Responses to the incidental parameter problem

Pua, A.A.Y.

Publication date
2016

Document Version
Final published version

Chapter 3

Simultaneous equations models for discrete outcomes: Coherence and completeness using panel data

3.1 Introduction

In this chapter, I show how to estimate a dynamic simultaneous equations panel data model with discrete outcomes. There are two main issues involved in this endeavor, namely, the manner in which time-invariant unobserved heterogeneity is introduced and the manner in which the nonexistence of a unique reduced form is addressed. Both these issues have implications for how the parameters of the simultaneous equations model are going to be identified and estimated. I show how both these issues can be tackled at the same time.

Researchers who want to estimate a dynamic simultaneous equations model with discrete outcomes using panel data would have to introduce an additive individual-specific fixed effect into the latent index. Unfortunately, time-invariant unobserved heterogeneity cannot be left unrestricted in dynamic nonlinear panel data models, especially when the number of time periods T is small (see Section 4 of Arellano and Bonhomme (2011)). Although we have bias reduction procedures for parameters of interest, they are motivated from a large-T perspective. Results of existing Monte Carlo simulations for dynamic nonlinear panel data models indicate that T has to be much larger than 10 in order to reap the gains from bias reduction (see Bester and Hansen (2009a), Carro (2007), Fernandez-Val (2009), and Hahn and Kuersteiner

(2011)). Furthermore, the fixed-T solution proposed by Bonhomme (2012) only applies to models without dynamics. As a compromise, we have to restrict some features of the distribution of time-invariant unobserved heterogeneity. Allowing for fixed effects in these models has not been explored fully, since most research has focused on either cross-sectional models, continuous outcomes (or the latent outcomes themselves), or random effects (for examples, see the work by Cornwell, Schmidt, and Wyhowski (1992), Leon-Gonzalez (2003), Matzkin (2008), Matzkin (2012), and Masten (2015)).

Even if T is large, coherency conditions would still have to be imposed. Coherency conditions effectively convert a model where the endogenous variables are jointly determined into a model which is triangular or recursive. A triangular model restricts the direction in which an endogenous variable affects other endogenous variables (for all observations). Triangularity implies that there are either zero or inequality restrictions on the parameters or functions of the parameters. For example, when there are two binary endogenous variables y_1 and y_2 that are jointly determined, y_2 should not enter the equation for y_1 or vice-versa. As a result, we have to choose beforehand how the coherency conditions should be imposed.

The literature on coherency conditions started with research aiming to extend the simultaneous equations approach of the Cowles Commission to endogenous variables that are subject to censoring or truncation. Some representative papers in this area include Maddala and Lee (1976), Heckman (1978), Gourieroux, Laffont, and Monfort (1980), and Schmidt (1981). Blundell and Smith (1993) summarize this strand of the literature. They have all shown that parameter restrictions are typically required to ensure the existence and uniqueness of the reduced form. As a result, ensuring coherency is a first step prior to discussing identification. Later research has focused more on how to avoid imposing the coherency conditions (see the contribution of Tamer (2003)).

In order to avoid imposing these coherency conditions, a separate strand of the literature has emphasized that the uniqueness of equilibrium in game-theoretic models has parallels with the problems involving uniqueness of the reduced form for simultaneous equation models with discrete outcomes. Early work in the estimation of game-theoretic models such as Bjorn and Vuong (1984), Bresnahan and Reiss (1991), and Kooreman (1994) attempt to overcome the possibility of multiple equilibria either by introducing a selection mechanism which assumes that players choose one of the equilibria at random or by fusing multiple equilibria as one outcome. Tamer (2003) shows that point identification and consistent estimation is still possible without imposing a set of auxiliary assumptions that resolve the underlying multiplicity. All that is needed is the presence of a regressor with large support. He suggests a semiparametric ML estimator that is more efficient than the ML estimator that fuses multiple equilibria as one outcome. In fact, Tamer (2003) introduces new terminology to differentiate models whose reduced form is nonexistent and models whose reduced form is nonunique. He calls these models incoherent and incomplete, respectively. Note that these cited papers focus on the incompleteness aspect because the signs of some parameters of game-theoretic models may be known a priori. These sign restrictions effectively rule out potential incoherence of the model.

Some like Dagenais (1999), Massacci (2010), and Hajivassiliou and Savignac (2011) have attempted to resolve both incompleteness and incoherence by imposing error-support restrictions. They all offer estimation methods that involve reweighting the likelihood contributions to reflect the restrictions on the error supports. In the most recent work by Chesher and Rosen (2012), they show how identified sets can be constructed without resorting to the restriction of error supports and a priori sign restrictions at all. In the process of constructing these identified sets, they were able to unify the different approaches that are available in the literature in the most general way possible. With the exception of Hajivassiliou and Savignac (2011), the preceding papers focus on either cross-sectional or time-series settings. On the other hand, Hajivassiliou and Savignac (2011) use panel data to estimate a model of the joint determination of a firm's decision to innovate and a firm's exposure to higher credit constraints but do not allow for fixed effects.

My proposal approaches the problem of incompleteness and incoherence from a different perspective. I exploit Lewbel's (2007) characterization of coherence and completeness when one of the endogenous variables is binary. He shows that a possible characterization involves indexing the direction of causality by a dummy variable which may be observable or modeled. In contrast, I do not restrict the dependence on observables but assume that this dependence is individual-specific. As a result, I allow for the estimation of a panel data simultaneous equations model with discrete outcomes where the individual-specific fixed effect can be interpreted as the manner in which the coherency condition holds or the direction in which causality flows from one discrete endogenous variable to another.

The paper is organized as follows. In Section 3.2, I provide a motivating example to demonstrate my proposal. In Section 3.3, I discuss how identification, estimation and inference may proceed in the model considered by Hajivassiliou and Ioannides (2007) (henceforth, HI). In Section 3.4, I revisit the empirical application of HI and cast doubt on the coherency conditions they have imposed. I conclude and suggest avenues for further work in Section 3.5.

3.2 A stylized example

3.2.1 Coherence and completeness

I start by introducing some terminology in the context of a simple example. Consider the situation where two dummy variables are jointly determined. This situation typically arises in many empirical applications, such as determining whether binary choices are substitutes or complements (Lewbel, 2007), estimating game-theoretic models with discrete actions (Bjorn and Vuong, 1984; Bresnahan and Reiss, 1991; Kooreman, 1994; Tamer, 2003; Hahn and Moon, 2010), modelling vote trading among congressmen for agricultural issues (Stratmann, 1992; Dagenais, 1999), and modelling fertility decisions among couples (Sobel and Arminger, 1992), to name a few.

Let (y_1, y_2) be two dummy endogenous variables jointly determined by the system

$$\begin{aligned} y_1^* &= y_2\alpha_1 + \epsilon_1,\ y_1 = 1\left\{y_1^* \geq 0\right\}, \qquad (3.2.1)\\ y_2^* &= y_1\alpha_2 + \epsilon_2,\ y_2 = 1\left\{y_2^* \geq 0\right\}, \qquad (3.2.2)\end{aligned}$$

where and (ϵ_1, ϵ_2) are the error terms. Only the signs of y_1^* and y_2^* are observable.[1] There are four possible outcomes for (y_1, y_2) and they arise according to the following rule:

$$(y_1, y_2) = \begin{cases} (1,1) & \text{if } \epsilon_1 > -\alpha_1,\ \epsilon_2 > -\alpha_2 \\ (1,0) & \text{if } \epsilon_1 > 0,\ \epsilon_2 \leq -\alpha_2 \\ (0,1) & \text{if } \epsilon_1 \leq -\alpha_1,\ \epsilon_2 > 0 \\ (0,0) & \text{if } \epsilon_1 \leq 0,\ \epsilon_2 \leq 0 \end{cases}.$$

Geometrically, the inequalities define regions in (ϵ_1, ϵ_2)-space. These regions will overlap when $\alpha_2\alpha_1 > 0$. As a result, y_1 may be assigned the value 0 or 1 in the overlapping region. This non-uniqueness of y_1 is called incompleteness. The model is indeterminate for y_1 for some (ϵ_1, ϵ_2).

On the other hand, the inequalities may lead to an empty region when $\alpha_2\alpha_1 < 0$. As a result, the model is unable to definitively assign a value for y_1 in the empty region. This nonexistence is called incoherence. Unfortunately, the data do not allow us to distinguish between the two cases unless we have prior information about the sign of $\alpha_2\alpha_1$ or we have a way of resolving how Nature (or perhaps the observed units) would assign (or choose) values for y_1 in those regions. We can overcome this by assuming $\alpha_2\alpha_1 = 0$. This restriction, called the coherency condition, assumes away the simultaneity initially posited for the endogenous variables.[2] However, even

[1] Blundell and Smith (1994) call the model above Type II because it is the observed indicators (y_1, y_2) that enter as right-hand side variables. In contrast, Type I models are models in which the latent variables (y_1^*, y_2^*) enter as right-hand side variables. In the latter case, standard simultaneous equation methods can be applied. Matzkin (2012) considers panel data versions of Type I models.

