting multiplicity. In particular, as compared with the limited maximum likelihood (ML) estimator that treats the outcomes (1, 0) and (0, 1) as one event, the semiparametric maximum likelihood estimator (SML) introduced below is shown to be more efficient.

The ML estimator ($\Delta_1 < 0$ and $\Delta_2 < 0$). I define here a maximum likelihood estimator that can be used to consistently estimate the parameter β. This inference strategy in models with multiple equilibria whereby the likelihood is written in terms of features of the model that are uniquely predicted is used in Bresnahan and Reiss (1990, 1991) and Berry (1992). Consider the model where we have three outcomes (0, 0), (1, 1) and (1, 0) or (0, 1). Given the outcome probabilities in (4), the log-likelihood is

$$L_{\mathrm{ML}}(b) = \sum_{i=1}^{n} \big[y_{i1}y_{i2}\log(P_1(x_i, b)) + (1 - y_{i1})(1 - y_{i2})\log(P_2(x_i, b))$$
$$((1 - y_{i1})y_{i2} + y_{i1}(1 - y_{i2}))\log(1 - P_1(x_i, b) - P_2(x_i, b))\big] \tag{7}$$

where the functions P_1 and P_2 are defined in (4) above. Using the usual MLE techniques the covariance matrix of the above-modified likelihood is

$$\Omega_{\mathrm{ML}} = E\left[\frac{\partial P_1 \partial P_1'}{P_1} + \frac{\partial P_2 \partial P_2'}{P_2} + \frac{(\partial P_1 + \partial P_2)(\partial P_1 + \partial P_2)'}{1 - P_1 - P_2}\right]^{-1}$$

where ∂P_1 and ∂P_2 are the derivative vectors of the functions $P_1(x, b)$ and $P_2(x, b)$ respectively, with respect to b evaluated at the true parameter value β.

The SML estimator ($\Delta_1 < 0$ and $\Delta_2 < 0$). To define this estimator, I replace the predicted probability of the (0, 1) outcome with its empirical counterpart. I have thus replaced the part of the model that is incomplete by a function which represents the population frequency of the (0, 1) event. Define the function $H(\cdot)$ as follows:

$$H(x) = \Pr[(0, 1)|x].$$

In the case where the function $H(\cdot)$ is known, a way to estimate β given the above assumptions is to use the method of maximum likelihood. The logarithm of the likelihood function of the model will be

$$L(\beta; H) = \frac{1}{N}\sum_{i=1}^{n}[y_{i1}y_{i2}\log(P_1(x_i, \beta)) + (1 - y_{i1})(1 - y_{i2})\log(P_2(x_i, \beta))$$
$$+ (1 - y_{i1})y_{i2}\log(H(x_i)) + y_{i1}(1 - y_{i2})\log(1 - P_1(x_i, \beta) - P_2(x_i, \beta) - H(x_i))].$$

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

Since the function $H(\cdot)$ is unknown, I replace it by $\hat{H}_n(\cdot)$ a function that locally approximates it. The resulting log-likelihood function becomes a semiparametric quasi-likelihood. For $b \neq \beta$, I have to make sure that the function $\hat{H}_n(\cdot)$ lies in the interval $[P_3(x,b), P_4(x,b)]$ otherwise we might get a negative value for the probability of the $(1,0)$ outcome. Hence I trim the estimated function $\hat{H}_n(\cdot)$ by replacing it with

$$\hat{H}_{tn}(x,b) = \min\{P_4(x,b), \max\{P_3(x,b), \hat{H}_n(x)\}\} \tag{8}$$

where $\hat{H}_{tn}(x,\beta) - \hat{H}_n(x) = o_p(1)$ for large enough n. Since $P_3(x,b)$ and $P_4(x,b)$ are known functions of x and b, the constrained function $\hat{H}_{tn}$ is easily updated at each iteration of the optimization procedure. Another way to do this is to maximize the likelihood[10] over a subset of the parameter space where $P_3(x,b) \leq \hat{H}_n(x) \leq P_4(x,b)$. Moreover, one needs to restrict the parameter space to make sure that Δ_1 and Δ_2 are negative.

The SML estimator $\hat{\beta}(\hat{H}_{tn})$ maximizes the following quasi-likelihood:

$$\begin{aligned} L(b;\hat{H}_{tn}) = \sum\nolimits_{i=1}^{n} &[y_{i1}y_{i2}\log(P_1(x_i,b)) \\ &+ (1-y_{i1})(1-y_{i2})\log(P_2(x_i,b)) + (1-y_{i1})y_{i2}\log(\hat{H}_{tn}(x_i,b)) \\ &+ y_{i1}(1-y_{i2})\log(1-P_1(x_i,b)-P_2(x_i,b)-\hat{H}_{tn}(x_i,b))]. \end{aligned} \tag{9}$$

