Identification In Binary Response Panel Data Models: Is Point-Identification More Common Than We Thought?∗

Bo E. Honoré† Ekaterini Kyriazidou‡

January 2019

Abstract

This paper investigates identification in binary response models with panel data. Conditioning on sufficient statistics can sometimes lead to a conditional maximum likelihood approach that can be used to identify and estimate the parameters of interest in such models. Unfortunately it is often difficult or impossible to find such sufficient statistics, and even if it is possible, the approach sometimes leads to conditional likelihoods that do not depend on some interesting parameters. Using a range of different data generating processes, this paper calculates the identified regions for parameters in panel data logit AR(2) and logit VAR(1) model for which it is not known whether the parameters are identified or not. We find that identification might be more common than was previously thought, and that the identified regions for non-identified objects may be small enough to be empirically useful.

1 Introduction

We consider models where for individual $i$, a (possibly vector valued) binary dependent variable in time period $t$, $y_{it}$, is modelled parametrically as a function of past values of the dependent variable, current and past values of a vector of explanatory variables, $x_{it}$, and of a variable, $\alpha_i$, that captures unobserved heterogeneity

$$ y_{it} \sim f \left( \cdot | x_{it}^t, y_{it}^{i-1}, \alpha_i; \theta \right). $$

(1)

∗This research was supported by the Gregory C. Chow Econometric Research Program at Princeton University and by the National Science Foundation (Grant Number SES-1530741). We have benefitted from discussions with Geert Dhaene and from comments and suggestions from Stéphane Bonhomme, Laurent Davezies and an anonymous referee. An earlier version of this paper was titled “Identification in Binary Response Panel Data Models.”

†Department of Economics, Princeton University, Princeton, NJ 08544, USA (honor@princeton.edu)

‡Corresponding Author: Department of Economics, Athens University of Economics and Business, Athens 10434, Greece (kyria@aeub.gr).
Here $y_t^s$ and $x_t^s$ denote the collection of $x$'s and $y$'s from period 1 up to and including period $s$. Some papers invoke a strict exogeneity assumption, in which case the parametric model specifies the distribution of $y_{it}$ conditional on the entire time path of the explanatory variables

$$y_{it} \sim f \left( \cdot \mid x_{i}^{T}, y_{i}^{t-1}, \alpha_{i}; \theta \right).$$  

We are interested in identification of the vector of parameters, $\theta$, from the distribution of $y_{it}^T$ given $x_{i}^{T}$ for a small number of time periods, $T$, without making assumptions on the relationship between the heterogeneity term $\alpha_{i}$ and the explanatory variables, $x_{i}^{T}$. This is sometimes referred to as a “fixed effects” approach.

There are two issues that make identification of (1) and (2) difficult. The first issue is the presence of the unobserved heterogeneity term $\alpha_{i}$. The second occurs when (1) and (2) model $y_{it}$ as a function of lagged $y$'s. The issue there is that (1) and (2) typically do not specify the distribution of the initial observations of $y$.

In a linear model where $\alpha_{i}$ enters linearly, $\alpha_{i}$ can be differenced away, typically through first differencing or by estimating the model in deviations from means form. This approach is not possible when $\alpha_{i}$ does not enter linearly, and one needs to explore alternative approaches in that case. Treating the $\alpha_{i}$’s as parameters to be estimated sometimes leads to a concentrated likelihood than can be used to estimate $\theta$, but that occurs relatively rarely. An alternative approach is to a look for a sufficient statistic, $S_{i}$, for $\alpha_{i}$. If such a sufficient statistic exists, the distribution of $y_{i}$ conditional on ($S_{i}, x_{i}, \alpha_{i}$) does not depend on $\alpha_{i}$, but it may depend on $\theta$, in which case it can be used to identify $\theta$. The corresponding maximum conditional likelihood estimator was studied in Andersen (1970). Unfortunately it is often difficult, or known to be impossible, to find the sufficient statistic $S_{i}$. In some cases, it is possible to come up with an identification strategy that can be used to estimate $\theta$ without making assumptions on $\alpha_{i}$ while also being nonparametric in the distribution of unobservable structural errors. See Manski (1987), Honoré (1992), Kyriazidou (1997) and Honoré and Kyriazidou (2000) for examples of different approaches in simple limited dependent variables models. Unfortunately, these strategies are model-specific and there does not seem to be a general approach that always works. Bonhomme (2012) proposes a functional differencing approach that applies to certain likelihood models including the static logit models. However, his approach generally fails for models with binary outcomes.\(^1\)

In this paper, we use the strategy of Honoré and Tamer (2006) to investigate identification in two models of the type (1) or (2) in which it is not known how to identify $\theta$. The two models are a panel data logit

---

\(^1\)More recently identification has been considered in panel multinomial choice models with multidimensional heterogeneity: Pakes and Porter (2016) obtain partial identification of common coefficients in semiparametric static and dynamic discrete choice models with index structure through a set of conditional moment inequalities. Shi, Shum, and Song (2018) propose a new semi-parametric identification and estimation approach based on the notion of cyclic monotonicity which, after integrating out the individual effects, in the semiparametric static multinomial choice panel model leads to a collection of conditional moment inequalities that are linear in the common parameters of interest.
AR(2) model, where the explanatory variables are the last two realizations of the dependent variables, and a VAR(1) panel data version of the simultaneous logit model proposed by Schmidt and Strauss (1975). Our approach to identification is computational. We will consider simple data generating processes and then calculate the identified region for the parameter, $\theta$. In the specific data generating processes we consider, our calculations suggest that point identification might be more common than was previously thought and that the identified region for non-identified parameters might be small enough for it to be useful for empirical analysis. Furthermore, we apply the same approach to construct bounds for marginal effects in certain specifications of a panel data logit AR(2) model where the parameter vector is not point identified. Our calculations again illustrate that the identified region can be small enough to be empirically useful.

Point identification is usually taken to mean that there is a unique parameter value that matches the data for any DGP in a large set of allowed DGPs. The findings of the paper should be understood under the caveat that our calculations only shed light on identification for particular DGPs.

Identification analysis is carried at the population level. In practice, identified parameters or identified regions need to be estimated from a finite sample. When a model is partially identified, as is the case in some of the models we consider in this paper, statistical inference is particularly delicate. For a comprehensive review of inference in partially identified models see, for example, Canay and Shaikh (2017) or Bontemps and Magnac (2017).

The paper is organized as follows: In Sections 2 and 3 we consider identification of common parameters in univariate and multivariate panel logit models, respectively. In Section 4 we consider computation of marginal effects. Section 5 concludes.

2 Identification in Univariate Panel Data Logit Models

2.1 Static Panel Logit Models

Rasch (1960) considered a static panel data model of the standard logit model,

$$P \left( y_{it} = 1 | x_{it}^T, y_{i}^{t-1}, \alpha_i \right) = \frac{\exp \left( x_{it}' \beta + \alpha_i \right)}{1 + \exp \left( x_{it}' \beta + \alpha_i \right)} = \Lambda \left( x_{it}' \beta + \alpha_i \right).$$

Note that the probability statement in (3) conditions on past, present and future values of the explanatory variables. In other words, the explanatory variables are strictly exogenous. In this case, the distribution of $(y_{1}, \ldots, y_{iT})$ conditional on $(x_{1}, \ldots, x_{iT})$ and on $s_i = \sum_{t=1}^{T} y_{it}$ does not depend on $\alpha_i$, but for $T \geq 2$ it does depend on $\beta$. The corresponding conditional likelihood can therefore be used to identify and estimate $\beta$. 
2.2 AR(1) Logit Models

The explanatory variables in (3) are strictly exogenous. This rules out an AR(1) specification in which the explanatory variable in time period \(t\) is the outcome from the previous period:

\[
P\left(y_{it} = 1|y_{i,t-1}, \alpha_i\right) = \frac{\exp(y_{i,t-1}\gamma + \alpha_i)}{1 + \exp(y_{i,t-1}\gamma + \alpha_i)}.
\]

(4)

Let \(s_i = \sum_{t=1}^{T} y_{it}\). Cox (1958) and Chamberlain (1985) showed that conditional on \((y_{it}, s_i, y_{iT})\), the distribution of \((y_{i1}, \ldots, y_{iT})\) does not depend on \(\alpha_i\), but for \(T \geq 4\), it does depend on \(\gamma\). The corresponding conditional likelihood can therefore be used to identify and estimate \(\gamma\).