[2] Another example is where $y_2^* = y_2$ is fully observable, as opposed to just the sign of y_2 being observable. Solving for y_1^* gives us

$$y_1^* = y_1\alpha_2\alpha_1 + \alpha_1\epsilon_2 + \epsilon_1.$$

There are only two possible observable values for y_1:

$$y_1 = \begin{cases} 0 & \text{if } \epsilon_1 + \alpha_1\epsilon_2 \leq 0 \\ 1 & \text{if } \epsilon_1 + \alpha_1\epsilon_2 > -\alpha_2\alpha_1 \end{cases}.$$

if we take the required condition that $\alpha_2\alpha_1 = 0$ seriously, it is not clear whether we should proceed with identification and estimation under $\alpha_1 = 0$ or $\alpha_2 = 0$. Because the coherency condition has to be imposed, most empirical applications would proceed by (a) producing two sets of results (depending on whether $\alpha_1 = 0$ or $\alpha_2 = 0$) (b) choosing to start with a triangular model from the onset (c) introducing the latent variables y_1^* and y_2^* instead of y_1 and y_2 in Equations (3.2.1) and (3.2.2).

Lewbel (2007) shows that it is possible to avoid setting either $\alpha_1 = 0$ or $\alpha_2 = 0$ by choosing a coherent and complete representation. In a nonseparable simultaneous equations model where one of the endogenous variables is binary, he shows that coherence and completeness restrict the manner in which some of the endogenous variables enter into the structural equations:

Theorem 3.2.1. *Assume that $y_1 \in \{0,1\}$, $y_2 \in \Psi$, and $w \in \Omega$ for some support sets Ψ and Ω.*[3] *The system*

$$\begin{aligned} y_1 &= H_1(y_1, y_2, w), \\ y_2 &= H_2(y_1, y_2, w) \end{aligned}$$

is coherent and complete if and only if there exists a function $g : \{0,1\} \times \Omega \to \Psi$ such that for all $w \in \Omega$, we have

$$\begin{aligned} H_1(0, g(0,w), w) &= H_1(1, g(1,w), w), \\ y_2 &= g(y_1, w). \end{aligned}$$

The proof of this theorem can be found in Lewbel (2007). As a result, the function H_1 should not depend on y_1. More importantly, the theorem allows us to choose g to ensure coherence and completeness without imposing sign restrictions or imposing error support restrictions that may be data-dependent. In the context of the model in (3.2.1) and (3.2.2), he shows that a coherent and complete representation can be chosen by defining a dummy function $d : \Omega \to \{0,1\}$ such that

$$\begin{aligned} y_1 &= 1\{(1-d(w))\,y_2\alpha_1 + \epsilon_1 \geq 0\}, \\ y_2 &= 1\{d(w)\,y_1\alpha_2 + \epsilon_2 \geq 0\}. \end{aligned}$$

The inclusion of d has intuitive appeal because some units may have $d = 0$ or $d = 1$ depending on the values of the observables and unobservables. As a result, y_1

Once again, we must have $\alpha_2\alpha_1 = 0$ so that a unique reduced form would exist. More empirical applications belong to this category of models. Examples include modelling the export-productivity relationship (Clerides, Lach, and Tybout, 1998), the household's bundle of appliance holdings and electricity consumption (Dubin and McFadden, 1984), and the measurement of recovery via output growth from crises (Cerra and Saxena, 2008).

[3] These support sets may contain either a finite or an infinite number of elements. The only requirement is that y_1 is binary.

depends on y_2 when $d = 0$ and y_2 depends on y_1 when $d = 1$. Unfortunately, Lewbel (2007) assumes that d is observable or could be modeled in some way.[4]

Note that Lewbel's (2007) approach gives a new interpretation for simultaneity when compared to its traditional usage in econometrics. Instead of simultaneity being a purely structural aspect of the model, simultaneity arises because of the econometrician's inability to distinguish the direction of dependence (from y_1 to y_2 or vice versa). Intuitively, information from repeated measurements can be useful in overcoming this inability without making parametric assumptions on the dummy function d.

In contrast to linear simultaneous equations models where there is a form of "bidirectional" causality (not in the usual Granger-causality sense) between two continuously distributed variables, the models I consider only allow for one-directional causality from y_1 to y_2 for a subset of observations and one-directional causality from y_2 to y_1 for the remaining set of observations. This is more credible than imposing the coherency conditions which will ultimately result in either one-directional causality from y_1 to y_2 for *all* observations or one-directional causality from y_2 to y_1 for *all* observations, but not both.

The model in (3.2.1) and (3.2.2) can be thought of as a static discrete game of complete information. In contrast to game-theoretic models, the models I consider allow for both incompleteness and incoherence. Incompleteness arising from multiple equilibria is a common feature in the estimation of game-theoretic models. Hahn and Moon (2010) estimate (3.2.1) and (3.2.2) using panel data of pairs of agents using Nash play. They assume that the model is incomplete by imposing sign restrictions $\alpha_1 < 0$ and $\alpha_2 < 0$ and that pairs of agents choose one of the two equilibria and stick to that same choice throughout time.

Since I am working with a complete and coherent representation by introducing d, it may be useful to motivate the representation in game-theoretic terms. One can view the representation as arising from the inability of the econometrician to observe how multiple equilibria or absence of equilibrium was resolved, possibly through an unmodelled communication or coordination device. Alternatively, the econometrician may also have neglected to model the sequential nature of the game and was unable to observe which player moved first (but the agents are aware of the sequential nature of the game).

[4]Introducing this dummy function is just one way to obtain a coherent and complete representation whenever one of the endogenous variables is binary. One could argue that this device may be useful in a linear simultaneous equations setting. A step in this direction is the introduction of random coefficients in a linear simultaneous equations model (see Masten (2015) for more details).

3.2.2 Why a cross section is not enough

Suppose we treat d as unobservable. For the moment, assume we only have a cross-section of observations which form a random sample. Consider the model

$$y_{1i} = 1\left((1-d_i)\, y_{2i}\alpha_1 + \epsilon_{1i} \geq 0\right), \tag{3.2.3}$$

$$y_{2i} = 1\left(d_i y_{1i}\alpha_2 + \epsilon_{2i} \geq 0\right), \tag{3.2.4}$$

for $i = 1, \ldots, n$. Let $\Pr(d_i = 1) = p_i$ and $\Pr(d_i = 0) = 1 - p_i$, where $p_i \in (0,1)$. I assume that d_i, ϵ_{1i} and ϵ_{2i} are i.i.d. draws from their joint distribution. There are four joint probabilities of the form $\Pr(y_{1i} = j, y_{2i} = k)$, where $(j,k) \in \{0,1\} \times \{0,1\}$, that are observable from the data but only three of these provide restrictions on the parameters of the model (since all these four probabilities should sum up to one). As a result, it is not possible to have point identification for all parameters, even if we set $p_i = p$ for all i. The main reason is that there are four parameters $(\alpha_1, \alpha_2, \rho, p)$ to identify given the three joint probabilities obtained from the data.

To illustrate and simplify things further, assume normality[5][6] for the errors given d_i for all i, i.e.

$$\left(\begin{array}{c} \epsilon_{1i} \\ \epsilon_{2i} \end{array} \right) \Bigg| d_i \sim N\left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array} \right) \right). \tag{3.2.5}$$

As a result, we have

$$\left(\begin{array}{c} \epsilon_{1i} \\ \epsilon_{2i} \end{array} \right) \Bigg| d_i = 0 \sim \left(\begin{array}{c} \epsilon_{1i} \\ \epsilon_{2i} \end{array} \right) \Bigg| d_i = 1,$$

i.e., $(\epsilon_{1i}, \epsilon_{2i})$ is independent of d_i. The parameters of interest are α_1, α_2, and ρ. Note that for all $(j,k) \in \{0,1\} \times \{0,1\}$, we have

$$\begin{aligned} \Pr(y_{1i} = j, y_{2i} = k) &= \Pr(y_{1i} = j, y_{2i} = k | d_i = 1)\, p_i \\ &\quad + \Pr(y_{1i} = j, y_{2i} = k | d_i = 0)(1 - p_i). \end{aligned}$$

Specifically, we have

$$\Pr(y_{1i} = 0, y_{2i} = 0) = \Pr(\epsilon_{1i} \leq 0, \epsilon_{2i} \leq 0;\, \rho), \tag{3.2.6}$$

$$\Pr(y_{1i} = 1, y_{2i} = 0) = \Pr(\epsilon_{1i} \geq 0, \epsilon_{2i} \leq -\alpha_2;\, \rho)\, p_i + \Pr(\epsilon_{1i} \geq 0, \epsilon_{2i} \leq 0;\, \rho)(1 - p_i), \tag{3.2.7}$$

$$\Pr(y_{1i} = 0, y_{2i} = 1) = \Pr(\epsilon_{1i} \leq -\alpha_1, \epsilon_{2i} \geq 0;\, \rho)(1 - p_i) + \Pr(\epsilon_{1i} \leq 0, \epsilon_{2i} \geq 0;\, \rho)\, p_i. \tag{3.2.8}$$

---

[5] Normality is not really required here. All that is required is a bivariate cdf that is strictly monotonic in ρ. In cases where ρ is not a correlation coefficient but some one-dimensional parameter that indexes dependence modelled via a copula. See Section 3.3 for more details.

[6] Bivariate normality of the errors can be thought of as a simple factor structure. In particular, ϵ_{1i} can be written as $\epsilon_{1i} = \rho\epsilon_{2i} + v_i$ where $v_i \sim N\left(0, 1 - \rho^2\right)$, $\epsilon_{2i} \sim N(0,1)$, and ϵ_{2i} is independent of v_i.

The left hand sides of (3.2.6), (3.2.7), and (3.2.8) are observable from the data. We are unable to observe the mixing probability p_i. As long as we observe outcomes of the form $(y_{1i} = 0, y_{2i} = 0)$ from the data,[7] (3.2.6) can be used to identify ρ because the bivariate normal cdf is strictly monotonic in ρ (hence, the bivariate normal cdf is invertible with respect to ρ). This is in contrast with identification problems associated with ρ as documented by Freedman and Sekhon (2010) and Meango and Mourifie (2013) in the context of triangular models with a dummy endogenous regressor.