Throughout, the following notation is used: $H_i = H(x_i)$, $\hat{H}_{tni} = \hat{H}_{tn}(x_i,\beta)$, $H_t(x,b) = \min\{P_4(x,b), \max\{P_3(x,b), H(x)\}\}$, $H_t = H(x,\beta) = H$ and $P_i = P_i(x,\beta)$ for $i = 1,\dots,4$. In the Appendix, I study the large sample behaviour of the SML estimator where I state the conditions and lemmas needed. The proofs to the lemmas are also collected in the Appendix. In particular, Theorem 5 shows that the SML estimator of β is consistent and normally distributed with an asymptotic variance Ω_{SML} given by

$$\Omega_{\text{SML}} = E_x\left\{\left[\frac{\partial P_1 \partial P_1'}{P_1} + \frac{\partial P_2 \partial P_2'}{P_2} + \frac{(\partial P_1 + \partial P_2)(\partial P_1 + \partial P_2)'}{1-P_1-P_2-H}\right]\right\}^{-1}.$$

The function $H(\cdot)$ enters the above covariance matrix and represents the efficiency gains of the SML estimator. The main result is summarized in the next theorem.

Theorem 3 (Efficiency Gains with Multiple Equilibria). *The SML estimator defined in* (9) *above is more efficient than the ML estimator defined in* (7) *if* $\Pr[(0,1)|x] > 0$.

Proof. If $H > 0$, where $H = \Pr((0,1)|X)$, then

$$1 - P_1 - P_2 - H < 1 - P_1 - P_2$$

implying that

$$\frac{(\partial P_1 + \partial P_2)(\partial P_1 + \partial P_2)'}{1-P_1-P_2-H} - \frac{(\partial P_1 + \partial P_2)(\partial P_1 + \partial P_2)'}{1-P_1-P_2}$$

is positive semi-definite. This means that

$$\Omega_{\text{ML}} - \Omega$$

is positive semi-definite.[11] ||

10. In practice, the estimation in the optimization routine needs to take into account the b in the likelihood as well as in the function $\hat{H}_{tn}$. When defining the objective function for a given b, one checks whether the inequalities that define $\hat{H}_n$ are binding and then replace $\hat{H}_n$ in the objective function and the score function appropriately.

11. Here we use the result that if $A - B$ is positive definite then $B^{-1} - A^{-1}$ is also positive definite.

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

Comment on efficiency gain:

- There are two estimators for $\Pr[(0, 1)|X]$. The first one is just the function $H(X)$. Another one is $1-P_1-P_2-G(X)$ where $G(X) = \Pr[(1, 0)|X]$. Since as we saw above estimation of this probability does not affect the asymptotic variance, one can combine the two estimators in practice to produce a more efficient estimator. This will depend on the proportion in the population of $(0, 1)$ and $(1, 0)$.

This result means that sharper estimates of our parameters can be obtained in the presence of multiple outcomes.

5. CONCLUSION

As discussed above, this paper analyses an incomplete bivariate discrete response model. The latter is a stochastic representation of a simultaneous game, and hence the multiple equilibria in the game maps into a region where the econometric model predicts a nonunique outcome. We have shed a new light on the "coherency" issue by arguing that these (coherency) restrictions should result only from conditions guaranteeing a well-defined economic optimization problem, otherwise the model would be misspecified. Moreover what we have found is that a lot of times the statistical coherency conditions or completeness conditions are not necessary for the model to be estimable. In the model we analyse, the recursivity issue results from the presence of multiple equilibria in the underlying game. We have shown how to obtain a conditions for identification in the presence of multiplicity and how to use the empirical distribution of the data to supplement the econometric model in the region of multiple outcomes, hence obtaining "observable implications" on equilibrium selection (Jovanovic, 1989). Moreover for the case where the model provides only bounds on the conditional probabilities, we show that the parameter is still identified. Exploiting the presence of multiple equilibria, we show how to obtain a consistent and $n^{1/2}$ normally distributed estimate that is more efficient than estimators commonly used. The framework for estimation relied on the fact that the model provided exact probability for two of the outcomes. In the case where the shift parameters have opposite signs, one has to use a different estimation technique. A step in this direction would be an estimator similar to the one used in Manski and Tamer (2002).

APPENDIX A

A.1. *Identification*

Identification Theorem 2. The following is a sketch of the proof. I show first that the coefficients on the continuous regressors are identified. Let β_{ik} be the coefficient that corresponds to the continuous regressor x_{ik} for $i = 1, 2$. Let b be such that $b_{ik} \neq \beta_{ik}$. For given x_{-ik} consider the following set:

$$\{(x_{1k}, x_{2k}) : -x_{1k}b_{1k} - x_{-k}b_{-1k} > -x_{1k}\beta_{1k} - x_{-1k}\beta_{-1k} - \Delta_1;\\ -x_{2k}b_2 - x_{-2k}b_{-2k} > -x_{2k}\beta_{2k} - x_{-2k}\beta_{-2k} - \Delta_2\}.$$

Under the support conditions in the theorem, this set has positive probability. As a result, the lower bound for $P_1(x, \beta)$ is higher than the upper bound for $P_1(x, b)$ implying that b_{ik} is identified relative to β_{ik}.