2.3 AR(2) Logit Models

Chamberlain (1985) considered an AR(2) version of the logit model in which the constant as well as the coefficient on the first lag of the dependent variable are individual specific:

\[
P\left(y_{it} = 1|y_{i,t-1}, \alpha_i, \gamma_{i1}\right) = \frac{\exp(\gamma_{i1}y_{i,t-1} + \gamma_{i2}y_{i,t-2} + \alpha_i)}{1 + \exp(\gamma_{i1}y_{i,t-1} + \gamma_{i2}y_{i,t-2} + \alpha_i)}.
\]

(5)

When the coefficient on \(y_{i,t-2}\) is 0, this corresponds to a Markov switching model where the transition out of each state is individual specific and unrestricted. Chamberlain (1985) showed that the sufficient statistic for \((\gamma_{i1}, \alpha_i)\) with \(T \geq 6\) is \((y_{i1}, y_{i2}, s_{i1}, s_{i11}, y_{i,T-1}, y_{iT})\), where \(s_{i1} = \sum_{t=1}^{T} y_{it}\) and \(s_{i11} = \sum_{t=2}^{T} y_{it}y_{i,t-1}\). (See also Magnac (2000).) Again, the corresponding conditional likelihood can be used to identify and estimate \(\gamma_2\). For example,

\[P\left(\{y_{it}\}_{t=1}^{6} = (1, 0, 1, 0, 0, 0) \mid \{y_{it}\}_{t=1}^{6} \in \{(1, 0, 1, 0, 0, 0) \cup (1, 0, 0, 1, 0, 0)\}, \alpha_i\right) = \frac{\exp(\gamma_2)}{1 + \exp(\gamma_2)}.
\]

This does not depend on \(\alpha_i\), so

\[P\left(\{y_{it}\}_{t=1}^{6} = (1, 0, 1, 0, 0, 0) \mid \{y_{it}\}_{t=1}^{6} \in \{(1, 0, 1, 0, 0, 0) \cup (1, 0, 0, 1, 0, 0)\}\right) = \frac{\exp(\gamma_2)}{1 + \exp(\gamma_2)},
\]

which can be used to estimate \(\gamma_2\).

In this section, we use the approach in Honoré and Tamer (2006) to ask what we can learn about the coefficient on \(y_{i,t-1}\) if we restrict it to be constant across individuals: \(\gamma_{i1} = \gamma_1\) for all \(i\). As far as we know, this is not known to be identified even if \(T = 6\).

2.3.1 Computational Strategy

We make no assumptions on the two initial observations.\(^2\) Denote them by \(d_1\) and \(d_2\). This means that the distribution of \(\alpha\) given \(y_1 = d_1\) and \(y_2 = d_2\) is unrestricted. We can identify the distribution of \((y_3, \ldots, y_T)\) given \((y_1 = d_1, y_2 = d_2)\) from the data. Denote this by \(p(d_3, \cdots, d_T | d_1, d_2)\).

\(^2\)Here and in all computational sections, we drop the individual subscript \(i\) for simplicity.
The identified region for $\gamma = (\gamma_1, \gamma_2)$ is the set of $g = (g_1, g_2)$ such that for every $(d_1, d_2)$ there exists a distribution for $\alpha$ conditional on the initial observations, $H(\alpha|y_1 = d_1, y_2 = d_2)$ such that

$$p(d_3, \cdots, d_T|d_1, d_2) = \int \Pi_g(d_3, \cdots, d_T|\alpha, d_1, d_2) dH(\alpha|y_1 = d_1, y_2 = d_2)$$

where $\Pi_g(d_3, \cdots, d_T|\alpha, d_1, d_2)$ is the conditional distribution for $(y_3, ..., y_T)$ given $(\alpha, y_1 = d_1, y_2 = d_2)$ implied by the model using the parameter value $g = (g_1, g_2)$.

$$\Pi_g(d_3, \cdots, d_T|\alpha, d_1, d_2) = \prod_{t=3}^{T} \Lambda (g_1 d_{t-1} + g_2 d_{t-2} + \alpha)^{d_t} (1 - \Lambda (g_1 d_{t-1} + g_2 d_{t-2} + \alpha))^{1-d_t}.$$ 

In practice, we approximate $H(\alpha|y_1 = d_1, y_2 = d_2)$ with a discrete distribution that puts mass $h_k(d_1, d_2)$ at $a_k$ for $k = 1, \cdots, K$, where $K$ is a large number. Equation (6) then becomes

$$p(d_3, \cdots, d_T|d_1, d_2) = \sum_{k=1}^{K} \Pi_g(d_3, \cdots, d_T|\alpha_k, d_1, d_2) h_k(d_1, d_2).$$

If, for a given $g$, one can find a probability distribution for each value of $(d_1, d_2)$ that satisfies (7), then that $g$ could be the true $(\gamma_1, \gamma_2)$. So that value of $g$ belongs to the identified set for $(\gamma_1, \gamma_2)$. If, for one of the values of $(d_1, d_2)$ it is not possible to solve (7) subject to the constraint that $\sum_{k=1}^{K} h_k(d_1, d_2) = 1$ and $h_k(d_1, d_2) \geq 0$ for $k = 1, \cdots, K$, then $g$ does not belong to the identified set.

To check whether (7) has a solution, we follow Honoré and Tamer (2006) who note that the problem is akin to the problem of checking whether the feasible set for a linear programming problem is empty. This leads to the following computational strategy.

Let $A_j$ $(j = 1, ..., 2^{T-2})$ denote all possible sequences of $d_3, \cdots, d_T$, and let

$$Q(g|d_1, d_2) = \min_{\{h_k, \{\nu_j\}\}} \sum_{k=1}^{2^{T-2}} \nu_k + \nu_0$$

subject to

$$\sum_{k=1}^{K} \Pi_g(A_j|\alpha_k, d_1, d_2) h_k + \nu_j = p(A_j|d_1, d_2),$$

$$\sum_{k=1}^{K} h_k + \nu_0 = 1,$$

$$\nu_j \geq 0,$$

$$h_k \geq 0.$$ 

Equation (7) is then satisfied subject to the constraint that $\sum_{k=1}^{K} h_k(d_1, d_2) = 1$ and $h_k(d_1, d_2) \geq 0$ for $k = 1, \cdots, K$ if, and only if, $Q(g|d_1, d_2) = 0$. The identified region for $(\gamma_1, \gamma_2)$ is therefore the set

$$\{g : Q(g) = Q(g|0, 0) + Q(g|1, 0) + Q(g|0, 1) + Q(g|1, 1) = 0\}.$$
2.3.2 Data Generating Processes and Results

We consider four combinations of \((\gamma_1, \gamma_2)\) and two underlying distributions of \(\alpha\) for each value of the initial conditions \((d_1, d_2)\). Since we are calculating all the probabilities rather than estimating them, we choose both distributions of \(\alpha\) to be discrete, but in the spirit of \(\alpha\) being unrestricted, we let the support of \(\alpha\) be large.

Let \(\alpha\) take values \((-3, -2.999, -2.998, ..., 3)\). The distribution \(G_1\) puts equal weight on each point while \(G_2\) puts probability proportional to \(\phi(u - 1) + \phi(u + 1)\) at a point \(u\). In words, \(G_1\) is essentially uniform while \(G_2\) is essentially a mixture of normals.

The four combinations of \((\gamma_1, \gamma_2)\) are \((-1, -1), (-1, 1), (1, -1), \) and \((1, 1)\). This gives a total of eight designs. Without modifications, these designs will lead to very different marginal distributions of \(y_t\). In order to make the results more comparable across designs, we therefore adjust \(\alpha\) (additively) so that the probability that \(y_0 = 1\) is approximately 0.6.

We start with \(T = 6\). Since \(\gamma_2\) is known to be identified in this case, we calculate the objective function, \(Q(g_1, \gamma_2)\), over a grid of \(g_1\). To do this we use the same points of support for \(\alpha\) as the data-generating process. Figures 1 to 4 show the results for each of the four combinations of parameter values. In the left hand side of the figures, the heterogeneity, \(\alpha\), is distributed according to the uniform distribution, while the right hand side of the figures display the results from the mixture of normals as explained above. Each plot displays the four functions \(Q(g_1, \gamma_2|0, 0)\), \(Q(g_1, \gamma_2|1, 0)\), \(Q(g_1, \gamma_2|0, 1)\), and \(Q(g_1, \gamma_2|1, 1)\) corresponding to the four possible initial conditions. The identified set for \(\gamma_1\) is the set of values for which the function equals 0.

Figures 1 to 4 strongly suggest that \(\gamma_1\) is point identified in each of the designs that we study. If this is indeed the case, then that would, to our knowledge, be a new result.

