

# Identification in Simple Binary Outcome Panel Data Models<sup>1</sup>

BO E. HONORÉ<sup>†</sup> AND ÁUREO DE PAULA<sup>‡</sup>

<sup>†</sup>*Department of Economics, Princeton University  
and The Dale T. Mortensen Centre, University of Aarhus*

E-mail: [honore@princeton.edu](mailto:honore@princeton.edu)

<sup>‡</sup>*Department of Economics, University College London,  
CeMMAP and IFS, London, UK.*

E-mail: [a.paula@ucl.ac.uk](mailto:a.paula@ucl.ac.uk)

**Summary** This paper first reviews some of the approaches that have been taken to estimate the common parameters of binary outcome models with fixed effects. We limit attention to situations in which the researcher has access to a data set with a large number of units (individuals or firms, for example) observed over a few time periods. We then apply some of the existing approaches to study fixed effects panel data versions of entry games, like the ones studied in Bresnahan and Reiss (1991) and Tamer (2003).

**Keywords:** *Panel Data, Fixed Effects, Discrete Choice, Entry Games.*

## 1. GENERAL SETUP

It is natural to model decisions made by individuals in terms of the information available to them when the choice is made. This motivates a general panel data setup in which the distribution of a dependent variable (or vector) in time period  $t$ ,  $y_{it}$ , can be modelled as a function of its own past values,  $y_i^{t-1} = \{y_{is}\}_{s < t}$ , a vector of explanatory variables up to time  $t$ ,  $x_i^t = \{x_{is}\}_{s \leq t}$ , and an unobserved individual-specific characteristic,  $\alpha_i$ . Consequently,

$$y_{it} \sim f(\cdot | x_i^t, y_i^{t-1}, \alpha_i; \theta), \quad (1.1)$$

where  $f$  is the distribution of  $y_{it}$  conditional on  $x_i^t$ ,  $y_i^{t-1}$ , and  $\alpha_i$ , and  $\theta$  is the vector of parameters. Throughout this paper, we treat  $\alpha_i$  as a “fixed effect” in the sense that its distribution is allowed to depend on the explanatory variables in an arbitrary way.

In specification (1.1), the explanatory variable is allowed to be predetermined so that future realizations of  $x$  may depend on the realization of  $y$  in the current period. This is attractive from an economic point of view when  $y$  is the outcome of a choice as indicated above. An individual makes a decision,  $y_{it}$ , based on her information at that point. The information set contains the covariates that she has observed until now,  $x_i^t$ , her past choices,  $y_i^{t-1}$ , and her time-invariant characteristics,  $\alpha_i$  (which are unobserved to the econometrician).

While it is possible to allow for predetermined explanatory variables in models where the fixed effect,  $\alpha_i$ , enters linearly or multiplicatively on the outcome variable,  $y_{it}$ , we are

<sup>1</sup>We thank Sharada Dharmasankar, Jaap Abbring and two anonymous referees for helpful comments. This research was supported by the Gregory C. Chow Econometric Research Program at Princeton University, by the National Science Foundation (Grant Numbers SES-1530741 and SES-1824131), and ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001

not aware of any results that allow for this in panel data discrete response models where the distribution of  $\alpha_i$  is left unrestricted. We therefore maintain throughout the stronger modelling setup where explanatory variables are strictly exogenous and dependence on the whole sequence of covariates is considered:

$$y_{it} \sim f(\cdot | x_i^T, y_i^{t-1}, \alpha_i; \theta). \quad (1.2)$$

The additional restrictions embedded in (1.2) rule out that individuals choose  $x$  in time period  $t$  in response to the outcomes of  $y$  in periods prior to  $t$ . As mentioned above, this will sometimes make it unattractive in economic applications. However, (1.2) makes it possible to make probability statements on the whole sequence (over time) of  $y_{it}$  conditional on the whole sequence of the explanatory variables.

It is important to recognize that knowing  $\theta$  in (1.1) or (1.2) is typically not sufficient for calculating counterfactual distributions or marginal effects. Those will depend on the distribution of  $\alpha_i$  as well as on  $\theta$  and they are typically not point-identified even if  $\theta$  is. For a discussion of this see, for example, Chernozhukov et al. (2013). On the other hand, it seems that point- or set-identifying and estimating  $\theta$  is a natural first step if one is interested in bounding, say, average marginal effects.

In Section 2 of this paper, we first review some approaches for estimating univariate binary outcome versions of (1.1). The traditional approach is to find a sufficient statistic for the fixed effects and then proceed by conditional maximum likelihood (conditioning on the sufficient statistic). This approach dates back to Rasch (1960, 1961), and a recent example includes Aguirregabiria et al. (2020). When it is not possible to find a sufficient statistic for the fixed effects, it is sometimes possible to construct moment equality conditions which must be satisfied at the true parameter value. See Johnson (2004) and Honoré and Weidner (2020) for an early and recent example, respectively. Significant progress has also been made by employing moment inequality conditions. See for example Manski (1987) and, more recently Pakes and Porter (2016) or Pakes et al. (2021). Section 3 discusses bivariate binary outcome models. We first describe some recent advances for reduced form models, and we then analyse a simple panel data version of an entry game. Section 4 concludes.

## 2. THE INCIDENTAL PARAMETERS PROBLEM

It is well understood that estimating the individual-specific effects,  $\{\alpha_i\}$ , along with the common parameter,  $\theta$ , typically (though not always) leads to inconsistent estimation of  $\theta$  in a panel where the number of time periods is fixed and to asymptotic bias in “large” panels, where both the number of time periods and the number of micro-units increase. This is known as the incidental parameters problem. See Neyman and Scott (1948).

There are many papers that attempt to eliminate the asymptotic bias in “large” panels. These include Hahn and Newey (2004), Arellano and Bonhomme (2009), Dhaene and Jochmans (2015), and Fernández-Val and Weidner (2016). These papers consider procedures that are justified asymptotically as the number of time periods grows with the number of individuals. See, for example, Fernández-Val and Weidner (2018) for a review of this literature. A different set of papers tries to construct methods that work when the panel contains observations for a large number of micro-units observed in a few time periods. This is the situation that we consider in this paper. Specifically, in this section, we briefly review three alternative approaches for dealing with individual-specific parameters in standard binary response models that have been explored in the

literature: conditional likelihood, construction of moment conditions, and moment inequalities. This list is by no means exhaustive. For example, a number of papers have explored the usefulness of restricting the relationship between individual-specific effects and one of the explanatory variables. This includes papers like Chen et al. (2016). A different set of papers places restrictions on the distribution of the fixed effects. For example, Bonhomme and Manresa (2015) assume that its marginal distribution is discrete with a finite number of points of support.

## 2.1. Conditional Likelihood

The traditional approach for obtaining consistent estimators of the common parameters in a parametric model with incidental parameters is to condition on a set of sufficient statistics for the individual-specific parameters. This was proposed by Rasch (1960, 1961) and studied in detail by Andersen (1970). Suppose that the distribution of  $y_i^T$  conditional of  $x_i^T$  and the individual-specific effects has been specified as a function of the common parameter,  $\theta$ . The idea behind conditional likelihood is that if there exists a (possibly vector-valued) function of the data for individual  $i$ ,  $S_i$ , such that (a) the distribution of  $y_i$  conditional on  $(S_i, x_i, \alpha_i)$  does not depend on  $\alpha_i$  (i.e.,  $S_i$  is a sufficient statistic for  $\alpha_i$ ), and (b) the distribution of  $y_i$  conditional on  $(S_i, x_i, \alpha_i)$  depends on  $\theta$ , then one can estimate  $\theta$  by maximum likelihood using the conditional distribution of the data given  $(S_i, x_i)$ . Andersen (1970) shows that the conditional maximum likelihood estimator is consistent and asymptotically normal under mild regularity conditions.