Here I use the exclusion of a continuous regressor to show that the constant terms are identified. Without lack of generality, suppose the vector x_i is composed of the continuous regressor x_{i1} and a vector of ones (constant term) for $i = 1, 2$. We look first at the probability of $(0, 0)|X$. The lower bound on this probability predicted by the model is $P_1(x, \beta) = \Pr[u_1 < -x_{11}\beta_{11} - \beta_{01}; u_2 < -x_{12}\beta_{12} - \beta_{02}]$ and the upper is $P_1(x, \beta) + P_{\text{square}}$, where P_{square} is defined in (6) above. Let b be such that $-b_{02} > -\beta_{02}$ (the case for $-b_{02} < -\beta_{02}$ is similar). There exists a set of (x_{11}, x_{12}) such that $\Pr[u_1 < -x_{11}\beta_{11} - b_{01}; u_2 < -x_{12}\beta_{12} - b_{02}] > P_1(X, \beta) + \Pr[u_1 > -x_{11}\beta_{11} - \beta_{01}] > P_1(x, \beta) + P_{\text{square}}$. Heuristically this is possible by driving $(-x_{11}\beta_{11} - \beta_{01}, -x_{12}\beta_{12} - \beta_{02})$ "far enough to the right" around the u_1 axis.

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

The same argument applies to β_{01} by driving $(-x_{11}\beta_{11} - \beta_{01}, -x_{12}\beta_{12} - \beta_{02})$ far enough up the u_2 axis. This implies that the constant terms are identified.

As for the coefficients on discrete regressors and the shift parameters, one can use the same arguments as above. For the shift parameters, one can use the probability of the $(1, 1)$ outcome instead of $(0, 0)$. ||

A.2. *Consistency and normality proofs*

I begin by stating the assumptions needed in the proofs. Then I introduce a semiparameteric maximum likelihood estimator that uses a kernel method to approximate the conditional probability of the $(0, 1)$ outcome. I study the properties of this estimator by establishing consistency and $\sqrt{n}$ normality. Throughout, I will use the following norm:

$$\|f\|_r = \max_{j<r} \|f^{(j)}\|$$

where $\|f\|$ is the usual sup norm, and $f^{(j)}$ denotes the j-th derivative and r an integer.

Assumptions

Assumption 5. *The true parameters $\beta = (\beta_1, \beta_2, \Delta_1, \Delta_2)$ and Ω lie in a compact set $\mathbf{\Theta}$. Moreover, (β, Ω) are point identified.*

Assumption 6. *Let the density of x $f(x)$ defined on $D \subset \mathrm{R}^d$ be continuously differentiable of order r such that $\|f\|_r < \infty$ such that $f(x) > 0$ for all x in D.*

Assumption 7. *Assume that* $\Pr[(0, 1) \mid x] = H(x)$, *where at the true parameter β, $P_3(x, \beta) \leq H(x) \leq P_4(x, \beta)$, for all $x \in D$.*

Assumption 8. *Assume that there exists a $\delta > 0$ such that $f(x) > \delta$ for all $x \in D$, the support of x.*

Assumption 9. *The function $H(x)f(x)$ is continuously differentiable of order r such that $\|H(x)f(x)\|_r < \infty$.*

Assumption 10. *$K(\cdot)$ is a Riemann integrable density in $L^2(\mathrm{R}^d)$ $(d > 2)$, differentiable of order r, with bounded derivatives. Moreover, the kernel obeys $\int u_1^{l_1} \dots u_d^{l_{s'}} K(u)du = 0$ for $l_1 + \dots + l_{s'} < m$, and $\int u_1^{l_1} \dots u_d^{l_{s'}} K(u)du \neq 0$ for $l_1 + \dots + l_{s'} = m$.*

Comment on assumptions above:

- In Assumption 5 I require that the parameters are point identified. One can also invoke the sufficient identification conditions in Theorem 1 above.
- One could have replaced the probability of $(1, 0)|x$ by a function $G(x)$ equally since the outcomes are symmetric. I show that estimation of the function $H(x)$ has no effect on the asymptotic distribution of $\hat{\beta}$ and hence one can exploit this symmetry in the problem at hand to obtain a more efficient estimator.
- Assumption 7 above is satisfied if the model is correctly specified and the random sample Assumption 1 holds.
- Assumption 8 is stronger than needed and it might restrict the use of regressors with infinite support. One could, for example, use sophisticated trimming techniques either by basing the trimming on the magnitude of the density of x like the ones used by Klein and Spady (1993) which then showed that the effect of this trimming is minimal on parameter estimates.