Since ρ is identified, we can treat it as known for the next step. In particular, we can now use (3.2.7) and (3.2.8) to identify whether (i) $\alpha_1 \leq 0$ or $\alpha_1 \geq 0$ and (ii) $\alpha_2 \leq 0$ or $\alpha_2 \geq 0$. Whenever the cross-sectional frequency of $(y_{1i} = 1, y_{2i} = 0)$ is less than or equal to $\Pr(\epsilon_{1i} \geq 0, \epsilon_{2i} \leq 0; \rho)$, we must have $\alpha_2 \geq 0$. Similarly, showing that the cross-sectional frequency of $(y_{1i} = 0, y_{2i} = 1)$ is less than or equal to $\Pr(\epsilon_{1i} \leq 0, \epsilon_{2i} \geq 0; \rho)$ allows us to conclude that $\alpha_1 \geq 0$. The other cases follow analogously.

Unfortunately, we need external information to determine the grouping of every ith unit. One possible route is to impose $p_i = \Pr(d_i = 1) = \Pr(d = 1) = p$ for all i. Knowing the signs of α_1, α_2 allows us to determine the group to which all units belong. There are four cases to consider as shown in Table 3.2.1.

Table 3.2.1: Group assignment rules for model (3.2.3) and (3.2.4)

Condition	Rule
$\alpha_1 \geq 0,\ \alpha_2 \geq 0$	$d = 0$ iff frequency of $y_{1i} = 0$ less than $\Pr(\epsilon_{1i} \leq 0)$ $d = 1$ iff frequency of $y_{2i} = 0$ less than $\Pr(\epsilon_{2i} \leq 0)$
$\alpha_1 \leq 0,\ \alpha_2 \leq 0$	$d = 0$ iff frequency of $y_{1i} = 0$ greater than $\Pr(\epsilon_{1i} \leq 0)$ $d = 1$ iff frequency of $y_{2i} = 0$ greater than $\Pr(\epsilon_{2i} \leq 0)$
$\alpha_1 \geq 0,\ \alpha_2 \leq 0$	$d = 0$ iff frequency of $y_{1i} = 0$ less than $\Pr(\epsilon_{1i} \leq 0)$ $d = 1$ iff frequency of $y_{2i} = 0$ greater than $\Pr(\epsilon_{2i} \leq 0)$
$\alpha_1 \leq 0,\ \alpha_2 \geq 0$	$d = 0$ iff frequency of $y_{1i} = 0$ greater than $\Pr(\epsilon_{1i} \leq 0)$ $d = 1$ iff frequency of $y_{2i} = 0$ less than $\Pr(\epsilon_{2i} \leq 0)$

Once we know whether $d = 0$ or $d = 1$ for all i, we are able to only point-identify either α_1 or α_2 but not both. Suppose $d = 0$ for the moment. Since $\Pr(y_{1i} = 0, y_{2i} = 1 | d = 0) = \Pr(\epsilon_{1i} \leq -\alpha_1, \epsilon_{2i} \geq 0; \rho)$ and $\Pr(\epsilon_{1i} \leq -\alpha_1, \epsilon_{2i} \geq 0; \rho)$ is strictly decreasing in α_1 for fixed ρ, we are able to point-identify α_1. In contrast, α_2 is set-identified because $d = 0$ for all observations and we know the sign

[7] The extreme case where we do not observe any other outcome aside from $(y_{1i} = 0, y_{2i} = 0)$ is ruled out.

of α_2. Similarly, we are able to point-identify α_2 from $\Pr(y_{1i}=1, y_{2i}=0|d=1) = \Pr(\epsilon_{1i} \geq 0, \epsilon_{2i} \leq -\alpha_2; \rho)$ but only set-identify α_1.

3.2.3 Why panel data may be useful

At this point, panel data may be useful for point identification of all the parameters. We can introduce individual-specific effects through d to allow for unrestricted dependence on the observables in a time-invariant manner. Note that I do not introduce individual-specific effects as intercepts in the linear predictors. As a result, time-invariant variables or even variables that do not have too much variation over time may be included in the model. In contrast, the usual way of introducing fixed effects additively precludes the inclusion of time-invariant variables of interest.

If we had panel data and imposed the assumption that p_i does not vary over time, we will be able to achieve point identification. Consider once again the model but this time adapted to panel data:

$$\begin{aligned} y_{1it} &= 1\left((1-d_i)\, y_{2it}\alpha_1 + \epsilon_{1it} \geq 0\right), \\ y_{2it} &= 1\left(d_i y_{1it}\alpha_2 + \epsilon_{2it} \geq 0\right), \end{aligned}$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. Let $\Pr(d_i = 1) = p_i$ and $\Pr(d_i = 0) = 1 - p_i$, where $p_i \in (0,1)$. I assume that d_i, ϵ_{1it} and ϵ_{2it} are i.i.d. draws from their joint distribution. Assume bivariate normality once again as in (3.2.5).[8] Analogously, we have

$$\Pr(y_{1it}=0, y_{2it}=0) = \Pr(\epsilon_{1it} \leq 0, \epsilon_{2it} \leq 0; \rho), \tag{3.2.9}$$

$$\Pr(y_{1it}=1, y_{2it}=0) = \Pr(\epsilon_{1it} \geq 0, \epsilon_{2it} \leq -\alpha_2; \rho)\, p_i + \Pr(\epsilon_{1it} \geq 0, \epsilon_{2it} \leq 0; \rho)(1-p_i), \tag{3.2.10}$$

$$\Pr(y_{1it}=0, y_{2it}=1) = \Pr(\epsilon_{1it} \leq -\alpha_1, \epsilon_{2it} \geq 0; \rho)(1-p_i) + \Pr(\epsilon_{1it} \leq 0, \epsilon_{2it} \geq 0; \rho)\, p_i. \tag{3.2.11}$$

We now follow the same steps in the identification argument of the previous subsection. Data on the observed frequencies of $(y_{1it}=0, y_{2it}=0)$ for all i and t point identify ρ from (3.2.9). After plugging in the value of ρ from the previous step, (3.2.10) and (3.2.11) for all i and t can be used to identify the signs of α_1, α_2 as before. Once the signs are identified, we can modify the group assignment rules in Table 3.2.1. Instead of using cross-sectional variation, we use time series variation of

[8] In footnote 6, bivariate normality can be rewritten as a factor structure. In the panel data case, the possibilities are richer. In particular, we may have $\epsilon_{1it} = \lambda_1\theta_i + \nu_{1it}$ and $\epsilon_{2it} = \lambda_2\theta_i + \nu_{2it}$, where θ_i is an individual-specific factor. A similar idea appears in Cameron and Taber (2004). They impose a random effects assumption on θ_i. It may be possible to modify the identification argument I present to allow for this. Unfortunately, the right hand side of Equation 3.2.9 will become an integral that depends on the distribution of θ_i. An identification argument based on the large-sample limit of the likelihood function for observations where $(y_{1it}=0, y_{2it}=0)$ may be used. It is unclear whether we can avoid the distributional assumption on θ_i. Preliminary research by Khan, Maurel, and Zhang (2015) points to the identifying power of factor structures in triangular discrete response models.

every unit to decide whether $d_i = 0$ or $d_i = 1$. Notice that d is now allowed to vary across cross-sectional units. Data on the observed frequencies of $(y_{1it} = 0, y_{2it} = 1)$ for all t and i such that $d_i = 0$ point-identify α_1. Similarly, data on the observed frequencies of $(y_{1it} = 1, y_{2it} = 0)$ for all t and i such that $d_i = 1$ point-identify α_2.

What I have shown is that point identification may be possible in a model where both (y_1, y_2) are dummy endogenous variables jointly determined by the system (3.2.1) and (3.2.2) without imposing sign restrictions, as long as we have access to panel data. The approach considered here is slightly different from the entry-exit game estimated by Hahn and Moon (2010) using panel data. They impose sign restrictions (motivated by economic theory) producing an incomplete game-theoretic model. In their model, the econometrician is unable to observe which equilibrium was selected by the players but whichever equilibrium is selected becomes fixed across time. The approach considered here is also different from Hajivassiliou and Savignac (2011). They do not have fixed effects and they restrict error supports conditional on the restriction that both α_1 and α_2 must not have the same sign.

The intuition behind the identification argument in both the cross-section and panel data case is to find a subset of the data unaffected by the mixing probability p_i. I use this subset of the data to identify the common parameters except for α_1 and α_2. Deciding whether $d_i = 0$ or $d_i = 1$ depends on time series variation. After deciding the grouping, α_1 and α_2 can now be point identified. It is important to note that the number of groups is known in advance. Lewbel's (2007) characterization of complete and coherent two-equation systems ensure that the number of groups is fixed at 2.

There are extensions of situations that follow essentially the identification argument above. For instance, the identification argument can be extended to the case where we have an intercept, strictly exogenous regressors, and lagged dependent variables. Another possibility is for the error terms to have other known marginal distributions linked by a parametric copula as in Han and Vytlacil (2015) but extended to the panel data case. Finally, we can accommodate other discrete choice models such as the multinomial logit/probit and ordered logit/probit.