The main limitation of the conditional likelihood approach is that in binary response settings, it is typically not possible to find a statistic,  $S_i$ , with the properties described above. The main exceptions include a number of logit models, some of which are discussed below.

*2.1.1. Simple Examples: Logit Models* Rasch (1960, 1961) considered a static panel data version of the standard logit model,

$$P(y_{it} = 1 | x_i^T, y_i^{t-1}, \alpha_i) = \frac{\exp(x_{it}'\beta + \alpha_i)}{1 + \exp(x_{it}'\beta + \alpha_i)} = \Lambda(x_{it}'\beta + \alpha_i), \quad (2.3)$$

where  $\Lambda(\cdot)$  is the logistic cumulative distribution function.

In this case, the distribution of  $(y_{i1}, \dots, y_{iT})$  conditional on  $(x_{i1}, \dots, x_{iT})$  and on  $S_i = \sum_{t=1}^T y_{it}$  does not depend on  $\alpha_i$ . If  $T \geq 2$  it nonetheless does depend on  $\beta$ :

$$P\left(\{y_{it}\}_{t=1}^T = \{c_{it}\}_{t=1}^T \mid \{x_{it}\}_{t=1}^T, \alpha_i, \sum_{t=1}^T y_{it} = \sum_{t=1}^T c_{it}\right) = \frac{\exp\left(\sum_{t=1}^T c_{it}x_{it}'\beta\right)}{\sum_{d_t \in \mathcal{B}_i} \exp\left(\sum_{t=1}^T d_t x_{it}'\beta\right)},$$

where  $c_{it} \in \{0, 1\}$  and

$$\mathcal{B}_i = \left\{ (d_1, \dots, d_T) : d_t \in \{0, 1\}, \sum_{t=1}^T d_t = \sum_{t=1}^T c_{it} \right\}.$$

As a result, the conditional likelihood can be used to identify and estimate  $\beta$  in the static panel data logit model. Unfortunately, this does not generalize to other simple models such as the probit model. Indeed, Chamberlain (2010) showed that in a model

of the form  $P(y_{it} = 1 | x_i^T, y_i^{t-1}, \alpha_i) = F(x_{it}'\beta + \alpha_i)$ , regular root- $n$  estimation of  $\beta$  without additional assumptions is only possible if  $F$  is the logistic cumulative distribution function.

The conditional likelihood approach can also be used to estimate some simple panel data autoregressive logit models. See, for example, Chamberlain (1985) and Magnac (2000). Consider the simple model

$$P(y_{it} = 1 | y_i^{t-1}, \alpha_i) = \frac{\exp(y_{it-1}\gamma + \alpha_i)}{1 + \exp(y_{it-1}\gamma + \alpha_i)}, \text{ for } t = 2, \dots, T. \quad (2.4)$$

Since (2.4) models an outcome in terms of its past value, we only insist that it applies starting in the second time period. The first observation,  $y_{i1}$ , is usually referred to as the initial condition.

Conditional on  $S_i = \left(y_{i1}, \sum_{t=1}^T y_{it}, y_{iT}\right)$ , the distribution of  $(y_{i1}, \dots, y_{iT})$  does not depend on  $\alpha_i$ . However, for  $T \geq 4$ , it does depend on  $\gamma$  when  $y_{i1} \neq y_{i4}$ . The corresponding conditional likelihood can therefore be used to identify and estimate  $\gamma$ .

This approach has been extended to an AR(2) model:

$$P(y_{it} = 1 | y_i^{t-1}, \alpha_i, \gamma_{i1}) = \frac{\exp(\gamma_{i1}y_{it-1} + \gamma_{i2}y_{it-2} + \alpha_i)}{1 + \exp(\gamma_{i1}y_{it-1} + \gamma_{i2}y_{it-2} + \alpha_i)}. \quad (2.5)$$

In this model there are two fixed effects,  $\gamma_{i1}$  and  $\alpha_i$ , and when the coefficient on  $y_{it-2}$  is 0, the model corresponds to a Markov switching model with individual-specific transition probabilities. Here, the initial conditions are  $y_{i1}$  and  $y_{i2}$ .

The sufficient statistic for  $(\gamma_{i1}, \alpha_i)$  in (2.5) with  $T \geq 6$  is  $(y_{i1}, y_{i2}, s_{i1}, s_{i11}, y_{iT-1}, y_{iT})$ , where  $s_{i1} = \sum_{t=1}^T y_{it}$  and  $s_{i11} = \sum_{t=2}^T y_{it}y_{it-1}$ , and the corresponding conditional likelihood can be used to estimate  $\gamma_2$ .<sup>1</sup> Magnac (2000) showed that this generalizes to AR( $p$ ) panel data logit models. In a model like (2.5), with  $p$  lags rather than 2, it is possible to find a vector of sufficient statistics,  $S_i$ , such that the distribution of  $(y_{i1}, \dots, y_{iT})$  conditional on  $S_i$  does not depend on  $(\alpha_i, \gamma_{i1}, \dots, \gamma_{ip-1})$ , but for  $T$  sufficiently large, it does depend on  $\gamma_p$ . The corresponding conditional likelihood can therefore be used to identify and estimate  $\gamma_p$  with no assumptions made on  $(\alpha_i, \gamma_{i1}, \dots, \gamma_{ip-1})$ .

While the model in (2.5) illustrates the usefulness of the conditional likelihood approach, it also illustrates its limitation. Suppose that one is willing to assume that  $\gamma_{i1}$  is homogeneous (so  $\gamma_{i1} = \gamma_1$  for all  $i$ ) so

$$P(y_{it} = 1 | y_i^{t-1}, \alpha_i, \gamma_{i1}) = \frac{\exp(\gamma_1 y_{it-1} + \gamma_2 y_{it-2} + \alpha_i)}{1 + \exp(\gamma_1 y_{it-1} + \gamma_2 y_{it-2} + \alpha_i)}. \quad (2.6)$$

In this case, the numerical calculations in Honoré and Kyriazidou (2019b) suggest that  $(\gamma_1, \gamma_2)$  is identified for  $T \geq 5$ . Specifically, Honoré and Kyriazidou (2019b) assume values of  $\gamma_1$  and  $\gamma_2$  and a distribution for  $\alpha_i$  conditional on the initial conditions,  $y_{i1}$  and  $y_{i2}$ . This implies a distribution,  $\tilde{P}$ , for  $(y_{i3}, y_{i4}, y_{i5})$  conditional on  $y_{i1}$  and  $y_{i2}$ . For a fine grid of potential values of  $\gamma_1$  and  $\gamma_2$ , they then ask whether one can find a heterogeneity

<sup>1</sup>The conditional likelihood approach can also be used to estimate models where the coefficient  $\gamma_2$  differs depending on the value of  $y_{it-1}$ . This is, for example, relevant if one does not want to tie the parameters that govern the transition out of employment to the parameters that govern the transition out of non-employment.

distribution (conditional on the initial conditions) that produces the probabilities,  $\tilde{P}$ , using the values on the grid. They find numerically that this is only possible when  $\gamma_1$  and  $\gamma_2$  take the true values. This suggests that  $\gamma_1$  and  $\gamma_2$  are both point identified. However, they also note that it seems that conditioning on any statistic that eliminates  $\alpha_i$  in a conditional likelihood will also eliminate  $\gamma_1$ . This suggests that even in a simple model like (2.6) where conditioning that eliminates  $\alpha_i$  is possible, there is additional information not captured by the conditional likelihood approach. We turn to this in the next subsection.