The next lemma states results for rates of convergence of kernel estimators of conditional expectations. These results are useful in the proofs below.

Lemma 4 (Lemma 8.10 Newey and McFadden (1994)). *Given Assumptions* 6, 9, 10, *and h is such that $h \to 0$, and $n^{1-(2/p)}h^d/\ln n \to \infty$ then*

$$\|\hat{H}_n(x) - H(x)\| = O_p[(\ln(n))^{1/2}(nh^{d+2r-2})^{-1/2} + h^m]$$

and

$$\|\hat{f}_n(x) - f(x)\| = O_p[(\ln(n))^{1/2}(nh^{d+r})^{-1/2} + h^m].$$

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

Let the kernel estimator of the conditional probability $\hat{H}(\cdot)$ be defined as

$$\hat{H}_n(x) = \frac{1}{nh^d \hat{f}_n(x)} \sum_{i=1}^{N} (1 - y_{i1}) y_{i2} K\left(\frac{x - x_i}{h}\right)$$

where $\hat{f}_n(x)$ is the density of x

$$\hat{f}_n(x) = \frac{1}{nh^d} \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right).$$

For a given parameter vector b, and as in (8) above let the trimmed function $\hat{H}_{tn}(x, b)$

$$\hat{H}_{tn}(x, b) = \min\{P_4(x, b), \max\{\hat{H}_n(x), P_3(x, b)\}\}$$

where in the following $H_t(x, b)$ is the same as $\hat{H}_{tn}(x, b)$ with $\hat{H}$ replaced with H. Also, I use H_t and H to designate $H(x)$, and P_{1i} to designate $P_1(x_i, \beta)$ and similarly for P_{2i}, P_{3i} and P_{4i}. Also, let $\hat{H}_{ni}$ designate $\hat{H}_n(x_i)$ and similarly for H_i. The next theorem summarizes the results on the asymptotic distribution of the SML estimator.

Theorem 5 (SML Estimator). *Given Assumptions* 1–3 *and* 5–7 *above and the conditions stated in Lemma* 4,

$$\hat{\beta}_n = \arg\max_p L_n[b; \hat{H}_{tn}(x, b)] \xrightarrow{p} \beta \qquad a.s. \qquad as\ n \to \infty.$$

Let the conditions in Lemmas 6 *and* 7 *hold. Then, the asymptotic distribution of* $\sqrt{n}(\hat{\beta} - \beta)$ *is* $N(0, \Omega)$, *where*

$$\Omega = E\left\{\left[\frac{\partial P_1 \partial P_1'}{P_1} + \frac{\partial P_2 \partial P_2'}{P_2} + \frac{(\partial P_1 + \partial P_2)(\partial P_1 + \partial P_2)'}{1 - P_1 - P_2 - H}\right]\right\}^{-1}.$$

Proof (Consistency Proof). Let

$$\begin{aligned} L_n(b; \hat{H}_{tn}) = \frac{1}{N} \sum_{i=1}^{n} [&y_{i1} y_{i2} \log(P_1(x_i, b)) + (1 - y_{i1})(1 - y_{i2}) \log(P_2(x_i, b)) \\ &+ (1 - y_{i1}) y_{i2} \log(\hat{H}_{tn}(x_i, b)) + y_{i1}(1 - y_{i2}) \log(1 - P_1(x_i, b) - P_2(x_i, b) - \hat{H}_{tn}(x_i, b)). \end{aligned}$$

Next, I show that the likelihood function evaluated at the estimated function $\hat{H}_{tn}$ is close to the likelihood function evaluated at $H_t(x, b)$ uniformly in b. This allows us to focus on consistency of the likelihood where I replace $\hat{H}$ with H:

$$\begin{aligned} |L_n[b; \hat{H}_{tn}(b)] - L_n[b; H_t(b)]| = \Bigg| &\frac{1}{N} \sum_{i=1}^{n} (1 - y_{i1}) y_{i2} [\log(\hat{H}_{tn}(x_i, b)) - \log(H_t(x_i, b))] \\ &+ y_{i1}(1 - y_{i2})[\log(1 - P_1(x_i, b) - P_2(x_i, b) - \hat{H}_{tn}(x_i, b)) - \log(1 - P_1(x_i, b) - P_2(x_i, b) - H_t(x_i, b))\Bigg|. \end{aligned}$$