Before elaborating on the possibility that \(\gamma_1\) is point identified with \(T = 6\), we calculate the identified regions for the same designs in the case where \(T = 5\). We calculate the objective function, \(Q(g_1, g_2)\) over a grid of values of \(g_1 = \gamma_1 - 0.1, \gamma_1 - 0.999, \gamma_1 - 0.998, ..., \gamma_1 + 0.1\) and \(g_2 = \gamma_2 - 0.1, \gamma_2 - 0.999, \gamma_2 - 0.998, ..., \gamma_2 + 0.1\). In each of the cases, it turns out that the true \((\gamma_1, \gamma_2)\) is the only point on our grid at which the objective function is below \(10^{-5}\). This suggest that \((\gamma_1, \gamma_2)\) might be point identified with as few as 5 time periods.

We finally repeat the exercise with \(T = 4\). Here, we search over a grid with values \(g_1 = \gamma_1 - 1.00, \gamma_1 - 0.99, \gamma_1 - 0.98, ..., \gamma_1 + 1.00\) and \(g_2 = \gamma_2 - 1.00, \gamma_2 - 0.99, \gamma_2 - 0.98, ..., \gamma_2 + 1.00\). Figures 5 to 8 show the combinations of \(g_1\) and \(g_2\) for which the objective function, \(Q(g_1, g_2)\) is less than \(10^{-5}\) (in which case it is virtually always exactly 0). It is clear from Figures 5 to 8 that \((\gamma_1, \gamma_2)\) is not point-identified when \(T = 4\), but even in that case the identified region is small enough to be informative.

As mentioned above, we do not believe that the literature contains a result that implies that \((\gamma_1, \gamma_2)\) is
point-identified when \( T = 5 \) or even that \( \gamma_1 \) is point-identified when \( T = 6 \). We therefore investigate this further.

The calculations above are limited by the fact that we only consider a finite grid of values for \( g_1 \) and \( g_2 \). It is possible that the identified region for the parameter \((\gamma_1, \gamma_2)\) is not a singleton, but that it is so small that it only intersects with the grid of values of \((g_1, g_2)\) at the true \((\gamma_1, \gamma_2)\). To investigate this, we consider a much finer grid, where the points are separated by 0.0001. It is also possible that approximating the distribution of \( \alpha \) by a finite distribution that takes values \((-3.999, -2.998, ..., 3)\) is too limiting. To address this, we therefore allowed \( \alpha \) to take the values \((-4, -3.999, -3.998, ..., 4)\). Finally, to try to gain a deeper understanding of identification in the model, we decomposed \( Q(g) \) into its components, \( Q(g|0,0) \), \( Q(g|1,0) \), \( Q(g|0,1) \), and \( Q(g|1,1) \). The results are presented in Table 1 for the designs for which \( \alpha \) is uniformly distributed and in Table 2 for the designs where it is distributed according to a mixture of normals. In each design, we found the values of \((g_1, g_2)\) for which \( Q(g|0,0) \), \( Q(g|1,0) \), \( Q(g|0,1) \), and \( Q(g|1,1) \) are 0. The lower and upper bounds for \( \gamma_1 \) and \( \gamma_2 \) based on those are in the first four blocks of rows in the tables. The last block of rows gives the bounds based on \( Q(g) = Q(g|0,0) + Q(g|1,0) + Q(g|0,1) + Q(g|1,1) \).

Tables 1 and 2 are consistent with \((\gamma_1, \gamma_2)\) being point-identified. Moreover, they suggest that the identification of \( \gamma_2 \) comes from sequences where the initial observations are \((0, 0)\) and \((1, 1)\).

To further investigate whether the apparent identification is an artifact of the computational approach discussed in Section 2.3.1, we made two additional changes using the design with uniform heterogeneity and where both parameters are 1. This is one of the designs where the initial conditions \((0, 0)\) and \((1, 1)\) give identified intervals for \( \gamma_2 \) rather than a single point, so it is likely to be one where there is room for detecting non-identification overall. We first doubled the number of points of support for the heterogeneity by making the distance between points 0.0005. This made no difference to the numbers presented in the last row of Table 1. We also took the original points of support and shifted them by half of the distance between the points. The effect of this is that the set of discrete distributions in Section 2.3.1 does not cover the true distribution of the heterogeneity. This also has no effect on the last row of Table 1.

We find it surprising if \((\gamma_1, \gamma_2)\) is point identified with \( T = 5 \). Consider, for example, the case where the initial observations are \( d_1 = 1 \) and \( d_2 = 0 \). There are eight possible sequences \((d_3, d_4, d_5)\) with conditional probabilities given by

\[
P((y_3, y_4, y_5) = (0, 0, 0)| (y_1, y_2) = (1, 0), \alpha) = \frac{1}{1 + \exp(\alpha + \gamma_2)} \frac{1}{1 + \exp(\alpha)} \frac{1}{1 + \exp(\alpha)}
\]

\[
P((y_3, y_4, y_5) = (0, 0, 1)| (y_1, y_2) = (1, 0), \alpha) = \frac{1}{1 + \exp(\alpha + \gamma_2)} \frac{1}{1 + \exp(\alpha)} \frac{\exp(\alpha)}{1 + \exp(\alpha)}
\]

7
\[ P ((y_3, y_4, y_5) = (1, 0, 0) | (y_1, y_2) = (1, 0), \alpha) = \frac{\exp(\alpha + \gamma_2)}{1 + \exp(\alpha + \gamma_2)} \frac{1}{1 + \exp(\alpha + \gamma_1)} \frac{1}{1 + \exp(\alpha + \gamma_2)} \]

\[ P ((y_3, y_4, y_5) = (1, 0, 1) | (y_1, y_2) = (1, 0), \alpha) = \frac{\exp(\alpha + \gamma_2)}{1 + \exp(\alpha + \gamma_2)} \frac{1}{1 + \exp(\alpha + \gamma_1)} \frac{\exp(\alpha + \gamma_2)}{1 + \exp(\alpha + \gamma_2)} \]

\[ P ((y_3, y_4, y_5) = (0, 1, 0) | (y_1, y_2) = (1, 0), \alpha) = \frac{1}{1 + \exp(\alpha + \gamma_2)} \frac{\exp(\alpha)}{1 + \exp(\alpha)} \frac{1}{1 + \exp(\alpha + \gamma_1)} \]

\[ P ((y_3, y_4, y_5) = (0, 1, 1) | (y_1, y_2) = (1, 0), \alpha) = \frac{1}{1 + \exp(\alpha + \gamma_2)} \frac{\exp(\alpha)}{1 + \exp(\alpha)} \frac{\exp(\alpha + \gamma_1)}{1 + \exp(\alpha + \gamma_1)} \]

\[ P ((y_3, y_4, y_5) = (1, 1, 0) | (y_1, y_2) = (1, 0), \alpha) = \frac{\exp(\alpha + \gamma_2)}{1 + \exp(\alpha + \gamma_2)} \frac{\exp(\alpha + \gamma_1)}{1 + \exp(\alpha + \gamma_1)} \frac{1}{1 + \exp(\alpha + \gamma_2 + \gamma_1)} \]

\[ P ((y_3, y_4, y_5) = (1, 1, 1) | (y_1, y_2) = (1, 0), \alpha) = \frac{\exp(\alpha + \gamma_2)}{1 + \exp(\alpha + \gamma_2)} \frac{\exp(\alpha + \gamma_1)}{1 + \exp(\alpha + \gamma_1)} \frac{\exp(\alpha + \gamma_1 + \gamma_2)}{1 + \exp(\alpha + \gamma_1 + \gamma_2)} \]

Among these sequences, there are no two sequences such that the ratio of their probabilities depends on one or both of the parameters of interest, \( \gamma_1 \) and \( \gamma_2 \), but not on \( \alpha \). This implies that there are no sequences \( S_1 \) and \( S_2 \) such that \( P (S_1 | S_1 \text{ or } S_2) \) does not depend on the distribution of \( \alpha \) but still depends on \( \gamma_1 \) or \( \gamma_2 \). Of course, it is possible that conditional probabilities of more complicated events could work. It is also possible that a completely different line of reasoning could deliver identification.