## 2.2. Moments

The observation that  $(\gamma_1, \gamma_2)$  appears to be identified in (2.6) is the inspiration for a recent paper by Honoré and Weidner (2020). The approach in that paper is to try to construct moment conditions that depend on  $(\gamma_1, \gamma_2)$ , but do not depend on the individual specific effects,  $\alpha_i$ . To do this, Honoré and Weidner (2020) follow the general approach in Bonhomme (2012). Bonhomme (2012) points out that models for discrete data generally cannot be dealt with using his approach. It is therefore “trial and error” to see whether it can be applied to models like (2.6).

To find a moment condition for  $(\gamma_1, \gamma_2)$  in (2.6) with  $T = 5$ , one needs to find functions,  $m$ , of the data and the parameters such that

$$E_{(\gamma_1, \gamma_2)} [m(y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}, \gamma_1, \gamma_2) | y_{i1}, y_{i2}, \alpha_i] = 0 \quad (2.7)$$

for all values of  $\alpha_i$ , and hence

$$E_{(\gamma_1, \gamma_2)} [m(y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}, \gamma_1, \gamma_2) | y_{i1}, y_{i2}] = 0$$

no matter what the true values of  $(\gamma_1, \gamma_2)$  are in the data generating process. The subscript  $(\gamma_1, \gamma_2)$  on the expectation is a reminder that the expectation is a function of  $\gamma_1$  and  $\gamma_2$ .

Since  $(y_{i3}, y_{i4}, y_{i5})$  can take eight values, (2.7) can be written as a sum over eight terms,

$$\begin{aligned} \sum_{(d_3, d_4, d_5) \in \{0,1\}^3} P_{(\gamma_1, \gamma_2)}(y_{i3} = d_3, y_{i4} = d_4, y_{i5} = d_5 | y_{i1} = d_1, y_{i2} = d_2, \alpha_i) \\ \times m(d_1, d_2, d_3, d_4, d_5, \gamma_1, \gamma_2) = 0. \end{aligned} \quad (2.8)$$

Honoré and Weidner (2020) approaches this and related problems by first fixing  $d_1, d_2, \gamma_1$ , and  $\gamma_2$  at a particular value and  $\alpha_i$  at  $q$  values for some  $q$ . At that point the probabilities are numbers, and the question becomes whether one can solve the  $q$  equations for the eight<sup>2</sup> unknown (the  $m$ 's) without making them all zero. If this is not possible, then there is no hope of finding an appropriate moment function,  $m$ .

After experimenting with various values for  $\gamma_1, \gamma_2$ , and the  $q$  values of  $\alpha_i$ , Honoré and Weidner (2020) conclude numerically that for each combination of the initial conditions, one can find a non-trivial moment condition. They obtain these analytically by solving (2.8) for a set of specific values of  $\alpha_i$  and then verifying that the obtained solution satisfies

<sup>2</sup>Actually, there are seven unknowns, since any multiple of a solution will also be a solution.

(2.8) generically. The moment conditions are

$$m_{(0,0)}(0, 0, d_3, d_4, d_5, \gamma) = \begin{cases} 1 & \text{if } (d_3, d_4, d_5) = (0, 1, 0), \\ e^{-\gamma_1} & \text{if } (d_3, d_4, d_5) = (0, 1, 1), \\ -1 & \text{if } (d_3, d_4) = (1, 0), \\ 0 & \text{otherwise,} \end{cases}$$

$$m_{(0,1)}(0, 1, d_3, d_4, d_5, \gamma) = \begin{cases} -1 & \text{if } (d_3, d_4) = (0, 1), \\ e^{\gamma_2 - \gamma_1} & \text{if } (d_3, d_4, d_5) = (1, 0, 0), \\ e^{\gamma_2} & \text{if } (d_3, d_4, d_5) = (1, 0, 1), \\ 0 & \text{otherwise,} \end{cases}$$

$$m_{(1,0)}(1, 0, d_3, d_4, d_5, y, \gamma) = \begin{cases} e^{\gamma_2} & \text{if } (d_3, d_4, d_5) = (0, 1, 0), \\ e^{\gamma_2 - \gamma_1} & \text{if } (d_3, d_4, d_5) = (0, 1, 1), \\ -1 & \text{if } (d_3, d_4) = (1, 0), \\ 0 & \text{otherwise,} \end{cases}$$

and

$$m_{(1,1)}(1, 1, d_3, d_4, d_5, \gamma) = \begin{cases} -1 & \text{if } (d_3, d_4) = (0, 1), \\ e^{-\gamma_1} & \text{if } (d_3, d_4, d_5) = (1, 0, 0), \\ 1 & \text{if } (d_3, d_4, d_5) = (1, 0, 1), \\ 0 & \text{otherwise.} \end{cases}$$

where the subscripts on  $m$  denote the initial values of  $y_1$  and  $y_2$ .

The functions  $m_{(0,0)}$  and  $m_{(1,1)}$  are both strictly monotone in  $\gamma_1$  if  $(d_3, d_4, d_5) = (0, 1, 1)$  or  $= (1, 0, 0)$ , respectively, and constant otherwise. It is therefore clear that as long as either  $P(y_{i1} = 0, y_{i2} = 0) > 0$  or  $P(y_{i1} = 1, y_{i2} = 1) > 0$ ,  $\gamma_1$  is identified from the moment conditions implied by either  $E[m_{(0,0)} | y_{i1} = 0, y_{i2} = 0]$  or  $E[m_{(1,1)} | y_{i1} = 1, y_{i2} = 1]$ .<sup>3</sup> Once  $\gamma_1$  has been identified,  $\gamma_2$  will be identified from  $E[m_{(0,1)} | y_{i1} = 0, y_{i2} = 1]$  or  $E[m_{(1,0)} | y_{i1} = 1, y_{i2} = 0]$  provided that either  $P(y_{i1} = 1, y_{i2} = 0) > 0$  or  $P(y_{i1} = 0, y_{i2} = 1) > 0$ .<sup>4</sup> In other words, if every combination of the initial conditions  $(y_{i1}, y_{i2})$  has positive probability,  $(\gamma_1, \gamma_2)$  is overidentified in the sense that there are four moment conditions (one corresponding to each of the initial conditions) and two parameters to be estimated. This partly solves the puzzle in Honoré and Kyriazidou (2019b) discussed above.

The strategy of looking for moment conditions developed in Bonhomme (2012) and explained above can be used for a number of other models. Honoré and Weidner (2020) present explicit expressions for such moment functions for AR( $p$ ) (for  $p = 1, 2$ , and 3) panel data logit models with strictly exogenous explanatory variables of the type

$$P(y_{it} = 1 | y_i^{t-1}, x_i^T, \alpha_i, \gamma_{i1}) = \frac{\exp\left(\sum_{j=1}^p \gamma_j y_{it-j} + x_{it}'\beta + \alpha_i\right)}{1 + \exp\left(\sum_{j=1}^p \gamma_j y_{it-j} + x_{it}'\beta + \alpha_i\right)}. \quad (2.9)$$

In this case, the moment functions will be functions of  $x_i^T$ , so the approach will yield

<sup>3</sup>Formally, this assumes that  $P(y_{i3} = 0, y_{i4} = 1, y_{i5} = 1 | y_{i1} = 1, y_{i2} = 1) > 0$  and/or  $P(y_{i3} = 1, y_{i4} = 0, y_{i5} = 0 | y_{i1} = 1, y_{i2} = 0) > 0$ . This will be true as long as  $\alpha_i$  take finite values with positive probability.