3.3 The model

3.3.1 Background

I now describe how to identify and estimate the parameters of the model to be used in the empirical application. HI (2007) construct a model of a household head living in finite time who chooses consumption and hours worked subject to a liquidity constraint and a quantity constraint on labor supply. As a consequence of liquidity constraints, household heads cannot hold negative wealth at any time over the life cycle. Furthermore, they may be subject to involuntary unemployment / underemployment, voluntary employment, or involuntary overemployment. As a result, they can be in one of the following situations – (a) they are able to work but are unable to reach their desired number of hours, (b) they are able to work at their desired number of hours, or (c) they are working beyond their desired level of hours. HI (2007) derive the solutions to the optimization problem faced by a representative household. The solutions to the optimization problem represent the optimal path of assets and hours worked over the life cycle of the household.

The econometrician is able to observe whether or not the household head is liquidity constrained, i.e., the liquidity constraint indicator S_{it} takes on the value 1 or 0, respectively. Household heads could be involuntarily overemployed ($E_{it} = -1$), voluntarily employed ($E_{it} = 0$), or involuntarily unemployed / underemployed ($E_{it} = 1$). The authors describe how these indicators were constructed in their earlier paper (see HI (1995)).

Since these optimal paths of assets and hours worked are determined jointly over the household's life cycle, the econometric treatment would have to acknowledge the underlying simultaneity. HI (2007) argue that one can either model the employment constraint indicator conditionally on the liquidity constraint indicator or vice versa. They further point out that this is consistent with the intertemporal two-stage budgeting of households described in Blundell and Walker (1986). Clearly, Lewbel's (2007) characterization can be exploited without us choosing the causal direction in advance or by presenting two sets of results.

HI (2007) specify their dynamic simultaneous equations model as follows:

$$\begin{aligned}
S_{it}^{*} &= \gamma_{11}S_{i,t-1} + \gamma_{12}S_{i,t-2} + \delta_0 E_{it} + \delta_1 E_{i,t-1} + \delta_2 E_{i,t-2} \\
&\quad + X_{1it}\beta^{bp} + \epsilon_{it}^{bp}, && (3.3.1)\\
E_{it}^{*} &= \gamma_{21}E_{i,t-1} + \gamma_{22}E_{i,t-2} + \kappa_0 S_{it} + \kappa_1 S_{i,t-1} + \kappa_2 S_{i,t-2} \\
&\quad + X_{2it}\beta^{op} + \epsilon_{it}^{op}, && (3.3.2)\\
S_{it} &= 1\left\{S_{it}^{*} \geq 0\right\}, \\
E_{it} &= -1\left\{E_{it}^{*} < \theta^{-}\right\} + 1\left\{E_{it}^{*} > \theta^{+}\right\}
\end{aligned}$$

where θ^{-} and θ^{+} are lower and upper thresholds. Just as in HI (2007), I also normalize $\theta^{+} = 0$. Observe that all the employment status indicators should really enter as two dummies because there are three categories. For example, $\delta_0 E_{it}$ can be decomposed into $\delta_{01}1\{E_{it} = -1\} + \delta_{02}1\{E_{it} = 1\}$. HI (1995; 2007) show that the coherency conditions are $(\delta_{01} + \delta_{02})\kappa_0 = 0$ and $\delta_{01}\delta_{02}\kappa_0 = 0$. As a result, we either have $\kappa_0 = 0$ or $\delta_{01} = \delta_{02} = 0$. If we exploit Lewbel's (2007) result, we need not impose these coherency conditions at all. Next, I show how to identify the parameters of their model without imposing coherency conditions.

3.3.2 Identification

To discuss the identification argument for the parameters in the model (3.3.1) and (3.3.2). I make the following assumptions:

A1 (Data generating process) For all i and t, S_{it} and E_{it} are generated by the model

$$
\begin{aligned}
S_{it}^{*} &= \gamma_{11}S_{i,t-1}+\gamma_{12}S_{i,t-2}+d_i\delta_{01}1\{E_{it}=-1\}+d_i\delta_{02}1\{E_{it}=1\} \\
&\quad +\delta_{11}1\left\{E_{i,t-1}=-1\right\}+\delta_{12}1\left\{E_{i,t-1}=1\right\}+\delta_{21}1\left\{E_{i,t-2}=-1\right\} \\
&\quad +\delta_{22}1\left\{E_{i,t-2}=1\right\}+X_{1it}\beta^{bp}+\epsilon_{it}^{bp}, \qquad (3.3.3)\\
E_{it}^{*} &= \gamma_{211}1\left\{E_{i,t-1}=-1\right\}+\gamma_{212}1\left\{E_{i,t-1}=1\right\}+\gamma_{221}1\left\{E_{i,t-2}=-1\right\} \\
&\quad +\gamma_{222}1\left\{E_{i,t-2}=1\right\}+(1-d_i)\kappa_0S_{it}+\kappa_1S_{i,t-1}+\kappa_2S_{i,t-2} \\
&\quad +X_{2it}\beta^{op}+\epsilon_{it}^{op} \qquad (3.3.4)\\
S_{it} &= 1\left\{S_{it}^{*}\geq 0\right\}, \\
E_{it} &= -1\left\{E_{it}^{*}<\theta^{-}\right\}+1\left\{E_{it}^{*}>0\right\},
\end{aligned}
$$

where S_{it}^{*} and E_{it}^{*} are latent variables. The parameters representing the simultaneous effects δ_{01}, δ_{02}, and κ_0 cannot all be jointly equal to zero.

The representation of the model in A1 is a result of applying Lewbel's (2007) characterization of a coherent and complete representation. Note that $X_{1it}\in\mathbb{R}^{p_1}$ and $X_{2it}\in\mathbb{R}^{p_2}$ may have common elements. The equation for S_{it}^{*} is a binary choice model while the equation for E_{it}^{*} is an ordered choice model. The superscripts bp and op refer to binary probability model and ordered probability model, respectively.

A2 (Exogeneity restrictions) Let

$$
\begin{aligned}
Z_i^t &= \left(S_{i,t-1},\ldots,S_{i0},S_{i,-1},E_{i,t-1},\ldots,E_{i0},E_{i,-1}\right),\\
Z_{it} &= \left(S_{i,t-1},S_{i,t-2},E_{i,t-1},E_{i,t-2}\right),
\end{aligned}
$$

$X_{1i}^T=\left(X_{1i,-t},X_{1it}\right)$, and $X_{2i}^T=\left(X_{2i,-t},X_{2it}\right)$. For all i and t, the error terms satisfy

$$
\left(\begin{array}{c}\epsilon_{it}^{bp}\\ \epsilon_{it}^{op}\end{array}\right)\Bigg|Z_i^t,X_{1i}^T,X_{2i}^T\sim\left(\begin{array}{c}\epsilon_{it}^{bp}\\ \epsilon_{it}^{op}\end{array}\right)\Bigg|Z_{it},X_{1it},X_{2it}.
$$

Assumption A2 establishes some notation adapted from the dynamic panel data and game theory literatures. The notation for X_{1i}^T splits $(X_{1i1},X_{1i2},\ldots,X_{1it},\ldots,X_{1iT})$ into a period t component X_{1it} and a component

$$
X_{1i,-t}=\left(X_{1i1},X_{1i2},\ldots,X_{1i,t-1},X_{1i,t+1}\ldots,X_{1iT}\right)
$$

representing all the other time periods except period t. I use the same notation for X_{2i}^T. Assumption A2 establishes that Z_{it} represents the predetermined regressors and X_{1it}, X_{2it} represent the strictly exogenous regressors.

A3 (Error distribution) The error terms are i.i.d. draws from the conditional distribution

$$\left.\begin{pmatrix} \epsilon_{it}^{bp} \\ \epsilon_{it}^{op} \end{pmatrix}\right| Z_{it}, X_{1it}, X_{2it} \sim C\left(F_{\epsilon^{bp}}\left(\epsilon^{bp}\right), F_{\epsilon^{op}}\left(\epsilon^{op}\right); \rho\right)$$

where $C(\cdot,\cdot;\rho)$ is a copula known up to a scalar parameter $\rho \in \Omega$ such that $C : (0,1)\times(0,1) \to (0,1)$ and Ω is an open subset of $\mathbb{R}$. The copula $C(u_1, u_2; \rho)$ is continuously differentiable everywhere in its domain $(u_1, u_2, \rho) \in (0,1) \times (0,1) \times \Omega$. $F_{\epsilon^{bp}}$ and $F_{\epsilon^{op}}$ are known marginal distribution functions for ϵ^{bp} and ϵ^{op}, respectively, that are strictly increasing, are absolutely continuous with respect to Lebesgue measure, and such that $\mathbb{E}\left(\epsilon^{bp}\right) = \mathbb{E}(\epsilon^{op}) = 0$ and $Var\left(\epsilon^{bp}\right) = Var(\epsilon^{op}) = 1$.

In contrast to the previous section where I imposed bivariate normality, I allow for a larger class of parametric models in A3. Furthermore, there is a large selection of copulas that are available (see the survey by Trivedi and Zimmer (2007), a textbook treatment by Nelsen (2006), and an application by Winkelmann (2012)). In contrast to Han and Vytlacil (2015), I do not impose any stochastic dominance assumptions on the selected copula. The assumptions on the marginal distributions $F_{\epsilon^{bp}}$ and $F_{\epsilon^{op}}$ are needed to ensure smoothness and invertibility. The restrictions on the moments of the error terms are typical normalizations in the discrete choice literature since the parameters are identified up to scale.

A4 (Finite support of fixed effects) The fixed effects d_i have known finite support $\{0,1\}$ for all i and are conditionally independent draws from some unknown distribution. Furthermore, $d_i \perp \left(\epsilon_{it}^{bp}, \epsilon_{it}^{op}\right) | S_{i0}, S_{i,-1}, E_{i0}, E_{i,-1}, X_{1i}^T, X_{2i}^T$ for all i and t.