The probability distribution of \((y_3, y_4, y_5)\) conditional on \((y_1, y_2) = (1, 0)\) does identify the sign of \( \gamma_1 \) and \( \gamma_2 \). To see that the sign of \( \gamma_1 \) is identified, compare the probabilities \( P \left((y_3, y_4, y_5) = (0, 0, 1) | (y_1, y_2) = (1, 0), \alpha \right) \) and \( P \left((y_3, y_4, y_5) = (0, 1, 0) | (y_1, y_2) = (1, 0), \alpha \right) \):

\[
\begin{align*}
P \left((y_3, y_4, y_5) = (0, 0, 1) | (y_1, y_2) = (1, 0), \alpha \right) & \leq P \left((y_3, y_4, y_5) = (0, 1, 0) | (y_1, y_2) = (1, 0), \alpha \right) \\
\frac{1}{1 + \exp(\alpha + \gamma_2)} \frac{1}{1 + \exp(\alpha + \gamma_1)} & \leq \frac{1}{1 + \exp(\alpha + \gamma_2)} \frac{1}{1 + \exp(\alpha + \gamma_1)} \\
\frac{1}{1 + \exp(\alpha + \gamma_2)} & \leq \frac{1}{1 + \exp(\alpha + \gamma_1)} \\
\gamma_1 & \leq 0
\end{align*}
\]
Since this holds for all \( \alpha \), the sign of \( \gamma_1 \) is identified from the sign of \( P \left( (y_3, y_4, y_5) = (0, 0, 1) \mid (y_1, y_2) = (1, 0) \right) - P \left( (y_3, y_4, y_5) = (0, 1, 0) \mid (y_1, y_2) = (1, 0) \right) \).

To identify the sign of \( \gamma_2 \) note that

\[
\frac{1}{1 + \exp(\alpha + \gamma_2)} \leq \frac{1}{1 + \exp(\alpha + \gamma_1)} \leq \frac{1 + \exp(\alpha + \gamma_2)}{1 + \exp(\alpha + \gamma_1)} \leq \frac{\exp(\alpha + \gamma_2)}{1 + \exp(\alpha + \gamma_2)} \leq (1 + \exp(\alpha)) \exp(\gamma_2) \leq \exp(\gamma_2) \leq 0 \leq \gamma_2.
\]

Since this holds for all \( \alpha \) the sign of \( \gamma_2 \) is identified from the sign of \( P \left( (y_3, y_4, y_5) = (0, 1, 0) \mid (y_1, y_2) = (1, 0) \right) - P \left( (y_3, y_4, y_5) = (1, 0, 0) \mid (y_1, y_2) = (1, 0) \right) \).

The fact that the signs are identified suggests that the identification problem gets more difficult for large values of \( \gamma_1 \) and \( \gamma_2 \). We therefore re-did the calculations that led to Tables 1 and 2 with the \( \gamma \)'s twice as large. The results are presented in Tables 3 and 4. The results in Tables 3 and 4 are consistent with those in Tables 1 and 2. \( \gamma_1 \) appears to be point identified from any of the possible initial conditions, while \( \gamma_2 \) is only identified from the probability distributions of \( (y_3, y_4, y_5) \) conditional on \( (y_1, y_2) = (1, 0) \) or on \( (y_1, y_2) = (0, 1) \).

In the designs where \( \gamma_1 = -\gamma_2 \), the distribution of \( y_t \) conditional on \( y_{t-1} = y_{t-2} = 0 \) and \( \alpha \), is the same as the distribution of \( y_t \) conditional on \( y_{t-1} = y_{t-2} = 1 \) and \( \alpha \). Even though we have not proved it, it is possible that this puts restrictions on the observed distributions (which are not conditional on \( \alpha \)), and that these restrictions have consequences for the identifiability of \( (\gamma_1, \gamma_2) \). Suppose, for example, that one could use the line of arguments above to show that the sign of \( \gamma_1 + \gamma_2 \) is identified, then identification in the cases where \( \gamma_1 = -\gamma_2 \) is not representative of the general identification issue. It is also possible that knife-edge results emerge when \( \gamma_1 = \gamma_2 \).

We therefore introduce a third set of data generating processes in which \( (\gamma_1, \gamma_2) \) take the values \((-1.5, -0.7), \ldots\)\(^3\)

---

\(^3\)The fact that the signs of \( \gamma_1 \) and \( \gamma_2 \) are identified, is also the reason why we do not use a design with \( \gamma_1 \) or \( \gamma_2 \) equal to 0.
(-1.5, 0.7), (1.5, -0.7), and (1.5, 0.7). The results are presented in Tables 5 and 6. In this case, the initial conditions (0, 1) and (1, 0) still seem to deliver point identification for all parameter combinations, but with initial conditions (0, 0) and (1, 1) now seem to not provide enough information to identify $\gamma_2$, even when $\gamma_1$ and $\gamma_2$ have opposite signs. This suggests that for these initial conditions, $\gamma_1 + \gamma_2 = 0$ is a special case.

As mentioned, we are not aware of any theoretical identification results for $(\gamma_1, \gamma_2)$ that apply to a version of (5) with homogeneous $\gamma_1$ and $T < 6$ (and no additional assumptions about the initial conditions). The numerical calculations presented in this section suggest that both $\gamma_1$ and $\gamma_2$ are identified with $T = 5$, but that neither is with $T = 4$. Moreover, it seems that $\gamma_1$ can be identified from any set of initial conditions, but that identification of $\gamma_2$ relies on the initial $y$’s being different from each other.

3 Identification in Multivariate Panel Data Logit Models

3.1 Simultaneous Static Panel Logit Models

Following the motivation in Schmidt and Strauss (1975) who considered cross sectional simultaneous logit models, Honoré and Kyriazidou (2018) consider the panel data version of the model with strictly exogenous covariates, which in the bivariate case may be written as

$$
P (y_{1, it} = 1 | y_{2, it}, y_{1, it-1}, x_{1, it}^T, x_{2, it}^T, \alpha_1, \alpha_2) = \Lambda (\alpha_1 + x_{1, it} \beta_1 + \rho_1 y_{2, it})
$$

and

$$
P (y_{2, it} = 1 | y_{1, it}, y_{1, it-1}, y_{2, it-1}, x_{1, it}^T, x_{2, it}^T, \alpha_1, \alpha_2) = \Lambda (\alpha_2 + x_{2, it} \beta_2 + \rho_2 y_{1, it}).
$$

Schmidt and Strauss (1975) show that $\rho_1$ must equal $\rho_2$, and Honoré and Kyriazidou (2018) show that in this case $\beta_1$, $\beta_2$ and $\rho$ are identified with $T = 2$.

3.2 Panel Logit Vector Autoregressions

Narendranthan, Nickell, and Metcalf (1985) introduced a first order vector autoregression version of the simple logit model with lagged dependent variables:

$$
y_{1, it} = 1 \{y_{1, it-1} \gamma_{11} + y_{2, it-1} \gamma_{12} + \alpha_{1, i} + \varepsilon_{1, it} \geq 0\}
$$

$$
y_{2, it} = 1 \{y_{1, it-1} \gamma_{21} + y_{2, it-1} \gamma_{22} + \alpha_{2, i} + \varepsilon_{2, it} \geq 0\},
$$
where \( \varepsilon_{1,it} \) and \( \varepsilon_{2,it} \) are logistic random variables that are independent of each other and independent over time. In this model

\[
P \left( y_{1,it} = d_1, y_{2,it} = d_2 \middle| y_{1,i}^{t-1}, y_{2,i}^{t-1}, \alpha_{1,i}, \alpha_{2,i} \right) = \Lambda \left( y_{1,it-1} \gamma_{11} + y_{2,it-1} \gamma_{12} + \alpha_{1,i} \right)^{d_1} \left( 1 - \Lambda \left( y_{1,it-1} \gamma_{11} + y_{2,it-1} \gamma_{12} + \alpha_{1,i} \right) \right)^{1-d_1}
\]

\[
P \left( y_{1,it} = 1 \middle| y_{2,it}, y_{1,i}^{t-1}, y_{2,i}^{t-1}, \alpha_{1,i}, \alpha_{2,i} \right) = \Lambda \left( \alpha_{1,i} + y_{1,it-1} \gamma_{11} + y_{2,it-1} \gamma_{12} + \rho y_{2,it} \right)
\]

and

\[
P \left( y_{2,it} = 1 \middle| y_{1,it}, y_{1,i}^{t-1}, y_{2,i}^{t-1}, \alpha_{1,i}, \alpha_{2,i} \right) = \Lambda \left( \alpha_{2,i} + y_{1,it-1} \gamma_{21} + y_{2,it-1} \gamma_{22} + \rho y_{1,it} \right)
\]

which implies that

\[
P \left( y_{1,it} = d_1, y_{2,it} = d_2 \middle| y_{1,i}^{t-1}, y_{2,i}^{t-1}, \alpha_{1,i}, \alpha_{2,i} \right) = \frac{\exp(d_1 c_{1,i} + d_2 c_{2,i} + d_1 d_2 \rho)}{1 + \exp(c_{1,i}) + \exp(c_{2,i}) + \exp(c_{1,i} + c_{2,i} + \rho)}
\]

where as above \( c_{1,i} = \alpha_{1,i} + y_{1,it-1} \gamma_{11} + y_{2,it-1} \gamma_{12} \) and \( c_{2,i} = \alpha_{2,i} + y_{1,it-1} \gamma_{21} + y_{2,it-1} \gamma_{22} \). When \( \rho = 0 \), this corresponds to the probabilities in (11).