<sup>4</sup>This follows because  $m_{(0,1)}$  and  $m_{(1,0)}$  are both monotone in  $\gamma_2$ .

conditional moment conditions, which can be turned into unconditional moment conditions for the purpose of estimation. While the conditional moment restrictions in Honoré and Weidner (2020) will not always point-identify the common parameters,  $\beta$  and the  $\gamma_j$ 's, the paper presents conditions under which the conditional moments can be turned into a finite number of unconditional moments which do identify the common parameters. Generalized method of moments estimation will deliver a root- $n$  consistent and asymptotically normal estimator in that case.<sup>5</sup>

When  $T = 4$  and  $p = 1$ , the moment functions in Honoré and Weidner (2020) yield moment conditions which are transformations of moment conditions that had previously been discovered by Kitazawa (2013, 2016). To apply the moment conditions, one needs a total of  $T \geq 2 + 2p$  periods of observations. Of these, the first  $p$  correspond to the initial conditions, and one therefore only needs to observe the explanatory variables in the last  $2 + p$  periods. Based on numerical calculations for various combinations of  $T$  and  $p$ , Honoré and Weidner (2020) conjecture that for each of the  $2^p$  combinations of the initial conditions, there are  $2^{T-p} - (T + 1 - 2p) 2^p$  linearly independent conditional moment conditions. For example, for an AR(2) panel data logit model with 10 time periods (two of which would provide the initial conditions), there are 228 conditional moment conditions. Intuitively, this implies that the model contains a lot of information about the parameters. However, the large number of moments also implies that one should be careful about blindly applying generalized method of moments estimation.

### 2.3. Inequalities

As mentioned above, Chamberlain (2010) shows that in a binary response model of the form  $P(y_{it} = 1 | x_i^T, y_i^{t-1}, \alpha_i) = F(x'_{it}\beta + \alpha_i)$ , where  $F$  is a known cumulative distribution function, regular root- $n$  estimation of  $\beta$  is only possible if  $F$  is the logistic cumulative distribution function. This suggests that it is also not possible to construct root- $n$  consistent estimators for dynamic models like (2.9) if one deviates from the logit model. Of course, this does not imply that it is not possible to construct useful consistent estimators or informative bounds for the common parameters in non-logit models. In this subsection, we discuss some of the progress that the literature has made in this direction.

Consider the panel data discrete choice model

$$y_{it} = 1 \{x'_{it}\beta + \alpha_i + \varepsilon_{it} \geq 0\} \quad t = 1, 2; \quad i = 1, \dots, n \quad (2.10)$$

where, conditional on  $(x_{it}, x_{is}, \alpha_i)$ ,  $\varepsilon_{it}$  and  $\varepsilon_{is}$  are identically distributed with unknown distribution function  $F_{(x_{it}, x_{is}, \alpha_i)}$ . This is a strict exogeneity assumption on the explanatory variables and a stationarity assumption of the errors. When  $F$  is the logistic distribution, this is the logit model studied by Rasch (1960, 1961). See Equation (2.3). Manski (1987) observed that if  $F_{(x_{it}, x_{is}, \alpha_i)}$  has support equal to the real line, then this implies that

$$P(y_{it} = 1 | x_{it}, x_{is}) > P(y_{is} = 1 | x_{it}, x_{is}) \iff x'_{it}\beta > x'_{is}\beta. \quad (2.11)$$

The key property is that the left hand side does not depend on  $\alpha_i$  and can be identified from the data, while the right hand side is a constraint on  $\beta$ .

<sup>5</sup>Honoré and Kyriazidou (2000) provide conditions under which a conditional likelihood approach can be used to estimate models like (2.9). In order to achieve root- $n$  consistency, that approach requires that there is positive probability that the explanatory variables are the same in two time periods. The sufficient conditions in Honoré and Weidner (2020) are weaker than that.

Equation (2.11) allowed Manski (1987) to define a conditional maximum score estimator,

$$\hat{\beta} = \arg \max_b \sum_{i=1}^n \text{sgn}(y_{i2} - y_{i1}) \text{sgn}((x_{i2} - x_{i1})'b).$$

With random sampling and assumptions on the support of the explanatory variables, this estimator is consistent, but its rate of convergence is  $n^{-1/3}$ .<sup>6</sup>

Honoré and Kyriazidou (2000) use Manski's insight to construct an estimator for a version of (2.10) that also has a lagged  $y$  as an explanatory variable. They assume that the errors are independent and identically distributed - and not just stationary as in Manski (1987) - and that the researcher has access to a sample with at least four time periods for each individual.<sup>7</sup> In order to get point identification, Honoré and Kyriazidou (2000) had to make the strong assumption that the vector  $x_{i4} - x_{i3}$  has support in a neighborhood of 0.

Other papers have been able to obtain bounds without the assumption that  $x_{i4} - x_{i3}$  has support in a neighborhood of 0. For example, Aristodemou (2020) also considers a version of (2.10) that has a lagged  $y$  as well as strictly exogenous regressors as explanatory variables. Consider an individual for whom  $y_{it}$  is observed in three time periods. Aristodemou (2020) observes that if the errors in periods two and three are independent of the explanatory variables conditional on the initial  $y_{i1}$ , then

$$\begin{aligned} & P(y_{i2} = 1, y_{i3} = 0 | x_{i2}, x_{i3}, y_{i1} = 0) \\ &= P(x'_{i2}\beta + \alpha_i + \varepsilon_{i2} \geq 0, x'_{i3}\beta + \gamma + \alpha_i + \varepsilon_{i3} < 0 | x_{i2}, x_{i3}, y_{i1} = 0) \\ &\leq P((x_{i2} - x_{i3})'\beta + (y_{i1} - 1)\gamma > -(\varepsilon_{i2} - \varepsilon_{i3}) | x_{i2}, x_{i3}, y_{i1} = 0) \\ &= F_{\varepsilon_{i3} - \varepsilon_{i2} | y_{i1}}((x_{i2} - x_{i3})'\beta - \gamma) \end{aligned} \tag{2.12}$$

and

$$\begin{aligned} & 1 - P(y_{i2} = 0, y_{i3} = 1 | x_{i2}, x_{i3}, y_{i1} = 0) \\ &= 1 - P(x'_{i2}\beta + \alpha_i + \varepsilon_{i2} < 0, x'_{i3}\beta + \alpha_i + \varepsilon_{i3} \geq 0 | x_{i2}, x_{i3}, y_{i1} = 0) \\ &\geq 1 - P((x_{i2} - x_{i3})'\beta < -(\varepsilon_{i2} - \varepsilon_{i3}) | x_{i2}, x_{i3}, y_{i1} = 0) \\ &= F_{\varepsilon_{i3} - \varepsilon_{i2} | y_{i1}}((x_{i2} - x_{i3})'\beta). \end{aligned} \tag{2.13}$$

For simplicity, suppose that  $x_{it}$  is one-dimensional and that  $\beta$  is normalized to 1 without loss of generality. Then each value,  $w$ , in the support of  $x_{i2} - x_{i3}$ , (2.12) provides a lower bound on  $F_{\varepsilon_{i2} - \varepsilon_{i1} | y_{i0}}(w - \gamma)$  while (2.13) gives an upper bound on  $F_{\varepsilon_{i2} - \varepsilon_{i1} | y_{i0}}(w)$ . This gives a bound on  $\gamma$ .

More recently, Khan et al. (2020) characterize the identified region for  $(\gamma, \beta)$  under the weaker assumptions that the errors are stationary conditional on the sequence of explanatory variables and on the individual specific effect. Like Aristodemou (2020), this paper does not maintain the strong assumption on the explanatory variables needed by Honoré and Kyriazidou (2000). Interestingly, Khan et al. (2020) show that it is sometimes

<sup>6</sup>Imposing additional smoothness assumptions, Horowitz (1992) shows that one can improve the rate of convergence by defining a smoothed maximum score estimator as  $\hat{\beta} = \arg \max_b \sum_{i=1}^n \text{sgn}(y_{i2} - y_{i1}) H\left(\frac{(x_{i2} - x_{i1})'b}{h_n}\right)$ , where  $H$  is a cumulative distribution function which plays the same role as a kernel in nonparametric estimation.