Lewbel's (2007) characterization ensures that the support of the fixed effects is finite and has cardinality equal to 2. Assumption A4 is an assumption in the spirit of fixed-effects models. The independence assumption, however, is much stronger than the zero correlation between the fixed effect and the idiosyncratic error one usually encounters in linear panel data models.

A5 (Support and rank conditions) For all i and t, there exists some regressor (say the kth regressor) X_{1itk} with $\beta_k^{bp} \neq 0$ such that the distribution of $X_{1itk}|X_{1it,-k}$ has an everywhere positive Lebesgue density where

$$X_{1it,-k} = \left(X_{1it1}, \ldots, X_{1it,k-1}, X_{1it,k+1}, \ldots, X_{1itp_1}\right).$$

For all i and t, the regressors X_{1it} and X_{2it} have full column rank. Furthermore, for all i and t, we have

$$\begin{aligned}\Pr\left(\operatorname{supp}\left(X_{1it}\beta^{bp}\right)\cap\operatorname{supp}\left(X_{1it}\beta^{bp}+s_1\right)\right) &> 0,\\ \Pr\left(\operatorname{supp}\left(X_{2it}\beta^{op}\right)\cap\operatorname{supp}\left(X_{2it}\beta^{op}+s_2\right)\right) &> 0,\end{aligned}$$

where $s_1 \in \{\gamma_{11},\gamma_{12},\delta_{11},\delta_{12},\delta_{21},\delta_{22}\}$ and $s_2 \in \{\gamma_{211},\gamma_{212},\gamma_{221},\gamma_{222},\kappa_1,\kappa_2\}$.

Assumption A5 imposes full rank on the regressors. It also assumes the existence of a regressor with large support only in the binary choice model. As a result, this strictly exogenous regressor X_{1itk} has to be continuous. In contrast, Tamer (2003) requires the existence of a regressor with large support in any of the two equations. Finally, the last set of conditions in A5 ensures that we can identify the coefficients of the predetermined regressors .

Let us now summarize the steps made for the identification argument. Note that the probability that $S_{it}=0$, $E_{it}=0$ is unaffected by the presence of the fixed effect d_i, just like the stylized example in the previous section. Although the stylized example considers the case where both endogenous variables are binary, the intuition underlying the identification argument remains the same. To see this, first let W_{it}^{bp} and W_{it}^{op} be the value of the linear predictor excluding the contemporaneous endogenous variables in (3.3.3) and (3.3.4), respectively, i.e.,

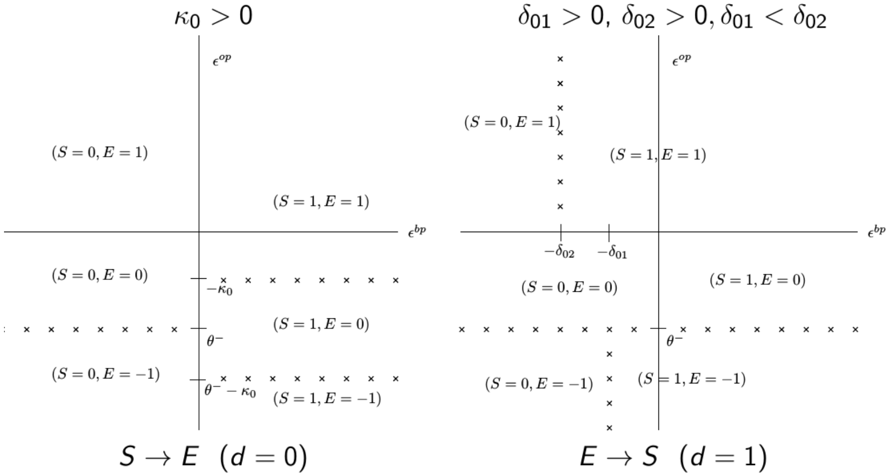
$$\begin{aligned}W_{it}^{bp} &= \gamma_{11}S_{i,t-1}+\gamma_{12}S_{i,t-2}+\delta_{11}1\left\{E_{i,t-1}=-1\right\}+\delta_{12}1\left\{E_{i,t-1}=1\right\}\\ &\quad +\delta_{21}1\left\{E_{i,t-2}=-1\right\}+\delta_{22}1\left\{E_{i,t-2}=1\right\}+X_{1it}\beta^{bp},\\ W_{it}^{op} &= \gamma_{211}1\left\{E_{i,t-1}=-1\right\}+\gamma_{212}1\left\{E_{i,t-1}=1\right\}+\gamma_{221}1\left\{E_{i,t-2}=-1\right\}\\ &\quad +\gamma_{222}1\left\{E_{i,t-2}=1\right\}+\kappa_1S_{i,t-1}+\kappa_2S_{i,t-2}+X_{2it}\beta^{op}.\end{aligned}$$

Next, we compute the probability that $S_{it}=0$, $E_{it}=0$ given the strictly exogenous regressors and predetermined regressors as follows:

$$\begin{aligned}&\Pr\left(S_{it}=0,E_{it}=0|Z_i^t,X_{1i}^T,X_{2i}^T\right)\\ &\stackrel{\mathbf{A1}}{=}\Pr\left(S_{it}^*\le 0,\theta^-\le E_{it}^*\le 0|Z_i^t,X_{1i}^T,X_{2i}^T\right)\\ &\stackrel{\mathbf{A1,A2,A4}}{=}\Pr\left(\epsilon_{it}^{bp}\le -W_{it}^{bp},\theta^- - W_{it}^{op}\le\epsilon_{it}^{op}\le -W_{it}^{op}\right)\\ &\stackrel{\mathbf{A3}}{=}\Pr\left(F_{\epsilon^{bp}}\left(\epsilon_{it}^{bp}\right)\le F_{\epsilon^{bp}}\left(-W_{it}^{bp}\right),F_{\epsilon^{op}}\left(\theta^- - W_{it}^{op}\right)\le F_{\epsilon^{op}}\left(\epsilon_{it}^{op}\right)\le F_{\epsilon^{op}}\left(-W_{it}^{op}\right)\right)\\ &\stackrel{\mathbf{A3}}{=}C\left(F_{\epsilon^{bp}}\left(-W_{it}^{bp}\right),F_{\epsilon^{op}}\left(-W_{it}^{op}\right);\rho\right)-C\left(F_{\epsilon^{bp}}\left(-W_{it}^{bp}\right),F_{\epsilon^{op}}\left(\theta^- - W_{it}^{op}\right);\rho\right).\end{aligned} \tag{3.3.5}$$

The probability computed in (3.3.5) is always positive since $\theta^- < 0$. Figure 3.3.1 confirms the calculation made in (3.3.5). In the figure, think of the origin as the ordered pair of linear predictors $\left(W_{it}^{bp},W_{it}^{op}\right)$. The probability mass over the region where we have $S_{it} = 0$, $E_{it} = 0$ is unaffected by the presence of the fixed effect given the assumptions I have imposed. Further observe that (3.3.5) can be thought of as a binary choice model where the outcomes are either the event $S_{it} = 0$, $E_{it} = 0$ or the event where $(S_{it} = 0, E_{it} = -1)$, $(S_{it} = 0, E_{it} = 1)$, $(S_{it} = 1, E_{it} = -1)$, or $(S_{it} = 1, E_{it} = 1)$, i.e. all the other configurations of (S_{it}, E_{it}).

Figure 3.3.1: An illustration of a case of (3.3.3) and (3.3.4)



The steps below summarize the identification argument. Steps 3 to 7 follow an argument similar to the stylized example in the previous section. The only steps that are new are the first two steps which account for how we will identify the coefficients of the lagged dependent variables and the coefficients of the strictly exogenous variables. The full details of the argument can be found in the Appendix.

Step 1. This step is the nonconstructive part of the identification argument. Take two points $\left(x_1^-, x_1\right) \in \operatorname{supp}\left(X_{1i}^T\right)$, $\left(x_2^-, x_2\right) \in \operatorname{supp}\left(X_{2i}^T\right)$, and $z \in \operatorname{supp}\left(Z_i^{t-1}\right)$. Collect the observed frequencies of $S_{it} = 0$, $E_{it} = 0$ conditional on $Z_i^{t-1} = z$, $Z_{it} = (0,0,0,0)$, $X_{1i}^T = \left(x_1^-, x_1\right)$, $X_{2i}^T = \left(x_2^-, x_2\right)$. Use an identification at infinity argument like the one used by Tamer (2003) to identify β^{bp} and β^{op}. We also identify θ^- in this step.

Step 2. Take another point $\tilde{x}_1 \in \operatorname{supp}(X_{1it})$. We use Manski's (1985; 1988) argument to identify (γ_{11}, κ_1) using the observed frequencies of $S_{it} = 0$, $E_{it} = 0$ conditional on $Z_{it} = (1,0,0,0)$, $X_{1it} = \tilde{x}_1$, $X_{2it} = \tilde{x}_2$ and the observed frequencies of $S_{it} = 0$, $E_{it} = 0$ conditional on $Z_{it} = (0,0,0,0)$, $X_{1it} = x_1$, $X_{2it} = x_2$. Repeat the argument to identify the coefficients of the other lagged dependent variables using the appropriate Z_{it}, i.e. (γ_{12}, κ_2), $(\delta_{11}, \gamma_{211})$, $(\delta_{12}, \gamma_{212})$, $(\delta_{21}, \gamma_{221})$, and $(\delta_{22}, \gamma_{222})$. For instance, we should set $Z_{it} = (0,0,1,0)$ to identify $(\delta_{12}, \gamma_{212})$.