Honorable and Kyriazidou (2018) prove that in the dynamic model considered here, \( (\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}) \) is identified with at least four time periods. However, the conditioning argument they use and which leads to identification of \( (\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}) \), eliminates the parameter \( \rho \) along with the heterogeneity terms \( \alpha_{1,i} \) and \( \alpha_{2,i} \). Here we address the question of what one can say about \( \rho \) for some specific data generating processes.

We will use the same methodology as in Section 2.3.\(^4\) We start by specifying a data generating process and then use that to calculate the probability of all possible sequences. For a grid of parameter values, we then solve the linear programming problem (8) to determine the parameter values that are consistent with the probability distribution of the data conditional on the initial observations.

We first consider a data generating process where the unobserved heterogeneity, \( (\alpha_{1}, \alpha_{2}) \), is uniformly distributed on the grid \((-3.0, -2.9, \ldots, 2.9, 3.0) \times (-3.0, -2.9, \ldots, 2.9, 3.0)\) conditional on \((y_{1,1}, y_{2,1}), (\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}) =

\(^4\)Again for simplicity, we drop the \( i \) notation in the remainder of the section.
The left hand side of Figure 9 shows the objective function in (8) for the four different initial values of \((y_{1,1}, y_{2,1})\) with \(T = 4\). Since \((\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})\) is identified in this case, we display the objective function as a function of \(\rho\) alone. It appears from the graph that \(\rho\) is point identified with \(T = 4\) conditional on any initial value of \((y_{1,1}, y_{2,1})\). The right hand side of Figure 9 shows the same graph for \(T = 3\). Close inspection reveals that the region for which the objective function is 0, is \([0.669, 0.728]\). In this case, it is not known whether \((\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})\) is identified. The interval \([0.669, 0.728]\) is the identified set that one would have if one knew \((\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})\). As a result it provides an optimistic view on the identifiability of \(\rho\).

We find it surprising that \(\rho\) seems to be point identified with four time periods. A conditional likelihood approach would lead one to compare two mutually exclusive sequences of \(\{(y_{1,t}, y_{2,t})\}_{t=1}^{4}\). The hope would then be that the probability of one of those conditional on the union and on the heterogeneity terms \((\alpha_1, \alpha_2)\), will depend on \(\rho\) (and possibly \((\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})\)) but not on the heterogeneity terms. Since this approach works off the conditional probability of one sequence conditional on a pair of sequences, one would calculate the ratio of the probabilities of the two sequences. Denoting the two sequences \(\{(b_{1t}, b_{2t})\}_{t=1}^{4}\) and \(\{(c_{1t}, c_{2t})\}_{t=1}^{4}\), the ratio of the probabilities would be

\[
\frac{P\left(\{(y_{1,t}, y_{2,t})\}_{t=1}^{4} \mid (b_{1t}, b_{2t})\}_{t=1}^{4} \mid \alpha_1, \alpha_2\right)}{P\left(\{(y_{1,t}, y_{2,t})\}_{t=1}^{4} \mid (c_{1t}, c_{2t})\}_{t=1}^{4} \mid \alpha_1, \alpha_2\right)}
\]

\[= \frac{p\left(y_{1,1} = b_{11}, y_{2,1} = b_{21} \mid \alpha_1, \alpha_2\right) p\left(y_{1,1} = c_{11}, y_{2,1} = c_{21} \mid \alpha_1, \alpha_2\right)}{\prod_{t=2}^{4} p\left(y_{1,t} = c_{1t}, y_{2,t} = c_{2t} \mid y_{1,t-1} = c_{1t-1}, y_{2,t-1} = c_{2t-1}, \alpha_1, \alpha_2\right)}
\]

\[= \frac{p\left(y_{1,1} = b_{11}, y_{2,1} = b_{21} \mid \alpha_1, \alpha_2\right) P_b}{p\left(y_{1,1} = c_{11}, y_{2,1} = c_{21} \mid \alpha_1, \alpha_2\right) P_c}
\]

where the terms \(P_b\) and \(P_c\) are given by (14).

The model does not specify the marginal probability distribution for the initial observations \((y_{1,1}, y_{2,1})\). This is the initial conditions problem, and it means that the function \(p\) in (15) is unknown. Hence \((b_{11}, b_{21})\) must be the same as \((c_{11}, c_{21})\) for \((\alpha_1, \alpha_2)\) to cancel in (15). Both \(P_b\) and \(P_c\) in (15) are the product of three ratios. Since \((b_{11}, b_{21}) = (c_{11}, c_{21})\), the first denominators (i.e. the denominators of the probabilities of the

\((-0.90, 0.80, 0.65, 0.50)\) and \(\rho = 0.70\). We have again shifted \((\alpha_1, \alpha_2)\) additively to make the marginal probabilities that \(y_{1,4}\) and \(y_{2,4}\) are 1 approximately 0.6.

The left hand side of Figure 9 shows the objective function in (8) for the four different initial values of \((y_{1,1}, y_{2,1})\) with \(T = 4\). Since \((\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})\) is identified in this case, we display the objective function as a function of \(\rho\) alone. It appears from the graph that \(\rho\) is point identified with \(T = 4\) conditional on any initial value of \((y_{1,1}, y_{2,1})\). The right hand side of Figure 9 shows the same graph for \(T = 3\). Close inspection reveals that the region for which the objective function is 0, is \([0.669, 0.728]\). In this case, it is not known whether \((\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})\) is identified. The interval \([0.669, 0.728]\) is the identified set that one would have if one knew \((\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})\). As a result it provides an optimistic view on the identifiability of \(\rho\).

We find it surprising that \(\rho\) seems to be point identified with four time periods. A conditional likelihood approach would lead one to compare two mutually exclusive sequences of \(\{(y_{1,t}, y_{2,t})\}_{t=1}^{4}\). The hope would then be that the probability of one of those conditional on the union and on the heterogeneity terms \((\alpha_1, \alpha_2)\), will depend on \(\rho\) (and possibly \((\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})\)) but not on the heterogeneity terms. Since this approach works off the conditional probability of one sequence conditional on a pair of sequences, one would calculate the ratio of the probabilities of the two sequences. Denoting the two sequences \(\{(b_{1t}, b_{2t})\}_{t=1}^{4}\) and \(\{(c_{1t}, c_{2t})\}_{t=1}^{4}\), the ratio of the probabilities would be

\[
\frac{P\left(\{(y_{1,t}, y_{2,t})\}_{t=1}^{4} \mid (b_{1t}, b_{2t})\}_{t=1}^{4} \mid \alpha_1, \alpha_2\right)}{P\left(\{(y_{1,t}, y_{2,t})\}_{t=1}^{4} \mid (c_{1t}, c_{2t})\}_{t=1}^{4} \mid \alpha_1, \alpha_2\right)}
\]

\[= \frac{p\left(y_{1,1} = b_{11}, y_{2,1} = b_{21} \mid \alpha_1, \alpha_2\right) p\left(y_{1,1} = c_{11}, y_{2,1} = c_{21} \mid \alpha_1, \alpha_2\right)}{\prod_{t=2}^{4} p\left(y_{1,t} = c_{1t}, y_{2,t} = c_{2t} \mid y_{1,t-1} = c_{1t-1}, y_{2,t-1} = c_{2t-1}, \alpha_1, \alpha_2\right)}
\]

\[= \frac{p\left(y_{1,1} = b_{11}, y_{2,1} = b_{21} \mid \alpha_1, \alpha_2\right) P_b}{p\left(y_{1,1} = c_{11}, y_{2,1} = c_{21} \mid \alpha_1, \alpha_2\right) P_c}
\]

where the terms \(P_b\) and \(P_c\) are given by (14).