<sup>7</sup>The first observation provides the initial condition. The model is not required to hold in this period.



possible to point-identify  $(\gamma, \beta)$  with as few as three time periods (including the initial condition).

### 3. BIVARIATE MODELS

As mentioned in Section 1, it is desirable to allow for predetermined – as opposed to strictly exogenous – explanatory variables in economic panel data settings. While doing this is an unsolved problem in general, it is possible to get results like those discussed above for a variety of models where a dependent variable and an explanatory variable are modelled jointly. We first illustrate this in a reduced form setting where two binary variables are modelled jointly. We next turn to a setting where they are the outcome of a simple game.

#### 3.1. Reduced Form Bivariate Models

Following Schmidt and Strauss (1975), who propose a cross-sectional bivariate binary response model, Honoré and Kyriazidou (2019a) considered the bivariate panel data model for two outcomes  $(y_{1,it}, y_{2,it})$

$$\begin{aligned} P(y_{1,it} = 1 | y_{2,it}, y_{1,i}^{t-1}, y_{2,i}^{t-1}, x_{1,i}^T, x_{2,i}^T, \alpha_{1,i}, \alpha_{2,i}) &= \Lambda(\alpha_{1,i} + x'_{1,it}\beta_1 + \rho y_{2,it}), \\ P(y_{2,it} = 1 | y_{1,it}, y_{1,i}^{t-1}, y_{2,i}^{t-1}, x_{1,i}^T, x_{2,i}^T, \alpha_{1,i}, \alpha_{2,i}) &= \Lambda(\alpha_{2,i} + x'_{2,it}\beta_2 + \rho y_{1,it}). \end{aligned} \quad (3.14)$$

Honoré and Kyriazidou (2019a) show that in this case  $\beta_1$ ,  $\beta_2$ , and  $\rho$  are identified with  $T = 2$ .

Honoré and Kyriazidou (2019a) also consider a vector autoregressive version of the simultaneous logit model in (3.14):

$$\begin{aligned} P(y_{1,it} = 1 | y_{2,it}, y_{1,i}^{t-1}, y_{2,i}^{t-1}, \alpha_{1,i}, \alpha_{2,i}) &= \Lambda(\alpha_{1,i} + y_{1,it-1}\gamma_{11} + y_{2,it-1}\gamma_{12} + \rho y_{2,it}), \\ P(y_{2,it} = 1 | y_{1,it}, y_{1,i}^{t-1}, y_{2,i}^{t-1}, \alpha_{1,i}, \alpha_{2,i}) &= \Lambda(\alpha_{2,i} + y_{1,it-1}\gamma_{21} + y_{2,it-1}\gamma_{22} + \rho y_{1,it}). \end{aligned} \quad (3.15)$$

When  $\rho = 0$ , this corresponds to the probabilities in the model proposed by Narendranthan et al. (1985):

$$\begin{aligned} y_{1,it} &= 1 \{y_{1,it-1}\gamma_{11} + y_{2,it-1}\gamma_{12} + \alpha_{1,i} + \varepsilon_{1,it} \geq 0\} \\ y_{2,it} &= 1 \{y_{1,it-1}\gamma_{21} + y_{2,it-1}\gamma_{22} + \alpha_{2,i} + \varepsilon_{2,it} \geq 0\}, \end{aligned}$$

where  $\varepsilon_{1,it}$  and  $\varepsilon_{2,it}$  are logistic random variables that are independent of each other and independent over time. Narendranthan et al. (1985) show that all parameters in this model are identified with a total of  $T = 4$  periods. Honoré and Kyriazidou (2019a) generalize this result by showing that  $(\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})$  is identified in the model given in (3.15) with at least four time periods.<sup>8</sup> However, the conditioning argument that leads to the identification eliminates the parameter  $\rho$  along with the heterogeneity terms  $\alpha_{1,i}$  and  $\alpha_{2,i}$ . On the positive side, this implies that one can allow the parameter  $\rho$  in (3.15) to be individual-specific. On the other hand,  $\rho$  may be the parameter of interest in many

<sup>8</sup>Honoré and Kyriazidou (2019a) also discuss how one can generalize the identification results in Narendranthan et al. (1985) and Honoré and Kyriazidou (2019b) to achieve identification if one also allows for strictly exogenous explanatory variables. The identification argument mimics that in Honoré and Kyriazidou (2000) and using the empirical counterpart for estimation will lead to estimators that converge at a rate slower than the usual  $\sqrt{n}$  if the strictly exogenous variables are continuously distributed.

applications. This makes it problematic that the conditioning argument eliminates it along with  $\alpha_{1,i}$  and  $\alpha_{2,i}$ . The calculations in Honoré and Kyriazidou (2019b) suggest that  $\rho$  might be identified, despite the fact that it drops out when one pursues a conditional likelihood approach to eliminate  $\alpha_{1,i}$  and  $\alpha_{2,i}$ . It would be interesting to know whether the results in Honoré and Weidner (2020) can be used to derive moment conditions that can be used to identify  $\rho$  the same way one can identify  $\gamma_1$  in (2.6).

### 3.2. Panel Data Games

The model in equations (3.14) and (3.15) is a natural generalization of classic linear simultaneous equations model to a logit framework. On the other hand, it is not straightforward to give a behavioral interpretation to the model. This is in contrast to single equation logit or probit models which can be interpreted in terms of threshold-crossing or utility maximization. We therefore turn to an alternative panel data version of the bivariate binary response models that are more inspired by economics.

Consider a game with two players  $i = 1, 2$ , each of whom takes a binary action,  $y \in \{0, 1\}$ , at instance  $t$  according to the best-response function:

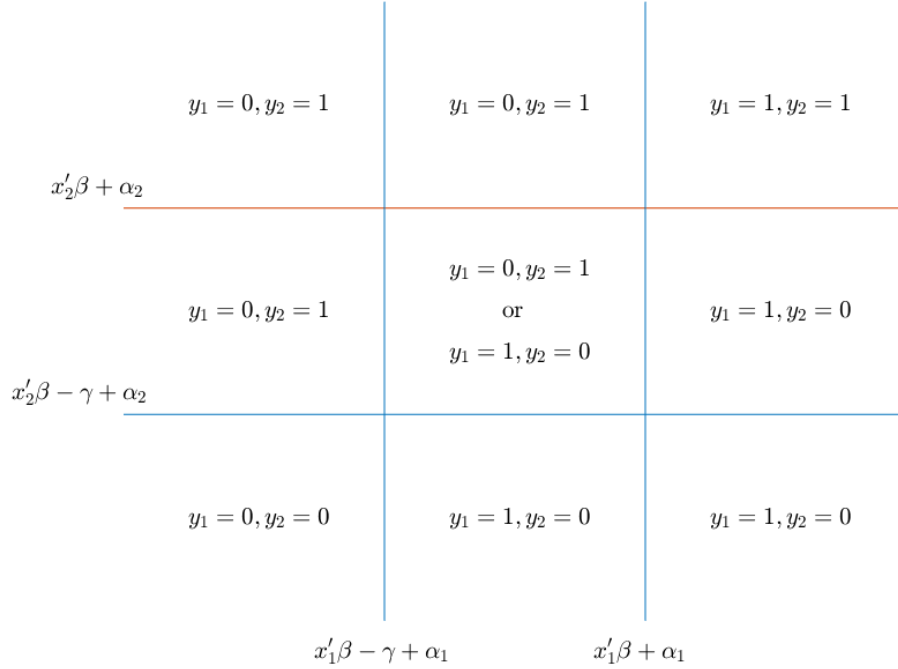
$$\begin{aligned} y_{1t} &= 1 \{x'_{1t}\beta - \gamma y_{2t} + \alpha_1 + \varepsilon_{1t} > 0\} \\ y_{2t} &= 1 \{x'_{2t}\beta - \gamma y_{1t} + \alpha_2 + \varepsilon_{2t} > 0\}, \end{aligned} \quad (3.16)$$