Step 3. Since (3.3.5) has the form of a fully parametric binary choice model as discussed earlier, the copula dependence parameter ρ can be identified immediately since the values of the parameters in Steps 1 and 2 are identified and can be taken as known.

Step 4. The signs of δ_{01}, δ_{02}, and κ_0 can now be identified.

Step 5. All the information from the previous steps can now be used to determine whether $d_i = 0$ or $d_i = 1$.

Step 6. Since the groupings are now identified, we can recover the values of δ_{01}, δ_{02}, and κ_0.

Steps 1 to 3 of the preceding argument can also be replaced by an alternative argument where we exploit the form of (3.3.5). (3.3.5) is a fully parametric binary choice model and a likelihood function formed from pooling all the cross-sectional and time series information can be used to identify all the parameters mentioned in Steps 1 to 3. This alternative avoids the rather nonconstructive nature of Step 1. Presenting the argument in these two ways is to set the stage for future work on weakening some of the parametric assumptions in Assumption A3. Furthermore, these two arguments may have different implications for estimation and inference. What I have shown is that point identification of all the common parameters is possible for (3.3.3) and (3.3.4).

3.3.3 Estimation and inference

Even though there would be incidental parameter bias when T is fixed and small, Hahn and Moon (2010) show that the incidental parameter bias disappears at a much faster rate. Although the context they have in mind is estimating a game-theoretic model where the fixed effect represents the equilibrium chosen by players (when there are multiple equilibria), the idea that the fixed effect takes only a finite number of values applies to my proposal.

In particular, they show that, under certain regularity conditions, the reduction in support is automatically bias-reducing under an asymptotic scheme where $n, T \to \infty$ with n typically growing as an exponential function of T. Since the asymptotic distribution of the MLE no longer has a noncentrality parameter (as opposed to the usual case where individual-specific effects are allowed to have full support over the real line; see Hahn and Kuersteiner (2011)), inferences can be justified without resorting to bias-reduction procedures.

In this subsection, we impose the Gaussian copula for C and standard normal cumulative distribution functions for the margins $F_{\epsilon^{bp}}$ and $F_{\epsilon^{op}}$ in Assumption A3, just as in HI (2007). As a result, the dependence parameter $\rho \in (-1, 1)$ coincides with the usual correlation coefficient of the bivariate normal distribution. I still impose all the assumptions required for identification here in this subsection. I use maximum likelihood for estimation and inference. Collect all the common parameters into a vector $\lambda \in \Lambda$ and treat $d_i \in \{0,1\}$ as a parameter to be estimated. As a result, the log-likelihood for an arbitrary i and t is given by

$$l_{it}(\lambda, d_i) = \sum_{j\in\{0,1\}} \sum_{k\in\{-1,0,1\}} \mathbf{1}(S_{it} = j, E_{it} = k) \log \Pr\left(S_{it} = j, E_{it} = k | Z_i^t, X_{1i}^T, X_{2i}^T; \lambda, d_i\right).$$

Aggregating over time for a fixed cross-sectional unit gives us the log-likelihood for the ith cross-sectional unit:

$$l_i(\lambda, d_i) = \sum_{t=1}^{T} l_{it}(\lambda, d_i).$$

Next, I impose the following additional assumptions:

E1 The parameters representing the simultaneous effects δ_{01}, δ_{02}, and κ_0 cannot all be jointly equal to zero.

E2 Let $y_{it} = \left(S_{it}, E_{it}, S_{i,t-1}, E_{i,t-1}, S_{i,t-2}, E_{i,t-2}, X_{1it}, X_{2it}\right)$ be the data for the ith unit and tth time period and $d_{i0} \in \{0,1\}$ be the true value of d_i. For each i, $\{y_{it} : t = 1, 2, \ldots\}$ is strictly stationary. The differences of the joint distribution of $\{y_{i1}, y_{i2}, \ldots\}$ across i is completely characterized by d_{i0}.

E3 Let

$$\varepsilon^* = \inf_i \left[G_{(i)}(\lambda_0, d_{i0}) - \sup_{\{d_i \neq d_{i0}\}} G_{(i)}(\lambda, d_i) \right] > 0,$$

where

$$G_{(i)}(\lambda, d) = \mathbb{E}_{(\theta_0, d_{i0})}\left[l_i(\lambda, d)\right].$$

For all $\eta > 0$,

$$\inf_i \left[G_{(i)}(\lambda_0, d_{i0}) - \sup_{\{|\lambda-\lambda_0|>\eta, d\}} G_{(i)}(\lambda, d) \right] > 0.$$

The parameter space Λ is compact. There exists some $M(y_{it})$ such that

$$\sup_{\lambda, d} \left| \frac{\partial^k l_{it}(\lambda, d_i)}{\partial \lambda^k} \right| \leq M(y_{it})$$

for $k = 0, 1$ and $\max_i \mathbb{E}\left[M(y_{it})\right]^2 < \infty$.

E4 Let $\varepsilon > 0$, $\eta > 0$ and θ be given. There exists some $h(T)$ strictly increasing in T, such that, for all $\left(d_i, d_i'\right)$ combinations, we have

$$\begin{aligned}
\Pr\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(l_{it}\left(\lambda, d_i\right) - \mathbb{E}\left[l_{it}\left(\lambda, d_i\right)\right]\right)\right| > \frac{\eta}{3}\right] &= o\left(\frac{1}{h(T)}\right), \\
\Pr\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(M\left(y_{it}\right) - \mathbb{E}\left[M\left(y_{it}\right)\right]\right)\right| > \frac{\eta}{3\varepsilon}\right] &= o\left(\frac{1}{h(T)}\right),
\end{aligned}$$

where the probability and the expectation are calculated with respect to the density of $(y_{i1}, \ldots, y_{iT})$ indexed by $\left(\lambda_0, d_i'\right)$.

Note that the individual-specific likelihood function under $d_i = 0$ becomes automatically distinguishable from the one under $d_i = 1$ provided that assumption E1 holds. If all these parameters representing simultaneous effects are jointly equal to zero, there is no way to use time series variation to differentiate between $d_i = 0$ and $d_i = 1$. This can be seen easily from Figure (3.3.1). Assumption E3 also holds because of the previous statements along with the point-identification result in the previous subsection. The compactness of Λ and the boundedness conditions on the likelihood and its score are standard regularity conditions imposed in maximum likelihood estimation. Note that the log-likelihood I consider are continuously differentiable over the compact parameter space. Furthermore, the parametric forms and time homogeneity assumed for the model in (3.3.3) and (3.3.4) ensures that the data for every cross-sectional unit are strictly stationary which satisfies Assumption E2. Finally, Assumption E4 is a technical condition required to identify the correct group assignment. This assumption has been used in the literature on discrete parameter models (refer to Choirat and Seri (2012) and its references). Hahn and Moon (2010) and Choirat and Seri (2012) show that $h(T)$ is typically an exponential function of T.

Since Assumptions E2 to E4 are the same conditions used by Hahn and Moon (2010), adapting their Theorem 1 gives us:

Theorem 3.3.1. *Let*

$$\begin{aligned}
\widehat{d_i}(\lambda) &= arg\max\left\{l_i\left(\lambda, d_i = 1\right), l_i\left(\lambda, d_i = 0\right)\right\}, \\
\widehat{\lambda} &= arg\max_{\lambda}\sum_{i=1}^{n}\sum_{t=1}^{T} l_{it}\left(\lambda, \widehat{d_i}(\lambda)\right), \\
\widetilde{\lambda} &= arg\max_{\lambda}\sum_{i=1}^{n}\sum_{t=1}^{T} l_{it}\left(\lambda, d_{i0}\right).
\end{aligned}$$

Suppose that $\sqrt{nT}\left(\widetilde{\lambda} - \lambda\right) \xrightarrow{d} N(0, \Sigma)$ *for some* Σ. *Under Assumptions E1 to E4, we have* $\sqrt{nT}\left(\widehat{\lambda} - \lambda\right) \xrightarrow{d} N(0, \Sigma)$ *if* $n \to \infty$ *and* $T \to \infty$ *such that* $n = O(h(T))$.

The theorem states that the substitution of a plug-in $\widehat{d_i}(\lambda)$ for d_i is asymptotically negligible. The covariance matrix Σ can either be the inverse of the Hessian or the covariance matrix based on the sandwich formula. To estimate all the common parameters, I use the following iterative approach:[9]

1. Set $s = 0$. Fix starting points for λ at $\lambda^{(0)}$.
2. Let $l_i(\lambda, d_i)$ be the log-likelihood for the ith unit. If $l_i\left(\lambda^{(s)}, 1\right) > l_i\left(\lambda^{(s)}, 0\right)$, then we set $\widehat{d_i^{(s)}} = 1$. Otherwise, $\widehat{d_i^{(s)}} = 0$.
3. Find the maximizer of $\sum_i l_i\left(\lambda^{(s)}, \widehat{d_i^{(s)}}\right)$ and call it $\theta^{(s+1)}$.
4. Set s to be $s + 1$. Repeat Steps 2 and 3 until convergence.

Note that Step 2 corresponds to the profiling out of the fixed effects and that Step 3 corresponds to finding the maximizer of the profile likelihood. The zigzag method proposed is slightly slow in the application because I have to estimate around 54 to 75 parameters. However, Step 2 is likely to be faster than the case where the fixed effect could take on any value.