The model does not specify the marginal probability distribution for the initial observations \((y_{1,1}, y_{2,1})\). This is the initial conditions problem, and it means that the function \(p\) in (15) is unknown. Hence \((b_{11}, b_{21})\) must be the same as \((c_{11}, c_{21})\) for \((\alpha_1, \alpha_2)\) to cancel in (15). Both \(P_b\) and \(P_c\) in (15) are the product of three ratios. Since \((b_{11}, b_{21}) = (c_{11}, c_{21})\), the first denominators (i.e. the denominators of the probabilities of the
second period) are the same and hence cancel. The ratio of the remaining denominators of \( P_b \) and \( P_c \) is

\[
\frac{1 + \exp (\alpha_1 + b_\rho' \gamma_1) + \exp (\alpha_2 + b_\rho' \gamma_2) + \exp (\alpha_1 + \alpha_2 + b_\rho' \gamma_2 + \rho)}{1 + \exp (\alpha_1 + c_\rho' \gamma_1) + \exp (\alpha_2 + c_\rho' \gamma_2) + \exp (\alpha_1 + c_\rho' \gamma_1 + \alpha_2 + c_\rho' \gamma_2 + \rho)}
\]

\[
\frac{1 + \exp (\alpha_1 + b_\rho' \gamma_1) + \exp (\alpha_2 + b_\rho' \gamma_2) + \exp (\alpha_1 + c_\rho' \gamma_1 + \alpha_2 + c_\rho' \gamma_2 + \rho)}{1 + \exp (\alpha_1 + c_\rho' \gamma_1) + \exp (\alpha_2 + c_\rho' \gamma_2) + \exp (\alpha_1 + c_\rho' \gamma_1 + \alpha_2 + c_\rho' \gamma_2 + \rho)}
\]

where \( a_1 = \exp (\alpha_1), a_2 = \exp (\alpha_2), \gamma_1 = (\gamma_{11}, \gamma_{12})', \gamma_2 = (\gamma_{21}, \gamma_{22})', b_t = (b_{1t}, b_{2t})' \) and \( c_t = (c_{1t}, c_{2t})' \).

For the unobserved heterogeneity to cancel from the conditional likelihood, the \( a \)'s must cancel from this expression. After performing the calculations in the expression above (too cumbersome to include here), we obtain terms involving \( a_1^2 \) and \( a_2^2 \). For them to cancel, we must have \( b_2 + b_3 = c_2 + c_3 \). On the other hand, for the heterogeneity to cancel from the numerators of \( P_b \) and \( P_c \), \( b_2 + b_3 + b_4 \) must equal \( c_2 + c_3 + c_4 \). Hence \( c_4 = b_4 \).

Recall that the aim is to identify \( \rho \). If it is indeed possible to cancel the effect of the \( a \)'s in the denominators of \( P_b \) and \( P_c \) as discussed above, then the resulting conditional likelihood can only identify \( \rho \) if \( \rho \) appears in the ratio of the numerators of \( P_b \) and \( P_c \). The numerators of these probabilities have the form \( \exp(d_1c_1 + d_2c_2 + d_1d_2\rho) \) (see (14)), so \( \rho \) only appears for periods where one observes \( (1, 1) \). This means that for \( \rho \) to not cancel in the ratio of the numerators of \( P_b \) and \( P_c \), there must be a different number of occurrences of \((1, 1)\) in the \( b \) and \( c \) sequences. Therefore (say) \( b \) must have \((1, 1)\) and \((0, 0)\) in periods 2 and 3, while \( c \) has \((1, 0)\) and \((0, 1)\). With this, (16) becomes

\[
\frac{1 + a_1 \exp (\gamma_{11} + \gamma_{12}) + a_2 \exp (\gamma_{21} + \gamma_{22}) + a_1a_2 \exp (\gamma_{11} + \gamma_{12} + \gamma_{21} + \gamma_{22} + \rho)}{1 + a_1 \exp (\gamma_{11}) + a_2 \exp (\gamma_{21}) + a_1a_2 \exp (\gamma_{11} + \gamma_{21} + \rho)}
\]

\[
\frac{1 + a_1 + a_2 + a_1a_2 \exp (\rho)}{1 + a_1 \exp (\gamma_{12}) + a_2 \exp (\gamma_{22}) + a_1a_2 \exp (\gamma_{12} + \gamma_{22} + \rho)}
\]

The coefficient on \( a_1a_2 \) in the numerator is

\[
\exp (\gamma_{11} + \gamma_{12} + \rho) + \exp (\gamma_{11} + \gamma_{12} + \gamma_{21} + \gamma_{22} + \rho) = \exp (\gamma_{11} + \gamma_{12} + \rho) \left( 1 + \exp (\gamma_{21} + \gamma_{22} + \rho) \right)
\]

while in the denominator it is

\[
\exp (\gamma_{11} + \gamma_{12} + \gamma_{22} + \rho) + \exp (\gamma_{11} + \gamma_{21} + \gamma_{12} + \rho) = \exp (\gamma_{11} + \gamma_{12} + \rho) \left( \exp (\gamma_{22}) + \exp (\gamma_{21} + \rho) \right)
\]

Since these are not equal, the unobserved heterogeneity does not cancel from the conditional likelihood. This is why we find it surprising that \( \rho \) appears to be identified (for the data generation considered here, at least).
Of course, it is possible that additional cancellations would occur for particular parameter configurations. This is why we chose the rather “unnatural” values of $\gamma$ in the calculation above. If some of the parameters had been the same, or had added up to 0, then the identified region for that design might not be representative of identification in general. It is also possible that $\rho$ happens to be identified because the $\alpha$’s are independent in the design above. We therefore turn to designs with dependent $\alpha$’s.

We consider one design where $\alpha_1$ and $\alpha_2$ are positively correlated and one where they are negatively correlated. Specifically, in the second and third data generating process, the probability at each value of $(\alpha_1, \alpha_2)$ is proportional to $\phi(u_1 - 1) \phi(u_2 - 1) + \phi(u_1 + 1) \phi(u_2 + 1)$ and $\phi(u_1 - 1) \phi(u_2 + 1) + \phi(u_1 + 1) \phi(u_2 - 1)$, respectively. Using the same parameter values for $\gamma$ and $\rho$ as above, the results are presented in Figures 10 and 11. As was the case for the first data generating process, $\rho$ appears to be identified in this case when $T = 4$. The identified regions that one would have with three time periods and known $\gamma$’s are $[0.680, 0.718]$ and $[0.682, 0.717]$, respectively.

4 Marginal Effects

Fixed effects approaches to nonlinear panel data models are often criticized on the grounds that they do not deliver marginal effects or counterfactual distributions. The reason is that even if the parameters of the model are identified, the underlying distribution of the individual heterogeneity cannot be, since the data will consist of a finite number of probabilities.

In this section, we demonstrate that the approach taken here can also be used to construct bounds for marginal effects. To simplify the exposition, we recycle the setup in Section 2.3.2.

Consider the AR(2) model given in (2.3.2) and suppose that the econometrician has access to the distribution of $(y_T, \ldots, y_T)$ conditional on $(y_1, y_2)$ as well as the probability distribution over the initial conditions, $(y_1, y_2)$. For concreteness, suppose we have a panel of length $T$, and we are interested in the difference in the probabilities that $y_{T+3}$ equals 1 depending on whether one “externally” sets $y_{T+1} = y_{T+2} = 0$ or $y_{T+1} = y_{T+2} = 1$. The difference in these probabilities is $E \left[ \frac{\exp(\gamma_1 + \gamma_2 + \alpha)}{1 + \exp(\gamma_1 + \gamma_2 + \alpha)} - \frac{\exp(\alpha)}{1 + \exp(\alpha)} \right]$, and it will depend on the parameters of the model, $(\gamma_1, \gamma_2)$, as well as on the distribution of the unobservable heterogeneity, $\alpha$.

As before, the identified region for $(\gamma_1, \gamma_2)$, $\mathcal{G}$, is the set of $g = (g_1, g_2)$ such that $Q(g) = 0$, where $Q(g) = Q(g|0,0) + Q(g|1,0) + Q(g|0,1) + Q(g|1,1)$ and $Q(g|d_1, d_2)$ is defined in the linear programming problem in (8). For a given $g$ in the identified set, and for a given probability distribution of $\alpha$, the difference

---

5Identification and estimation of such effects has been the focus of a number of recent papers such as Chernozhukov, Fernández-Val, Hahn, and Newey (2013) and Torgovitsky (2016).
in the probabilities above is

\[
E \left[ \frac{\exp (g_1 + g_2 + \alpha)}{1 + \exp (g_1 + g_2 + \alpha)} - \frac{\exp (\alpha)}{1 + \exp (\alpha)} \right],
\]

where the expectation is with respect to the distribution of \(a\).