where  $\varepsilon_{1t}$  and  $\varepsilon_{2t}$  are error terms. Except for the  $\alpha$  terms, this is the canonical model considered in Tamer (2003). If players are firms contemplating their presence in a particular market, it is natural to assume that  $\gamma > 0$ . One can envision observing their entry decisions across different periods for the same market or over distinct geographic markets. Our aim is to study identification and estimation of  $\beta$  and  $\gamma$  in panel data versions of this model with the  $\alpha$ 's being firm-market specific effects.<sup>9</sup>

The econometric model above can also be seen as a dyadic network formation model defining *directed* connections between  $(i, j)$  pairs of individuals, households, firms, or countries. Here  $t$  indexes node pairs and  $y_{it}$  indicates whether person  $i$  sends a link to person  $j$ . The individual effect  $\alpha_i$  would in turn encode the ‘gregariousness’ of individual  $i$ . Charbonneau (2017), for example, considers such a model for directed networks with  $\gamma = 0$  and an additional individual effect for the ‘target’ node  $j$ , which can be interpreted as this node’s ‘attractiveness’. See also Graham (2017). A specification with  $\gamma < 0$  would in turn allow for  $i$  to have a tendency to reciprocate the link decision of its counterpart  $j$  (see footnote 5 in de Paula (2020)). For expositional ease, we nevertheless assume that  $\gamma \geq 0$  for the remainder of this section and refer to players as firms and the game as a market.

It is well understood that the model in (3.16) is incomplete in the sense that there is no unique mapping from  $(x_{1t}, x_{2t}, \alpha_1, \alpha_2, \varepsilon_{1t}, \varepsilon_{2t})$  to  $(y_{1t}, y_{2t})$ . For example, when  $\gamma > 0$ , certain realizations for  $(x_{1t}, x_{2t}, \alpha_1, \alpha_2, \varepsilon_{1t}, \varepsilon_{2t})$  are consistent with  $(y_{1t}, y_{2t}) = (1, 0)$  or  $(y_{1t}, y_{2t}) = (0, 1)$  as depicted in Figure 1, and this leads to difficulties in the conventional panel data manipulations discussed so far. It is easiest to explain our ideas in a setting where the errors are stationary, independent of  $(x_{1t}, x_{2t}, \alpha_1, \alpha_2)$ , and independent over

<sup>9</sup>If the same two firms interact repeatedly in the same market or across several different geographic markets, it is plausible that their actions are related through time in a dynamic game or across space in a larger game. We view the considerations here as starting points for the analysis of more complex dynamic or spatial games with individual effects, which remain an interesting avenue for research.

**Figure 1.** Nash Equilibria in  $-\epsilon_1 \times -\epsilon_2$  space

time, but the derivations below suggest that these assumptions can be relaxed considerably. Clearly, it would be interesting to allow for dynamics (i.e., lagged dependent variables) in the model. To illustrate the main idea, we will nonetheless abstract from dynamics, but combining the insights from the literature discussed previously (as well as the possibility of forward looking behavior discussed later) would be an important angle on which to expand the ideas delineated below.

### 3.3. Identification of $\beta$

In this subsection, we discuss the potential for identifying the  $\beta$  in (3.16) when the distribution of  $(\varepsilon_1, \varepsilon_2)$  left unspecified.

Letting  $N^t$  be the number of entrants in a market in period  $t$ , conventional calculations (see de Paula (2013)) deliver:

$$P\left(N^t = 2 \mid \{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_1, \alpha_2\right) = F(x'_{1t}\beta - \gamma + \alpha_1, x'_{2t}\beta - \gamma + \alpha_2),$$

where  $F$  is the cumulative distribution function of  $(-\varepsilon_1, -\varepsilon_2)$ . Note that the probability above is monotone in  $(x'_{1t}\beta, x'_{2t}\beta)$ .

Suppose there are two time periods or instances, and that  $x$  is market-specific, so

$x_{1t} = x_{2t} = x_t$ . Then

$$P\left(N^t = 2 \mid \{x_s\}_{s=1}^2, \alpha_1, \alpha_2\right) = F(x'_t\beta - \gamma + \alpha_1, x'_t\beta - \gamma + \alpha_2),$$

which is increasing in  $x'_t\beta$ . Consequently,

$$\begin{aligned} P\left(N^1 = 2 \mid \{x_s\}_{s=1}^2, \alpha_1, \alpha_2\right) &\gtrless P\left(N^2 = 2 \mid \{x_s\}_{s=1}^2, \alpha_1, \alpha_2\right) \\ \text{if (and only if) } &x'_1\beta \gtrless x'_2\beta. \end{aligned} \quad (3.17)$$

Now condition on the event that  $N^t$  equals 2 in exactly one of the two periods. A maximum score argument like that in Manski (1987) (see above) applied to the event  $N^1 = 2$  with  $x'_1 - x'_2$  as the explanatory variables can then be used to identify and estimate  $\beta$  (up to scale).<sup>10</sup>

More specifically, conditional on  $\{x_s\}_{s=1}^2, \alpha_1, \alpha_2$  and  $\mathbf{1}(N^2 = 2) \neq \mathbf{1}(N^1 = 2)$ , the variable  $\mathbf{1}(N^2 = 2) - \mathbf{1}(N^1 = 2)$  is a Bernoulli random variable with the median given by

$$\begin{aligned} &\text{sgn}\left(\frac{P(N^2 = 2, N^1 \neq 2 \mid \{x_s\}_{s=1}^2, \alpha_1, \alpha_2) - P(N^2 \neq 2, N^1 = 2 \mid \{x_s\}_{s=1}^2, \alpha_1, \alpha_2)}{P(\mathbf{1}(N^2 = 2) \neq \mathbf{1}(N^1 = 2) \mid \{x_s\}_{s=1}^2, \alpha_1, \alpha_2)}\right) \\ &= \text{sgn}\left(P(N^2 = 2, N^1 \neq 2 \mid \{x_s\}_{s=1}^2, \alpha_1, \alpha_2) - P(N^2 \neq 2, N^1 = 2 \mid \{x_s\}_{s=1}^2, \alpha_1, \alpha_2)\right) \\ &= \text{sgn}\left(P(N^2 = 2 \mid \{x_s\}_{s=1}^2, \alpha_1, \alpha_2) - P(N^1 = 2 \mid \{x_s\}_{s=1}^2, \alpha_1, \alpha_2)\right) \\ &= \text{sgn}((x_2 - x_1)'\beta). \end{aligned}$$

The last equality follows since  $P(N^1 = 2 \mid \{x_s\}_{s=1}^2, \alpha_1, \alpha_2) \gtrless P(N^2 = 2 \mid \{x_s\}_{s=1}^2, \alpha_1, \alpha_2)$  if (and only if)  $x'_1\beta \gtrless x'_2\beta$ . Then, under the assumptions delineated in Manski (1987), one obtains that  $\beta = \text{argmax}_b E[\text{sgn}((x_2 - x_1)'b)(\mathbf{1}(N^2 = 2) - \mathbf{1}(N^1 = 2))]$  as established in that paper and discussed previously. Note also that this will work even if  $\gamma$  is market- and/or player-specific. On the other hand, it is crucial for the argument that  $\beta$  is the same for the two players. Using a similar argument, we can also recover  $\beta$  by conditioning on the event  $N^1 = 0$  or  $N^2 = 0$ , but not both.