Notice[12] that $|\log(\hat{H}_{tn}(x_i, b)) - \log(H_t(x_i, b))| \le |\log(\hat{H}_n(x_i)) - \log(H(x_i))|$. By consistency of the kernel estimator (Lemma 4) we have that $|\log(\hat{H}_n(x_i)) - \log(H(x_i))| \xrightarrow{p} 0$. This means that $|L_n[b; \hat{H}_{tn}(b)] - L_n[b; H_t(b)]|$ is $o_p(1)$ uniformly in b almost surely. Given the usual regularity conditions, $L_n[b; H_t(b)]$ converges to its expectation. To prove that the expectation of the likelihood above is maximized at the true parameter value, consider the following:

$$\begin{aligned} E[L(b; H_t) - L(\beta; H)] &= E\left[\begin{array}{c} \ln\left[\frac{P_1(b)}{P_1(\beta)}\right]^{1(y_1, y_2)[(1,1)]} \left[\frac{P_2(b)}{P_2(\beta)}\right]^{1(y_1, y_2)[(0,0)]} \left[\frac{H_t(b)}{H}\right]^{1(y_1, y_2)[(0,1)]} \\ \left[\frac{1 - P_1(b) - P_2(b) - H_t(b)}{1 - P_1(\beta) - P_2(\beta) - H}\right]^{1(y_1, y_2)[(1,0)]} \end{array}\right] \\ &\le \ln E_x\left[\frac{P_1(b)}{P_1(\beta)} P_1(\beta) + \frac{P_2(b)}{P_2(\beta)} P_2(\beta) + \frac{H_t(b)}{H} H \right. \\ &\quad \left. + \frac{1 - P_1(b) - P_2(b) - H_t(b)}{1 - P_1(\beta) - P_2(\beta) - H} 1 - P_2(\beta) - P_2(\beta) - H\right] \\ &= \ln E_x[1] = 0. \end{aligned}$$

Here we have used Jensen's inequality for convex functions. This implies that the objective function is maximized at the true parameter value. Uniform convergence added with the compactness of the parameter space and the usual regularity conditions will imply consistency. ||

12. This follows from $|\log\max(a, c) - \log\max(a, b)| \le |\log c - \log b|$. The same also holds for the min function.

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

For deriving the asymptotic distribution of the SML estimator consider the Taylor series expansion for the gradient of the quasi-likelihood around the true value of the parameter β

$$\sqrt{n}(\hat{\beta}-\beta) \simeq -[\partial_\beta^2 L_n(\bar{\beta};\hat{H}_{tn})]^{-1}\sqrt{n}\partial_\beta L_n(\beta;\hat{H}_{tn}) \tag{A.1}$$

where $\bar{\beta}$ is the mean value lying on the line joining $\hat{\beta}$ to β. Since the score function is evaluated at the true parameter vector, and for large enough n, $\hat{H}_{tn}(x,\beta)$ is close to $H_t(x,\beta)$ which is equal to $H(x)$. Hence for large n, $\hat{H}_{tn}(x,\beta)$ will not depend on β and I will replace the function $\hat{H}_{tn}(x,\beta)$ with $\hat{H}_n(x)$.[13] Let $S_i(\beta;\hat{H}_n(x_i))$ be defined as

$$\begin{aligned}\frac{1}{n}\sum_i S(\beta;\hat{H}_n(x_i)) &= \partial_\beta L_n[\beta;\hat{H}_n]\\ &= \frac{1}{n}\sum_{i=1}^N\left[\frac{\partial_\beta P_{1i}}{P_{1i}}1_{(y_{i1},y_{i2})}(1,1)+\frac{\partial_\beta P_{2i}}{P_{2i}}1_{(y_{i1},y_{i2})}(0,0)\right.\\ &\qquad\left.-\frac{(\partial_\beta P_{1i}+\partial_\beta P_{2i})}{(1-P_{1i}-P_{2i}-\hat{H}_n(x_i))}1_{(y_{i1},y_{i2})}(1,0)\right]\end{aligned}$$

where $P_{ji}=P_j(x_i,\beta)$, and $1_{(y_{i1},y_{i2})}(1,1)=\begin{cases}1 & \text{if } (y_{i1},y_{i2})=(1,1)\\ 0 & \text{otherwise.}\end{cases}$

I now concentrate on the gradient term $\sum S(z_i;\beta;\hat{H}_{ni})/\sqrt{n}$ in (A.1). Define the following:

$$\Gamma(\hat{H}_n)=\frac{1}{N}\sum S(z_i;\hat{H}_{ni})-\frac{1}{N}\sum S(z_i;H_i).$$

I show that

$$\sqrt{n}\Gamma(\hat{H}_n)=o_p(1)$$

and hence that the asymptotic distribution does not depend on the estimation of the infinite-dimensional parameter. Then I use a central limit type of result to find the asymptotic distribution of $\frac{1}{N}\sum S(z_i;H_i)$.