For each \((g_1, g_2)\) in \(G\), a linear programming problem like (8) can be used to bound (17) provided that, as above, one is willing to approximate the distribution of \(\alpha\) by a discrete distribution with a large number of points of support. These bounds will have the form of bounds conditional on the initial conditions,\(^6\) \((y_1, y_2)\). Specifically, if the initial conditions are \((y_1, y_2) = (d_1, d_2)\) then the lower bound using parameters \(g = (g_1, g_2)\) in \(G\), is

\[
P_L (g|d_1, d_2) = \min_{\{h_k\}} \sum_{k=1}^{K} \left( \frac{\exp (g_1 + g_2 + \alpha)}{1 + \exp (g_1 + g_2 + \alpha)} - \frac{\exp (\alpha)}{1 + \exp (\alpha)} \right) h_k
\]

subject to

\[
\sum_{k=1}^{K} \Pi_g (\alpha_k; A_j|d_1, d_2) h_k = p (A_j|d_1, d_2) \\
\sum_{k=1}^{K} h_k = 1 \\
h_k \geq 0
\]

where \(p (A_j|d_1, d_2)\) is the true probability that \((y_3, y_4, ..., y_T) = A_j\) conditional on \((y_1, y_2) = (d_1, d_2)\) and \(\Pi_g (\alpha_k; A_j|d_1, d_2)\) is the same probability according to the model (conditional on \(\alpha = \alpha_k\)) if the parameter is \(g\). Let \(P_U (g|d_1, d_2)\) be the corresponding maximum.

The identified region for the expected value of \(\frac{\exp (\gamma_1 + \gamma_2 + \alpha)}{1 + \exp (\gamma_1 + \gamma_2 + \alpha)} - \frac{\exp (\alpha)}{1 + \exp (\alpha)}\) conditional on \((y_1, y_2) = (d_1, d_2)\) is therefore [\(\min_{g \in G} P_L (g|d_1, d_2)\), \(\max_{g \in G} P_U (g|d_1, d_2)\)]. We illustrate this in Table 7 for \(T = 4\). Here we calculate these bounds for the designs that correspond to Figures 5 to 8. They range from 0.08 to 0.2. To put the numbers in Table 7 in context, we estimated the standard deviation of \(\frac{\exp (\gamma_1 + \gamma_2 + \alpha)}{1 + \exp (\gamma_1 + \gamma_2 + \alpha)} - \frac{\exp (\alpha)}{1 + \exp (\alpha)}\) that one would get if one assumes that \(\alpha\) is constant across observations and estimated \((\alpha, \gamma_1, \gamma_2)\) by maximum likelihood based on 1000 observations of \((y_1, y_2, y_3, y_4)\) from each of the models. These standard deviations range from 0.027 to 0.032 across all eight models that we consider, suggesting 95% confidence intervals of length approximately 0.12. In other words, in this case the length of the identified region is of the same order of magnitude as the confidence intervals would be in a much simpler misspecified model.

5 Concluding Remarks

A number of papers in different areas of economics have demonstrated that the identified regions of non-identified parameters can be small enough to be useful in applications. The papers by Haile and Tamer

\(^6\)Unconditional bounds can then be obtained by averaging over the probability distribution of the initial conditions.
(2003), Honoré and Lleras-Muney (2006), and Blundell, Gosling, Ichimura, and Meghir (2007) are early examples of this. In this paper, we have studied the identified regions for the parameters of two variations of the panel data logit model, a dynamic model with two lags and a dynamic panel data version of Schmidt and Strauss (1975)’s simultaneous logit model. Both of the models have the feature that part of the parameter vector is identified for some values of $T$, while the identifiability of the other part of the parameter vector is unknown. Our calculations suggest that the latter might indeed be point identified, although we argue that it would be surprising if a conditional likelihood argument can be used to prove this. Moreover, our calculations give some hints about what to look at (and what to not look at) in order to formally prove identification. Doing this is clearly an important topic for future research. Furthermore, our calculations of the marginal effects in Section 4 illustrate that the identified region for a parameter vector that is not point identified can be small enough to be empirically useful.

One weakness of the models studied here is that they are dynamic in a mechanical way. Ongoing research by Aguirregabiria, Gu, and Luo (2018) modifies the models to make them consistent with simple economic models of dynamic decision making. We speculate that combining the ideas illustrated here with the models studied in Aguirregabiria, Gu, and Luo (2018) will lead to new insights.

References


Tables with identified regions in an AR(2) logit with $T = 5$. $\gamma_1 = \pm 1$, $\gamma_2 = \pm 1$.

Table 1: Identified Regions. Heterogeneity Uniform.

| Source of identification | $Q((g_1, g_2)|0,0)$ | $Q((g_1, g_2)|1,0)$ | $Q((g_1, g_2)|0,1)$ | $Q((g_1, g_2)|1,1)$ | $Q(g_1, g_2)$ |
|--------------------------|----------------------|----------------------|----------------------|----------------------|----------------|
| True $(\gamma_1, \gamma_2)$ |                      |                      |                      |                      |                |
| $\gamma_1 = -1$          | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$   |
| $\gamma_2 = -1$          | $[-1.0086, -0.9942]$  | $\{-1.000\}$         | $\{-1.000\}$         | $[-1.0090, -0.9940]$  | $\{-1.000\}$   |
| $\gamma_1 = -1$          | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$   |
| $\gamma_2 = 1$           | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$    |
| $\gamma_1 = 1$           | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$    |
| $\gamma_2 = -1$          | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$   |
| $\gamma_1 = 1$           | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$    |
| $\gamma_2 = 1$           | $[0.9973, 1.0055]$    | $\{1.000\}$          | $\{1.000\}$          | $[0.9965, 1.0068]$    | $\{1.000\}$    |

Table 2: Identified Regions. Heterogeneity Mixture of Normals.

| Source of identification | $Q((g_1, g_2)|0,0)$ | $Q((g_1, g_2)|1,0)$ | $Q((g_1, g_2)|0,1)$ | $Q((g_1, g_2)|1,1)$ | $Q(g_1, g_2)$ |
|--------------------------|----------------------|----------------------|----------------------|----------------------|----------------|
| True $(\gamma_1, \gamma_2)$ |                      |                      |                      |                      |                |
| $\gamma_1 = -1$          | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$   |
| $\gamma_2 = -1$          | $[-1.0086, -0.9942]$  | $\{-1.000\}$         | $\{-1.000\}$         | $[-1.0090, -0.9940]$  | $\{-1.000\}$   |
| $\gamma_1 = -1$          | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$   |
| $\gamma_2 = 1$           | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$    |
| $\gamma_1 = 1$           | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$    |
| $\gamma_2 = -1$          | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$         | $\{-1.000\}$   |
| $\gamma_1 = 1$           | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$          | $\{1.000\}$    |
| $\gamma_2 = 1$           | $[0.9973, 1.0055]$    | $\{1.000\}$          | $\{1.000\}$          | $[0.9965, 1.0068]$    | $\{1.000\}$    |
Tables with identified regions in an AR(2) logit with $T = 5$. $\gamma_1 = \pm 2$, $\gamma_2 = \pm 2$.

Table 3: Identified Regions. Heterogeneity Uniform.

| Source of identification | $Q((g_1, g_2)|0,0)$ | $Q((g_1, g_2)|1,0)$ | $Q((g_1, g_2)|0,1)$ | $Q((g_1, g_2)|1,1)$ | $Q(g_1, g_2)$ |
|--------------------------|-----------------------|-----------------------|-----------------------|-----------------------|----------------|
| True ($\gamma_1, \gamma_2$) | | | | | |
| $\gamma_1 = -1, \gamma_2 = 1$ | {-2.0000} | {-2.0000} | {-2.0000} | {-2.0000} | {-2.0000} |
| $\gamma_1 = -1, \gamma_2 = -1$ | [-2.0011, -1.9989] | {-2.0000} | {-2.0000} | [-2.0011, -1.9989] | {-2.0000} |
| $\gamma_1 = 1, \gamma_2 = -1$ | {-2.0000} | {-2.0000} | {-2.0000} | {-2.0000} | {-2.0000} |
| $\gamma_1 = 1, \gamma_2 = 1$ | {2.0000} | {2.0000} | {2.0000} | {2.0000} | {2.0000} |
| $\gamma_1 = 1, \gamma_2 = -1$ | {2.0000} | {2.0000} | {2.0000} | {2.0000} | {2.0000} |
| $\gamma_1 = 1, \gamma_2 = 1$ | [1.9959, 2.0057] | {2.0000} | {2.0000} | [1.9903, 2.0119] | {2.0000} |
| $\gamma_1 = 1, \gamma_2 = -1$ | [1.9978, 2.0044] | {2.0000} | {2.0000} | [1.9941, 2.0101] | {2.0000} |