3.4 Revisiting the results of HI (1995; 2007)

3.4.1 Similarities and differences

Using PSID data[10] from Waves 1 to 20, the authors estimate an econometric model based on the simultaneous determination of (S_{it}, E_{it}) as seen in (3.3.1) and (3.3.2). They estimate both a binary probit and an ordered probit model where both indicators are jointly determined. Both their 1995 and 2007 papers impose the coherency conditions $\kappa_0 = 0$ or $\delta_{01} = \delta_{02} = 0$. Therefore, they will have two sets of results – a set of results based on $\delta_{01} = \delta_{02} = 0$ and another based on $\kappa_0 = 0$. In contrast, I jointly estimate (3.3.3) and (3.3.4) without imposing the coherency conditions.

They incorporate dynamic effects in the model by introducing lagged values of the corresponding indicators. The other regressors are variables that represent characteristics of the household head and the labor market to which the household head was exposed. The list of regressors used in both the 1995 and 2007 papers can be found in Table 3.6.2 found in the Appendix. I exclude the cube of age in the list of regressors because the resulting Hessian was singular.[11] The 1995 paper makes use of exclusion restrictions when estimating (3.3.1) and (3.3.2). On the other hand,

[9] The algorithm is not exactly an application of the EM algorithm. The fixed effects d_i I introduce into the model are not just labels. The values that d_i take have a direct interpretation.

[10] I use data made available by the authors in the Journal of Applied Econometrics data archive.

[11] Freedman and Sekhon (2010) document some of the numerical issues involved in estimating systems of equations with discrete endogenous variables even with very few regressors.

the 2007 paper does not have any exclusion restrictions at all. For instance, the age of the household head may influence both S_{it} and E_{it}, but being a union member may influence E_{it} but not S_{it}. Meango and Mourifie (2013) show that some parameters are only partially identified in two-equation probit models with a dummy endogenous regressor when there are no exclusion restrictions. In contrast, I apply the exclusion restrictions in HI (1995).[12]

In addition to (3.3.1) and (3.3.2), they assume that the error terms $\left(\epsilon_{it}^{bp}, \epsilon_{it}^{op}\right)$ have an AR(1) structure:

$$\begin{aligned} \epsilon_{it}^{bp} &= \eta_i^{bp} + \varsigma_{it}^{bp},\ \varsigma_{it}^{bp} = \rho^{bp}\varsigma_{i,t-1}^{bp} + \xi_{it}^{bp},\ \left|\rho^{bp}\right| < 1, && (3.4.1) \\ \epsilon_{it}^{op} &= \eta_i^{op} + \varsigma_{it}^{op},\ \varsigma_{it}^{op} = \rho^{op}\varsigma_{i,t-1}^{op} + \xi_{it}^{op},\ \left|\rho^{op}\right| < 1, && (3.4.2) \end{aligned}$$

where $\left(\eta_i^{bp}, \eta_i^{op}\right)$ represent time-invariant unobserved heterogeneity and $\left(\xi_{it}^{bp}, \xi_{it}^{op}\right)$ are both i.i.d. Gaussian random variables with mean zero, variance 1 and have nonzero correlation ρ conditional on the strictly exogenous regressors $X_i = (X_{i1}, \ldots, X_{iT})$. They use the Mundlak-Chamberlain device to model $\left(\eta_i^{bp}, \eta_i^{op}\right)$. In particular, they assume that $\eta_i^{bp}|X_i \sim N\left(\bar{X}_{i.}\theta^{bp}, \sigma_{bp}^2\right)$ and $\eta_i^{op}|X_i \sim N\left(\bar{X}_{i.}\theta^{op}, \sigma_{op}^2\right)$. They also model the initial conditions using an analogous assumption. In contrast, I use d_i to represent time-invariant unobserved heterogeneity that may be arbitrarily correlated with X_i. All my results are conditional on the initial observation. I allowed for a similar AR(1) structure but results were not forthcoming as will be discussed below.

Strictly speaking, the model I consider neither encompasses nor generalizes the model that HI (1995; 2007) consider. HI (1995; 2007) use large-n, fixed-T asymptotics to justify their results. In contrast, I use large-n, large-T asymptotics to justify my results. Introducing additive fixed effects and modelling these effects using the Mundlak-Chamberlain device may require substantial changes to the identification argument and the justification of the estimation procedure. Even without resorting to the Mundlak-Chamberlain device, it is not clear how to justify existing bias reduction procedures meant for panel data models with fixed effects that have full support. Despite these concerns, the results I present point to the conclusion that imposing coherency conditions may be inappropriate.

3.4.2 Results

There are a total of 32408 observations on 2410 male household heads observed for an average number of 14 periods. Complete spells accounted for 528 out of the 2410 male household heads. I exclude all household heads with spells of length 1 in

[12]Nevertheless, I estimate the model without the exclusion restrictions because recent work by Han and Vytlacil (2015) point to the possibility of point identification even if there are common exogenous regressors and there are no exclusion restrictions. For now, the identification argument in Section 3.3 require exclusion restrictions. The regressors with large support in the binary choice equation that HI (1995; 2007) use are food needs (`fneed`) and the growth of food needs (`gfneed`).

the analysis. Table 3.6.1 in the Appendix has the distribution of spell length for the household heads in the sample.

I compute all the results in this section using `R` (R Core Team, 2014). I use the `optimx` package, the accompanying BFGS algorithm, and the programmed tests for the Karush-Kuhn-Tucker optimality conditions (see Nash and Varadhan (2011) and Nash (2014) for more details). All results presented below have passed these tests. I use estimates found in HI (1995; 2007) as possible starting points for the algorithm.[13]

Given the discussion in the previous subsection, I estimate 3 different specifications:

1. Specification A uses the list of regressors in HI (1995) but only allows for own state dependence, i.e. $\delta_{11} = \delta_{12} = \delta_{21} = \delta_{22} = 0$ in (3.3.3) and $\kappa_1 = \kappa_2 = 0$ in (3.3.4). There are 56 parameters to be estimated in this case.

2. Specification B uses the list of regressors in HI (1995) but removes the restrictions just mentioned. As a result, I account for some form of spillover effect. There are 62 parameters to be estimated in this case.

3. Specification C uses the same set of regressors in both (3.3.3) and (3.3.4). The restrictions just mentioned are also removed. There are 73 parameters to be estimated in this case.

Furthermore, I consider two samples – Sample 1 consists of observations from Waves 1 to 14 while Sample 2 consists of observations from Waves 1 to 20. We present the coefficient estimate and their corresponding standard errors immediately below. Coefficients that are statistically significant at the 1% level are in **bold**.

Table 3.4.1: Main Results

	Specification A		Specification B		Specification C	
	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
δ_{01}	**0.693**	**0.293**	**0.489**	**0.283**	**0.489**	**0.309**
	(0.084)	(0.057)	(0.078)	(0.058)	(0.078)	(0.058)
δ_{02}	0.042	**−0.402**	**−0.721**	**−0.612**	**−0.770**	**−0.601**
	(0.053)	(0.036)	(0.054)	(0.041)	(0.055)	(0.040)
κ_0	**1.326**	**1.128**	**1.429**	**1.282**	**1.413**	**1.281**
	(0.034)	(0.025)	(0.036)	(0.028)	(0.036)	(0.028)
θ^-	**−5.697**	**−5.645**	**−5.714**	**−5.668**	**−5.719**	**−5.680**
	(0.008)	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)
ρ	−0.007	0.001	0.000	−0.001	0.001	−0.000
	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)

[13] These starting points are based on Tables 8 and 9 of HI (1995) and Tables VI and VII of HI (2007). I find that the results are not sensitive to these starting points.

I impose the AR(1) error structure used by the authors. However, the estimated first-order autocorrelation coefficients are extremely small (with sizes around 10^{-6}).[14] In contrast, the authors found first-order autocorrelation coefficients around the range of 0.40 to 0.68 and are significantly different from zero. Therefore, I set aside the AR(1) structure for the rest of the calculations.

The results in Table 3.4.1 indicate that imposing the coherency condition may not be appropriate. Note that δ_{01}, δ_{02}, and κ_0 are significantly different from zero across all specifications and samples (except for the one found in Specification A, Sample 1). Furthermore, the signs are very different from their results. For instance, the immediate effect of being involuntarily unemployed/underemployed on the probability of being liquidity constrained is negative while they estimate it as positive. The absolute values of the coefficients are much larger compared to the results by the authors. For instance, their estimates of κ_0 range from 0.12 to 0.13. There are some differences in the estimates for δ_{01} and δ_{02} in Specification A relative to the other specifications because Specification A does not include lagged spillover effects. There might also be indications of parameter nonconstancy as one moves from Sample 1 to Sample 2. Nevertheless, the results are qualitatively unchanged.

Table 3.4.2: Results on the effects of state dependence for Specification A

	Equation for S_{it}		Equation for E_{it}	
	Sample 1	Sample 2	Sample 1	Sample 2
$S_{i,t-1}$	**1.512**	**1.535**		
	(0.031)	(0.025)		
$S_{i,t-2}$	**0.313**	**0.459**		
	(0.031)	(0.025)		
$\mathbf{1}\{E_{i,t-1}=-1\}$			**−1.868**	**−1.885**
			(0.051)	(0.040)
$\mathbf{1}\{E_{i,t-2}=-1\}$			**−0.811**	**−0.883**
			(0.053)	(0.042)
$\mathbf{1}\{E_{i,t-1}=1\}$			**0.923**	**0.983**
			(0.028)	(0.021)
$\mathbf{1}\{E_{i,t-2}=1\}$			**0.523**	**0.539**
			(0.028)	(0.022)

It is clear from Table 3.4.1 that the coherency conditions imposed by HI (1995; 2007) are unlikely to be true for all household heads. The estimated lower threshold associated with involuntary unemployment/underemployment relative to voluntary is twice the value estimated by the authors. This means that the lower threshold is not as tight as HI (1995; 2007) estimate. The estimated correlation ρ between

[14] Even with different starting values, as noted in the preceding footnote, the estimates for the first-order autocorrelation coefficients are also very near zero.

the error terms $\left(\epsilon_{it}^{bp}, \epsilon_{it}^{op}\right)$ is not significantly different from zero, while the authors estimate this correlation at around 0.34 to 0.43 and are significantly different from zero. It is may be possible that the nonzero correlation of the error terms estimated by HI (1995; 2007) is an artifact of imposing the coherency conditions.