When the  $x$ s are not market-specific, we can use the same argument by conditioning on  $x_{21} = x_{22} = x_2$  (i.e., player 2 has the same  $x$  in two periods; needless to say, this assumes that  $x_{21} - x_{22}$  has support in a neighborhood around 0). In that case,

$$P\left(N^t = 2 \mid \{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_{1j}, \alpha_2\right) = F(x'_{1t}\beta - \gamma + \alpha_1, x'_2\beta - \gamma + \alpha_2),$$

so that

$$P\left(N^1 = 2 \mid \{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_1, \alpha_2\right) \gtrless P\left(N^2 = 2 \mid \{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_1, \alpha_2\right)$$

if (and only if)  $x'_{11}\beta \gtrless x'_{12}\beta$ . As a result, we can identify and estimate  $\beta$  by conditioning on markets where  $N^1 = 2$  or  $N^2 = 2$ , but not both.

<sup>10</sup>Since  $F$  is not specified, there is no scope for identifying the scale of  $(\beta, \gamma)$ .

3.4. Bounds on  $\gamma$ 

Even if the distribution of  $(\varepsilon_{1t}, \varepsilon_{2t})$  in Tamer (2003) is known, the model's incompleteness does not allow us to represent the probability distribution of  $(y_{1t}, y_{2t})$  conditional on  $(x_{1t}, x_{2t}, \alpha_1, \alpha_2)$  as a function of  $(\beta, \gamma)$  (see Figure 1). On the other hand, the model does provide bounds on the probabilities for each outcome as a function of  $(\beta, \gamma)$  (conditional on  $(x_{1t}, x_{2t}, \alpha_1, \alpha_2)$ ) and, analogously, bounds on the probabilities for each outcome as a function of  $(\beta, \gamma)$  conditional on  $(x_{1t}, x_{2t})$ . See, for example, Tamer (2003).

For a given period  $t$  we can establish that

$$\begin{aligned} P\left((y_{1t}, y_{2t}) = (1, 1) \mid \{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_1, \alpha_2\right) &= F(x'_{1t}\beta - \gamma + \alpha_1, x'_{2t}\beta - \gamma + \alpha_2), \\ P\left((y_{1t}, y_{2t}) = (0, 0) \mid \{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_1, \alpha_2\right) &= 1 - F(x'_{1t}\beta + \alpha_1) - F(x'_{2t}\beta + \alpha_2) \\ &\quad + F(x'_{1t}\beta + \alpha_1, x'_{2t}\beta + \alpha_2), \end{aligned}$$

whereas

$$\begin{aligned} &P\left((y_{1t}, y_{2t}) = (0, 1) \mid \{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_1, \alpha_2\right) \\ &\leq 1 - P\left((y_{1t}, y_{2t}) = (1, 1) \mid \{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_1, \alpha_2\right) \\ &\quad - P\left((y_{1t}, y_{2t}) = (0, 0) \mid \{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_1, \alpha_2\right) \end{aligned} \quad (3.18)$$

and

$$\begin{aligned} &P\left((y_{1t}, y_{2t}) = (0, 1) \mid \{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_1, \alpha_2\right) \\ &\geq 1 - P\left((y_{1t}, y_{2t}) = (1, 1) \mid \{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_1, \alpha_2\right) \\ &\quad - P\left((y_{1t}, y_{2t}) = (0, 0) \mid \{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_1, \alpha_2\right) \\ &\quad - P\left((\varepsilon_1, \varepsilon_2) \in \times_{i=1,2} [x'_{it}\beta - \gamma + \alpha_i, x'_{it}\beta + \alpha_i] \mid \{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_1, \alpha_2\right), \end{aligned} \quad (3.19)$$

and similarly for  $(y_{1t}, y_{2t}) = (1, 0)$ . Since  $(\varepsilon_{1t}, \varepsilon_{2t})$  is independent across time, one can thus obtain probability bounds on  $((y_{11}, y_{21}), (y_{12}, y_{22}))$  (conditional on  $(\{x_{1s}, x_{2s}\}_{s=1}^2, \alpha_1, \alpha_2)$ ) by taking products of the above bounds. Finally, to obtain probability bounds conditional on  $\{x_{1s}, x_{2s}\}_{s=1}^2$  one can integrate the above equalities and inequalities against the distribution  $H(\alpha_1, \alpha_2 \mid \{x_{1s}, x_{2s}\}_{s=1}^2)$  for  $\alpha_1$  and  $\alpha_2$ . An identified set for the unknown parameters<sup>11</sup> is then the set of parameters that is consistent with the above bounds (for some admissible distribution  $H(\alpha_1, \alpha_2 \mid \{x_{1s}, x_{2s}\}_{s=1}^2)$ ). For example,

$$\begin{aligned} P\left((y_{11}, y_{21}) = (1, 1) \text{ and } (y_{12}, y_{22}) = (1, 1) \mid \{x_{1s}, x_{2s}\}_{s=1}^2\right) &= \\ &\int F(x'_{11}\beta - \gamma + \alpha_1, x'_{21}\beta - \gamma + \alpha_2) \\ &\quad F(x'_{12}\beta - \gamma + \alpha_1, x'_{22}\beta - \gamma + \alpha_2) dH(\alpha_1, \alpha_2 \mid \{x_{1s}, x_{2s}\}_{s=1}^2) \end{aligned} \quad (3.20)$$

<sup>11</sup>If the data generating process satisfies the assumptions that one needs to apply maximum score above, then one needs only search over  $\gamma$ . If not, it is necessary to search over both  $\gamma$  and  $\beta$ .

Since the support of  $((y_{11}, y_{21}), (y_{12}, y_{22}))$  has sixteen points, there are thus two probability equalities (for the events  $((y_{11}, y_{21}), (y_{12}, y_{22})) = ((1, 1), (1, 1))$  and  $((y_{11}, y_{21}), (y_{12}, y_{22})) = ((0, 0), (0, 0))$ ) and twenty-eight inequalities, two for each of the remaining fourteen events (given covariates). To operationalize this, we would need to compute those restrictions across all possible distributions,  $H$ , for the individual specific (“fixed”) effects. If the data generating process satisfies the assumptions that one needs to apply maximum score above, then one only needs to bound  $\gamma$  using the inequalities in (3.18) and (3.19). If not, then one could combine the restrictions implied by the inequalities in (3.18) and (3.19) with, for example, (3.17), to obtain bounds for  $\beta$  and  $\gamma$ . The same approach can be used if the  $\gamma$ ’s are different for units one and two.

The approach for bounding the model parameters above can be combined with parametric assumptions on the distribution  $H$ . Since the model is not dynamic, this will not lead to potential internal inconsistencies. Alternatively, one can proceed more nonparametrically. For example, Honoré and Tamer (2006) approximate the distribution of the individual-specific effects by a discrete distribution with many points of support. To determine whether a particular parameter value belongs to the identified set, they use linear programming to check whether there exists a distribution of the individual-specific effects such that the probability distribution calculated from the econometric model matches the probability distribution in the data. This approach seems reasonable when the individual specific effect is one-dimensional and one does not need to condition on additional covariates. On the other hand, when that is not the case, the necessary number of support points is likely to be unreasonably large. Theorem 2.1 in Winkler (1988), on the other hand, implies that to match  $m$  probabilities (adding to 1), there is no loss of generality in considering discrete distributions with  $m + 1$  points of support. This or similar results have been used in statistics and econometrics (see, e.g., Lindsay (1995), Chernozhukov et al. (2013), d’Haultfoeuille and Rathelot (2017)). Winkler’s result suggests a hybrid algorithm where one searches over the location of the points of support using nonlinear methods and then solves for the implied probabilities using linear programming. The result, that there is no loss of generality in considering a discrete distribution for the unobservable, is similar to a result in Honoré and Lleras-Muney (2006), except that in that instance, the structure of the problem determined their location. Here, searching over those locations will be part of the computational challenge.