We have

$$\Gamma(\hat{H}_n)=\frac{1}{N}\sum_{i=1}^N\left(\frac{(\partial_\beta P_{1i}+\partial_\beta P_{2i})}{(1-P_{1i}-P_{2i}-\hat{H}_{ni})}-\frac{(\partial_\beta P_{1i}+\partial_\beta P_{2i})}{(1-P_{1i}-P_{2i}-H_i)}\right)1_{(y_{i1},y_{i2})}(1,0)$$

where the function $\hat{H}$ is of the following form:

$$\hat{H}_{ni}(X)=\frac{\sum 1_{(y_{i1},y_{i2})}(0,1)k\left(\frac{X-X_i}{h}\right)}{\sum k\left(\frac{X-X_i}{h}\right)}=\frac{\hat{m}(X)}{\hat{f}(X)}$$

where $\hat{m}(X)=\frac{1}{Nh^d}\sum 1_{(y_{i1},y_{i2})}(0,1)k\left(\frac{X-X_i}{h}\right)$ is the kernel estimate of the joint distribution of X and $(0,1)$, and $\hat{f}(X)=\frac{1}{Nh^d}\sum k\left(\frac{X-X_i}{h}\right)$ is the kernel estimator of the density of the vector X. To be able to deal with the ratio, we need to linearize $\Gamma(\hat{H}_n)$ as a function of $\hat{m}(X)$ and $\hat{f}(X)$ functions using the following:

$$\frac{a}{b}-\frac{a_0}{b_0}\cong\frac{1}{b_0}(a-a_0)-\frac{a_0}{b_0^2}(b-b_0)$$

where the remainder is of order $O((a-a_0)^2,(b-b_0)^2)$. This implies that

$$\begin{aligned}\Gamma(\hat{H}_n) &= \frac{1}{N}\sum_{i=1}^N\frac{(\partial_\beta P_{1i}+\partial_\beta P_{2i})}{(1-P_{1i}-P_{2i}-H_i)}\left(\frac{\hat{f}_i-f_i}{f_i}\right)\\ &\quad+\frac{(\partial_\beta P_{1i}+\partial_\beta P_{2i})(1-P_{1i}-P_{2i})}{(1-P_{1i}-P_{2i}-H_i)((1-P_{1i}-P_{2i})f_i-m_i)}(\hat{f}_i-f_i)\\ &\quad-\frac{(\partial_\beta P_{1i}+\partial_\beta P_{2i})}{(1-P_{1i}-P_{2i}-H_i)((1-P_{1i}-P_{2i})f_i-m_i)}(\hat{m}_i-m_i)\\ &\quad+\text{remainder}\\ &= (1)+(2)+(3)+\text{remainder}\end{aligned}$$

where the remainder is of the following order:

$$\text{Rem}=O(\|\hat{m}_n-m\|^2,\|\hat{f}_n-f\|^2).$$

13. Alternatively, we could drop those observations for which $\hat{H}_n(x_i)>P_4(x_i,\beta)$ $\hat{H}_n(x_i)<P_3(x_i,\beta)$.

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

I show below that the three terms (1), (2) and (3) and the remainder tend to zero faster than root n which allows us to replace the estimated $\hat{H}$ with H in (A.1) and use a central limit theorem. The lemma below deals with the first term. The other two are similar.

Lemma 6. *Given Assumptions* 8–10, *we have that*

$$\sqrt{n}(\text{term }(1)) = o_p(1).$$

The same applies for terms (2) *and* (3).

Let us write term (1) in a different manner:

$$\begin{aligned}\sqrt{n}(\text{term }(1)) &= \sum_{i=1}^{N} N^{-1/2}\frac{(\partial_\beta P_{1i} + \partial_\beta P_{2i})}{(1 - P_{1i} - P_{2i} - H_i)}\left(\frac{\hat{f}_i - f_i}{f_i}\right)\\ &= \sum_{i=1}^{N} N^{-1/2}T(z_i, \beta)(\hat{f}_i - f_i).\end{aligned}$$

It suffices to show that the above converges in mean-square:

$$\begin{aligned}E\left[\sum_{i=1}^{N} N^{-1/2}T(z_i,\beta)(\hat{f}_i - f_i)\right]^2 &= E\left[\sum_{i=1}^{N} T(z_i,\beta)(\hat{f}_i - f_i)\right]^2\Big/N\\ &= E\left[\sum T^2(z_i,\beta)(\hat{f}_i - f_i)^2/N\right]\\ &\quad + E\left[\sum_{i\neq j} T(z_i,\beta)(\hat{f}_i - f_i)(\hat{f}_j - f_j)T(z_j,\beta)/N\right].\end{aligned}$$

Since $T^2(z,\beta)(\hat{f} - f)^2 \xrightarrow{p} 0$, and by a uniform boundedness condition, the first term above goes to zero. The second condition goes to zero by first using the law of iterated expectation and conditioning on z_i and then z_j. The same applies to the other two terms (2) and (3). ||

The next lemma guarantees that the remainder has the appropriate rate.