Table 4: Identified Regions. Heterogeneity Mixture of Normals.

| Source of identification | $Q((g_1, g_2)|0,0)$ | $Q((g_1, g_2)|1,0)$ | $Q((g_1, g_2)|0,1)$ | $Q((g_1, g_2)|1,1)$ | $Q(g_1, g_2)$ |
|--------------------------|-----------------------|-----------------------|-----------------------|-----------------------|----------------|
| True ($\gamma_1, \gamma_2$) | | | | | |
| $\gamma_1 = -2, \gamma_2 = 2$ | {-2.0000} | {-2.0000} | {-2.0000} | {-2.0000} | {-2.0000} |
| $\gamma_1 = -2, \gamma_2 = -2$ | [-2.0214, -1.9854] | {-2.0000} | {-2.0000} | [-2.0252, -1.9834] | {-2.0000} |
| $\gamma_1 = 2, \gamma_2 = 2$ | {-2.0000} | {-2.0000} | {-2.0000} | {-2.0000} | {-2.0000} |
| $\gamma_1 = 2, \gamma_2 = -2$ | {2.0000} | {2.0000} | {2.0000} | {2.0000} | {2.0000} |
| $\gamma_1 = 2, \gamma_2 = 2$ | {2.0000} | {2.0000} | {2.0000} | {2.0000} | {2.0000} |
| $\gamma_1 = 2, \gamma_2 = -2$ | [1.9978, 2.0044] | {2.0000} | {2.0000} | [1.9941, 2.0101] | {2.0000} |
Tables with identified regions in an AR(2) logit with $T = 5$. $\gamma_1 = \pm 1.5$, $\gamma_2 = \pm 0.7$.

Table 5: Identified Regions. Heterogeneity Uniform.

| Source of identification | $Q((g_1, g_2)|0, 0)$ | $Q((g_1, g_2)|1, 0)$ | $Q((g_1, g_2)|0, 1)$ | $Q((g_1, g_2)|1, 1)$ | $Q(g_1, g_2)$ |
|--------------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------|
| $\gamma_1 = -1.5$       | $[-1.5000]$           | $[-1.5000]$           | $[-1.5000]$           | $[-1.5000]$           | $[-1.5000]$       |
| $\gamma_2 = 0.7$        | $[-0.7042, -0.6966]$  | $[-0.7000]$           | $[-0.7000]$           | $[-0.7061, -0.6951]$  | $[-0.7000]$       |
| $\gamma_1 = -1.5$       | $[-1.5000]$           | $[-1.5000]$           | $[-1.5000]$           | $[-1.5000]$           | $[-1.5000]$       |
| $\gamma_2 = 0.7$        | $[0.6979, 0.7017]$    | $[0.7000]$            | $[0.7000]$            | $[0.6969, 0.7025]$    | $[0.7000]$        |
| $\gamma_1 = 1.5$        | $[1.5000]$            | $[1.5000]$            | $[1.5000]$            | $[1.5000]$            | $[1.5000]$        |
| $\gamma_2 = -0.7$       | $[-0.7020, -0.6976]$  | $[-0.7000]$           | $[-0.7000]$           | $[-0.7018, -0.6978]$  | $[-0.7000]$       |
| $\gamma_1 = 1.5$        | $[1.5000]$            | $[1.5000]$            | $[1.5000]$            | $[1.5000]$            | $[1.5000]$        |
| $\gamma_2 = 0.7$        | $[0.6967, 0.7041]$    | $[0.7000]$            | $[0.7000]$            | $[0.6968, 0.7040]$    | $[0.7000]$        |

Table 6: Identified Regions. Heterogeneity Mixture of Normals.

| Source of identification | $Q((g_1, g_2)|0, 0)$ | $Q((g_1, g_2)|1, 0)$ | $Q((g_1, g_2)|0, 1)$ | $Q((g_1, g_2)|1, 1)$ | $Q(g_1, g_2)$ |
|--------------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------|
| $\gamma_1 = -1.5$       | $[-1.5000]$           | $[-1.5000]$           | $[-1.5000]$           | $[-1.5000]$           | $[-1.5000]$       |
| $\gamma_2 = 0.7$        | $[-0.7032, -0.6978]$  | $[-0.7000]$           | $[-0.7000]$           | $[-0.7044, -0.6971]$  | $[-2.0000]$       |
| $\gamma_1 = -1.5$       | $[-1.5000]$           | $[-1.5000]$           | $[-1.5000]$           | $[-1.5000]$           | $[-1.5000]$       |
| $\gamma_2 = 0.7$        | $[0.6985, 0.7010]$    | $[0.7000]$            | $[0.7000]$            | $[0.6974, 0.7017]$    | $[2.0000]$        |
| $\gamma_1 = 1.5$        | $[1.5000]$            | $[1.5000]$            | $[1.5000]$            | $[1.5000]$            | $[1.5000]$        |
| $\gamma_2 = -0.7$       | $[-0.7012, -0.6983]$  | $[-0.7000]$           | $[-0.7000]$           | $[-0.7011, -0.6983]$  | $[-2.0000]$       |
| $\gamma_1 = 1.5$        | $[1.5000]$            | $[1.5000]$            | $[1.5000]$            | $[1.5000]$            | $[1.5000]$        |
| $\gamma_2 = 0.7$        | $[0.6981, 0.7029]$    | $[0.7000]$            | $[0.7000]$            | $[0.6979, 0.7031]$    | $[2.0000]$        |
$Q(g_1, \gamma_2|d_1, d_1)$ as a function of $g_1$ in the AR(2) logit with $T = 6$. $\gamma_1 = \pm 1$, $\gamma_2 = \pm 1$.

Figure 1: Objective Functions when $(\gamma_1, \gamma_2) = (-1, -1)$.

Figure 2: Objective Functions when $(\gamma_1, \gamma_2) = (-1, 1)$.

Figure 3: Objective Functions when $(\gamma_1, \gamma_2) = (1, -1)$.

Figure 4: Objective Functions when $(\gamma_1, \gamma_2) = (1, 1)$.
Identified regions in an AR(2) and $T = 4$. $\gamma_1 = \pm 1$, $\gamma_2 = \pm 1$.

Figure 5: Identified Region when $\gamma_1 = -1$ and $\gamma_2 = -1$

Figure 6: Identified Region when $\gamma_1 = -1$ and $\gamma_2 = 1$

Figure 7: Identified Region when $\gamma_1 = 1$ and $\gamma_2 = -1$

Figure 8: Identified Region when $\gamma_1 = 1$ and $\gamma_2 = 1$
$Q(\rho|y_{1,i1}, y_{1,i2})$ in the AR(1) Schmidt and Strauss Model

Figure 9: Objective Functions with Independent $\alpha$’s

Figure 10: Objective Functions with Positively Correlated $\alpha$’s

Figure 11: Objective Functions with Negatively Correlated $\alpha$’s
Table 7: Marginal Effects: $T = 4$

<table>
<thead>
<tr>
<th>Heterogeneity: Uniform</th>
<th>Heterogeneity: Mixture of Normals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1 = -1$ $\gamma_2 = -1$</td>
<td>$(-0.34, -0.26)$</td>
</tr>
<tr>
<td>$\gamma_1 = -1$ $\gamma_2 = 1$</td>
<td>$(-0.14, 0.05)$</td>
</tr>
<tr>
<td>$\gamma_1 = 1$ $\gamma_2 = -1$</td>
<td>$(-0.05, 0.05)$</td>
</tr>
<tr>
<td>$\gamma_1 = 1$ $\gamma_2 = 1$</td>
<td>$(0.20, 0.40)$</td>
</tr>
<tr>
<td>$\gamma_1 = -1$ $\gamma_2 = -1$</td>
<td>$(-0.37, -0.29)$</td>
</tr>
<tr>
<td>$\gamma_1 = -1$ $\gamma_2 = 1$</td>
<td>$(-0.13, 0.05)$</td>
</tr>
<tr>
<td>$\gamma_1 = 1$ $\gamma_2 = -1$</td>
<td>$(-0.05, 0.04)$</td>
</tr>
<tr>
<td>$\gamma_1 = 1$ $\gamma_2 = 1$</td>
<td>$(0.25, 0.42)$</td>
</tr>
</tbody>
</table>