### 3.5. Generalizations

The argument above combines the setup in (3.16) with simple static panel data insights. In most economic applications, it will be important to also allow for dynamics. One may consider “non-structural (myopic) dynamics” as in, for example, Honoré and Kyriazidou (2000) as well as “structural dynamics” as in Aguirregabiria et al. (2020), where utility maximizing agents realize that their choice today has an effect on their utility tomorrow. In addition, one may explore the econometric consequences of restricting how the equilibrium selection mechanism evolves over time.

## 4. CONCLUSIONS

Much of the literature on nonlinear panel data models has been inspired by standard cross sectional models. Historically, these models have been made dynamic by including lagged dependent variables as explanatory variables. While this is natural in some settings, it

is important to recognize that it has implications if one wants to interpret the estimated model in terms of some implicit underlying economic model. For example, if one wants to motivate a logit model with lagged dependent variables in terms of a random utility model in which the utility of an option in one period depends on the choice in the previous period, then one typically implicitly rules out that the agents are forward looking. In a recent paper, Aguirregabiria et al. (2019) demonstrates that in a particular example, it is possible to adapt some of the conditioning arguments for logit models to more natural economic models. Investigating whether this is true for panel data discrete choice models with fixed effects more generally is an interesting topic for future research.

## REFERENCES

- Aguirregabiria, V., J. Gu, and Y. Luo (2019). Sufficient statistics for unobserved heterogeneity in dynamic structural logit models. Working Paper.
- Aguirregabiria, V., J. Gu, and Y. Luo (2020). Sufficient statistics for unobserved heterogeneity in structural dynamic logit models. *Journal of Econometrics*.
- Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society Series B*, 32, 283–301.
- Arellano, M. and S. Bonhomme (2009). Robust priors in nonlinear panel data models. *Econometrica* 77(2), 489–536.
- Aristodemou, E. (2020). Semiparametric identification in panel data discrete response models. *Journal of Econometrics*.
- Bonhomme, S. (2012). Functional differencing. *Econometrica* 80(4), 1337–1385.
- Bonhomme, S. and E. Manresa (2015). Grouped patterns of heterogeneity in panel data. *Econometrica* 83(3), 1147–1184.
- Bresnahan, T. and P. Reiss (1991). Empirical models of discrete games. *Journal of Econometrics* 48.
- Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In J. J. Heckman and B. Singer (Eds.), *Longitudinal Analysis of Labor Market Data*, Number 10 in Econometric Society Monographs series, pp. 3–38. Cambridge, New York and Sydney: Cambridge University Press.
- Chamberlain, G. (2010). Binary response models for panel data: Identification and information. *Econometrica* 78(1), 159–168.
- Charbonneau, K. B. (2017, October). Multiple fixed effects in binary response panel data models. *Econometrics Journal* 20(3), S1–S13.
- Chen, S., S. Khan, and X. Tang (2016). Informational content of special regressors in heteroskedastic binary response models. *Journal of Econometrics* 193(1), 162 – 182.
- Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013). Average and quantile effects in nonseparable panel models. *Econometrica* 81(2), 535–580.
- de Paula, A. (2013, January). Econometric analysis of games with multiple equilibria. *Annual Review of Economics* 5, 107–131.
- de Paula, A. (2020, August). Econometric models of network formation. *Annual Review of Economics* 12.
- Dhaene, G. and K. Jochmans (2015, 02). Split-panel Jackknife Estimation of Fixed-effect Models. *The Review of Economic Studies* 82(3), 991–1030.
- d’Haultfoeuille, X. and R. Rathelot (2017). Measuring segregation on small units: A partial identification analysis. *Quantitative Economics* 8, 39–73.
- Fernández-Val, I. and M. Weidner (2016). Individual and time effects in nonlinear panel models with large  $n$ ,  $t$ . *Journal of Econometrics* 192(1), 291 – 312.

- Fernández-Val, I. and M. Weidner (2018). Fixed effects estimation of large-tpanel data models. *Annual Review of Economics* 10(1), 109–138.
- Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica* 85(4), 1033–1063.
- Hahn, J. and W. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72(4), 1295–1319.
- Honore, B. E. and E. Kyriazidou (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica* 68, 839–874.
- Honore, B. E. and E. Kyriazidou (2019a). Chapter 8 - panel vector autoregressions with binary data. In M. Tsionas (Ed.), *Panel Data Econometrics*, pp. 197 – 223. Academic Press.
- Honore, B. E. and E. Kyriazidou (2019b). Identification in binary response panel data models: Is point-identification more common than we thought? *Annals of Economics and Statistics* (134), 207–226.
- Honore, B. E. and A. Lleras-Muney (2006). Bounds in competing risks models and the war on cancer. *Econometrica* 74(6), 1675–1698.
- Honore, B. E. and E. T. Tamer (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica* 74(3), 611–629.
- Honore, B. E. and M. Weidner (2020). Moment conditions for dynamic panel logit models with fixed effects. <https://arxiv.org/abs/2005.05942>.
- Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica* 60, 505–531.
- Johnson, E. G. (2004). Identification in discrete choice models with fixed effects.
- Khan, S., M. Ponomareva, and E. Tamer (2020). Identification of dynamic binary response models. Working Paper.
- Kitazawa, Y. (2013). Exploration of dynamic fixed effects logit models from a traditional angle. Technical report, No. 60, Kyushu Sangyo University Faculty of Economics.
- Kitazawa, Y. (2016). Root-n consistent estimations of time dummies for the dynamic fixed effects logit models: Monte carlo illustrations. Technical report, No. 72, Kyushu Sangyo University Faculty of Economics.
- Lindsay, B. (1995). Mixture models: Theory, geometry and applications. *NSF-CBMS Regional Conference Series in Probability and Statistics* 5, 1–163.
- Magnac, T. (2000). State dependence and unobserved heterogeneity in youth employment histories. *Economic Journal* 110, 805–837.
- Manski, C. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica* 55, 357–62.
- Narendranathan, W., S. Nickell, and D. Metcalf (1985). An investigation into the incidence and dynamic structure of sickness and unemployment in Britain, 1965-75. *Journal of the Royal Statistical Society. Series A (General)* 148(3), 254–267.
- Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica* 16, 1–32.
- Pakes, A. and J. Porter (2016, January). Moment inequalities for multinomial choice with fixed effects. Working Paper 21893, National Bureau of Economic Research.
- Pakes, A., J. Porter, M. Shepard, and S. Calder-Wang (2021). State dependence, unobserved heterogeneity, and health plan choices. Unpublished Working Paper In Progress.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Denmark's Pædagogiske Institut.



- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Contributions to Biology and Problems of Medicine*, Berkeley, Calif., pp. 321–333. University of California Press.
- Schmidt, P. and R. P. Strauss (1975). Estimation of models with jointly dependent qualitative variables: A simultaneous logit approach. *Econometrica* 43(4), 745–755.
- Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies* 70.
- Winkler, G. (1988). Extreme points of moment sets. *Mathematics of Operations Research* 13(4), 581–587.