Lemma 7. *If* (i) $\sqrt{n}h^{2m} \to 0$, (ii) $\sqrt{n}\ln(n)/nh^{2r+d} \to 0$, (iii) $nh^d \to \infty$, *and* $nh^d/\ln(n) \to \infty$ *we get*

$$\sqrt{n}\|\hat{m}_n - m\|^2 = o_p(1)$$

and

$$\sqrt{n}\|\hat{f}_n - f\|^2 = o_p(1).$$

Proof. Basically we need condition (i) to control the bias, but at the same time we need (iv) to make sure the variance goes to zero. Condition (ii) guarantees uniform convergence. These conditions will be satisfied for a range of bandwidth sequences h_n if the kernel we use is of high enough order ($m \gg 0$) and the density has enough derivatives ($r \gg 0$). ||

The next lemma deals with the Hessian term in (A.1).

Lemma 8. *Given the assumptions above,*

$$-[\partial^2 L_n[b;\hat{H}_n(x,b)]] \xrightarrow{p} E\left\{\left[\frac{\partial P_1 \partial P_1'}{P_1} + \frac{\partial P_2 \partial P_2'}{P_2} + \frac{(\partial P_1 + \partial P_2)(\partial P_1 + \partial P_2)'}{1 - P_1 - P_2 - H}\right]\right\}_{b\to\beta}.$$

Proof. First taking the derivative of the score function with respect to b we get

$$\begin{aligned}\partial^2 L_n[b;\hat{H}_{tn}(x,b)] = \frac{1}{n}\sum_i \Bigg[&\frac{P_{1i}\partial^2 P_{1i} - \partial P_{1i}\partial P_{1i}'}{P_{1i}^2} 1_i(1,1) + \frac{P_{2i}\partial^2 P_{2i} - \partial P_{2i}\partial P_{2i}'}{P_{2i}^2} 1_i(0,0)\\ &- \frac{(\partial^2 P_{1i} + \partial^2 P_{2i})(1 - P_{1i} - P_{2i} - \hat{H}_n(x_i)) + (\partial P_{1i} + \partial P_{2i})(\partial P_{1i} + \partial P_{2i})'}{(1 - P_{1i} - P_{2i} - \hat{H}_n(x_i))^2} 1_i(1,0)\Bigg] \quad \text{(A.2)}\end{aligned}$$

where for example $1_i(1,1) = 1$ if $(y_{1i}, y_{2i}) = (1,1)$. Using similar arguments to ones used above, one can show that

$$|\partial^2 L_n[b;\hat{H}_{tn}(x,b)] - \partial^2 L_n[b;H_t(x,b)]| = o_p(1)$$

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

uniformly in b in a shrinking neighbourhood around β. Using a law of large number type result we get that since $\hat{\beta}$ is consistent the convergence results above insure that for $\bar{\beta} \in [\hat{\beta}, \beta]$,

$$\partial^2 L_n[\bar{\beta}; H_t(x, \bar{\beta})] \xrightarrow{p} E\partial^2 L_n[\beta; H] \text{ as } n \to \infty$$

where by conditioning first on x and taking expectations in (A.2) we get the result of the lemma. ||

Normality Proof:

The above implies that

$$\sqrt{n}\left\| \frac{1}{N}\sum S_i(\beta, \hat{H}_{tn}(x_i, \beta)) - \frac{1}{N}\sum S_i(\beta, H(x_i)) \right\| = o_p(1)$$

which implies that $\sum S_i(\beta; \hat{H}_{tni})/\sqrt{n}$ has the same asymptotic distribution as $\sum S_i(\beta; H_i)/\sqrt{n}$. We have

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta) &= [-\partial^2 L(\beta; H)]^{-1}\frac{1}{\sqrt{n}}\sum\nolimits_i S_i(\beta, H_i) + o_p(1)\\ &= [-\partial^2 L(\beta; H)]^{-1}\frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^N \left[\frac{\partial P_{1i}}{P_{1i}} 1_{(y_{i1}, y_{i2})}(1, 1) + \frac{\partial P_{2i}}{P_{2i}} 1_{(y_{i1}, y_{i2})}(0, 0)\right.\\ &\qquad \left. - \frac{(\partial P_{1i} + \partial P_{2i})}{(1 - P_{1i} - P_{2i} - \hat{H}_{ni})} 1_{(y_{i1}, y_{i2})}(1, 0)\right] + o_p(1).\end{aligned}$$

By a central limit theorem,

$$\frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^N \left[\frac{\partial P_{1i}}{P_{1i}} 1_{(y_{i1}, y_{i2})}(1, 1) + \frac{\partial P_{2i}}{P_{2i}} 1_{(y_{i1}, y_{i2})}(0, 0) - \frac{(\partial P_{1i} + \partial P_{2i})}{(1 - P_{1i} - P_{2i} - H_i)} 1_{(y_{i1}, y_{i2})}(1, 0)\right]$$

tends to a normal distribution with mean zero and the variance matrix equals the variance of

$$\left[\frac{\partial P_1}{P_1} 1_{(y_1, y_2)}(1, 1) + \frac{\partial P_2}{P_2} 1_{(y_1, y_2)}(0, 0) - \frac{(\partial P_1 + \partial P_2)}{(1 - P_1 - P_2 - H)} 1_{(y_1, y_2)}(1, 0)\right].$$

This is equal to

$$E\left[\frac{\partial P_1 \partial P_1'}{P_1^2} 1_{(1,1)} + \frac{\partial P_2 \partial P_2'}{P_2^2} 1_{(0,0)} + \frac{(\partial P_1 + \partial P_2)(\partial P_1 + \partial P_2)'}{(1 - P_1 - P_2 - H)^2} 1_{(1,0)}\right].$$

Hence taking first the expectation with respect to y conditional on x and using Lemma 6, we get that

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N\left(0, E\left\{\left[\frac{\partial P_1 \partial P_1'}{P_1} + \frac{\partial P_2 \partial P_2'}{P_2} + \frac{(\partial P_1 + \partial P_2)(\partial P_1 + \partial P_2)'}{1 - P_1 - P_2 - H}\right]\right\}^{-1}\right). \quad ||$$

Acknowledgements. I would like to thank Tim Conley, Ian Domowitz, Rob Porter, Tom Severini, and especially Chuch Manski and Rosa Matzkin for constructive and insightful comments. I am grateful to Orazio Attanasio and three anonymous referees for providing valuable comments that improved the paper. I would also like to thank seminar participants at BU, Caltech, Columbia GSB, Duke, MIT, UPenn, Princeton and Yale. All errors are mine. Financial support from the Alfred P. Sloan Dissertation Fellowship in Economics is gratefully acknowledged.

REFERENCES

BERRY, S. (1992), "Estimation of a Model of Entry in the Airline Industry", *Econometrica*, **60** (4), 889–917.

BJORN, P. A. and VUONG, Q. H. (1985), "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation" (Working Paper No. 537, Caltech, Pasadena, CA).

BLUNDELL, R. and SMITH, J. R. (1993), "Simultaneous Microeconometric Models with Censored or Qualitative Dependent Variables", in G. S. Maddala, C. R. Rao and H. D. Vinod (eds.) *Handbook of Statistics*, Vol. 11 (New York: North-Holland) 117–143.

BLUNDELL, R. and SMITH, R. J. (1994), "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables", *Journal of Econometrics*, **64**, 355–373.

BRESNAHAN, T. F. and REISS, P. C. (1990), "Entry in Monopoly Markets", *Review of Economic Studies*, **57** (4), 57–81.

BRESNAHAN, T. F. and REISS, P. C. (1991), "Empirical Models of Discrete Games", *Journal of Econometrics*, **48**, 57–81.

GOURIEROUX, J., LAFONT, J. J. and MONFORT, A. (1981), "Coherency Conditions in Simultaneous Linear Equation Models with Endogenous Switching Regimes", *Econometrica*, **48** (3), 675–696.

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019

HECKMAN, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System", *Econometrica*, **46**, 931–959.

JOVANOVIC, B. (1989), "Observable Implications of Models with Multiple Equilibria", *Econometrica*, **57**, 1431–1438.

KLEIN, R. W. and SPADY, R. H. (1993), "An Efficient Semiparametric Estimator for Binary Response Models", *Econometrica*, **61**, 387–421.

KOOREMAN, P. (1994), "Estimation of Econometric Models of Some Discrete Games", *Journal of Applied Econometrics*, **9**, 255–268.

MADDALA, G. S. (1983) *Limited-dependent and Qualitative Variables in Econometrics* (Cambridge: Cambridge University Press).

MANSKI, C. (1988) *Analog Estimation Methods in Econometrics* (New York: Chapman and Hall).

MANSKI, C. and TAMER, E. (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome", *Econometrica*, **70** (2), 519–546.

NEWEY, W. K. (1994), "The Asymptotic Variance of Semiparametric Estimators", *Econometrica*, **62**, 1349–1382.

NEWEY, W. K and McFADDEN, D. (1994), "Large Sample Estimation", in *Handbook of Econometrics*, Vol. 4.

SCHMIDT, P. (1981), "Constraints on the Parameters in Simultaneous Tobit and Probit Models", in C. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications* (Cambridge, MA: MIT Press) 422–434.

VAN SOEST, A. and KOOREMAN, P. (1990), "Coherency of the Indirect Translog Demand System with Binding Nonnegativity Constraints", *Journal of Econometrics*, **44**, 391–400.

WALES, T. and WOODLAND, A. (1983), "Estimation of Consumer Demand Systems with Binding Nonnegativity Constraints", *Journal of Econometrics*, **21**, 263–285.

Downloaded from https://academic.oup.com/restud/article-abstract/70/1/147/1560993 by University of Toronto user on 07 January 2019