The results in Tables 3.4.2, 3.4.3, and 3.4.4 indicate that there are statistically significant effects of state dependence. In particular, the existence of own state dependence is a major feature common to the three tables. As a result, household heads that were liquidity constrained in the previous periods are more likely to be liquidity constrained now. I find a similar result for employment status. In particular, household heads who were overemployed in previous periods are more likely to be overemployed now.

Tables 3.4.3 and 3.4.4 indicate the possibility of lagged spillover effects, especially for employment status.[15] In particular, household heads that were liquidity constrained in previous periods are more likely to be overemployed now. The results also indicate that past employment status (except for household heads who were involuntarily unemployed/underemployed one period ago) may not be a significant indicator for determining whether household heads are more or less likely to be liquidity constrained now. Since being liquidity constrained or being involuntarily overemployed or unemployed for two periods in the past still has a significant effect on the current state of the household head, even after controlling for contemporaneous effects, the results paint quite a negative picture of the lasting effects of liquidity and employment constraints.

Table 3.4.3: Results on the effects of state dependence for Specification B

	Equation for S_{it}		Equation for E_{it}	
	Sample 1	Sample 2	Sample 1	Sample 2
$S_{i,t-1}$	**1.496**	**1.519**	**−0.385**	**−0.385**
	(0.031)	(0.025)	(0.038)	(0.030)
$S_{i,t-2}$	**0.297**	**0.444**	**−0.128**	**−0.115**
	(0.031)	(0.025)	(0.035)	(0.028)
$\mathbf{1}\left\{E_{i,t-1} = -1\right\}$	−0.018	−0.013	**−1.856**	**−1.876**
	(0.059)	(0.046)	(0.051)	(0.040)
$\mathbf{1}\left\{E_{i,t-2} = -1\right\}$	−0.097	−0.042	**−0.821**	**−0.882**
	(0.058)	(0.046)	(0.053)	(0.042)
$\mathbf{1}\left\{E_{i,t-1} = 1\right\}$	**0.124**	**0.123**	**0.941**	**1.008**
	(0.032)	(0.025)	(0.028)	(0.021)
$\mathbf{1}\left\{E_{i,t-2} = 1\right\}$	0.046	0.059	**0.529**	**0.561**
	(0.032)	(0.025)	(0.029)	(0.022)

[15] Note that these lagged spillover effects do not exactly represent the absence of Granger non-causality since the model includes contemporaneous terms.

Table 3.4.4: Results on the effects of state dependence for Specification C

	Equation for S_{it}		Equation for E_{it}	
	Sample 1	Sample 2	Sample 1	Sample 2
$S_{i,t-1}$	**1.480**	**1.507**	**−0.382**	**−0.359**
	(0.031)	(0.025)	(0.038)	(0.030)
$S_{i,t-2}$	**0.291**	**0.441**	**−0.134**	**−0.112**
	(0.031)	(0.025)	(0.035)	(0.028)
$\mathbf{1}\{E_{i,t-1}=-1\}$	−0.012	−0.010	**−1.856**	**−1.870**
	(0.059)	(0.047)	(0.051)	(0.040)
$\mathbf{1}\{E_{i,t-2}=-1\}$	−0.095	−0.041	**−0.814**	**−0.868**
	(0.058)	(0.046)	(0.053)	(0.042)
$\mathbf{1}\{E_{i,t-1}=1\}$	**0.109**	**0.101**	**0.939**	**1.009**
	(0.033)	(0.026)	(0.028)	(0.021)
$\mathbf{1}\{E_{i,t-2}=1\}$	0.039	0.051	**0.526**	**0.572**
	(0.033)	(0.025)	(0.029)	(0.022)

Note that the estimates in the preceding tables are not directly interpretable. Marginal effects would have to be computed and I leave this to future work. I conjecture that no bias correction would be required when estimating marginal effects unlike the case where the fixed effects have full support (for example, see Bester and Hansen (2009a)). Furthermore, the estimators for these marginal effects may be much slower when the fixed effects have full support, as documented by Fernandez-Val and Weidner (2013). It is unclear whether this will be the case when the fixed effects have finite support.

An alternative to estimating marginal effects is to estimate the ratio of the coefficient estimates. The ratio of the coefficient estimates is usually the ratio of marginal effects. Stewart (2004) show in the context of an ordered probit model that the ratios of the coefficient estimates can be interpreted as slopes of indifference curves. If we apply the idea to my context, the slope of this indifference curve represents the required tradeoff in one regressor so that a change in a different regressor will not alter the state of the household head. Unfortunately, these ratios cannot be obtained from the ratios of marginal effects because the probability in (3.3.5) is a joint probability and involves a difference of two probabilities.

The estimated fixed effects $\widehat{d_i}$ can also be obtained and be used to describe which of the household heads have a particular direction of causality. I calculate the estimated distribution of the fixed effects in Table 3.4.5. Compared to Hajivassiliou and Ioannides (1995; 2007), either all 2410 males have only a single direction of causality (say from S_{it} to E_{it}) or all of them have the other direction. Since we allow for the direction of causality to vary across males, we are able to count how many of these household heads have a pattern where S_{it} affects E_{it} and vice-versa. I find that around half of the 2410 males have a pattern where S_{it} affects E_{it} across specifications and across different samples. I also find that some of these males change patterns from Sample 1 to Sample 2. In particular, around 12% of the males from Sample 1 change patterns once we observed them for more time periods.

Table 3.4.5: Estimated distribution of the fixed effects $\widehat{d_i}$

	Specification A		Specification B		Specification C	
	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
$S_{it} \to E_{it}$	1101	1214	1100	1221	1103	1209
$E_{it} \to S_{it}$	1001	1186	1002	1179	999	1191
Total	2102	2400	2102	2400	2102	2400

Finally, the tables in the appendix contain results for the coefficients of the strictly exogenous regressors across different specifications and samples. Apart from a few coefficients changing signs across specifications and samples (in particular, household age and some of the dummies representing residence, ethnicity, and religion), the results are quite similar to one another and are consistent with expectations. However, most of the positive coefficients in Sample 1 are larger than those in Sample 2. Similarly, most of the negative coefficients in Sample 1 are larger in absolute value than those in Sample 2.

3.5 Concluding remarks

In this chapter, I have developed a route toward identification, estimation, and inference in dynamic simultaneous equations models with discrete outcomes when panel data is available. These models are subject to the incidental parameter problem when individual-specific fixed effects are included and are also subject to incoherence and incompleteness. I introduce a specific type of individual-specific fixed effect so that the coherency condition need not be imposed across all observations. This proposal allows us to avoid imposing sign restrictions or to avoid restricting error supports.

Specifically, I use a subset of the observables unaffected by the individual-specific fixed effect to identify the common parameters of the model. I then use time series variation to identify the individual-specific fixed effect. This fixed effect represents the direction of causality from one endogenous variable to another. Knowing the direction allows us to identify the coefficients of the endogenous variables. Consistent estimation and correct inference without any need for bias reduction follows from the large-n, large-T asymptotic theory. I revisit the empirical application of Hajivassiliou and Ioannides (1995; 2007) and find strikingly different results with respect to the contemporaneous interaction and dynamic structure of employment status and liquidity constraints.

Future work may consider the computation of certain types of marginal effects defined in Lewbel, Dong, and Yang (2012). Future work in this area may also allow the specification of individual-specific effects to be time-varying as well, just as Bonhomme and Manresa (2015) do for groupings in the linear model. There seems to be some slight evidence in the empirical application that cast some doubt on the assumption that the direction of causality is time-invariant. However, it may well be the case that we have restricted time-invariant unobserved heterogeneity too much. Introducing another fixed effect in the linear predictor may be fruitful but is beyond the scope of this chapter. Although the approach would seem fruitful, bias-reduction procedures have to be adapted for the case I considered. A natural alternative would be to use each cross-section to set-identify the parameters of the model (as seen in Section 2) and find methods to combine these set estimates across different time periods.

3.6 Appendix

Details of identification argument in Section 3.3.2

Let us now examine the details behind each step of the identification argument. In Step 1, we need to calculate the probability that $S_{it} = 0$ and $E_{it} = 0$ conditional on $Z_{it} = (0,0,0,0), X_{1i}^T = \left(x_1^-, x_1\right), X_{2i}^T = \left(x_2^-, x_2\right)$:

$$
\begin{aligned}
&\Pr\left(S_{it}^* \leq 0, \theta^- \leq E_{it}^* \leq 0 | Z_i^{t-1} = z, Z_{it} = (0,0,0,0), X_{1i}^T = \left(x_1^-, x_1\right), X_{2i}^T = \left(x_2^-, x_2\right)\right) \\
&\overset{\mathbf{A2}}{=} \Pr\left(S_{it}^* \leq 0, \theta^- \leq E_{it}^* \leq 0 | Z_{it} = (0,0,0,0), X_{1it} = x_1, X_{2it} = x_2\right) \\
&\overset{\mathbf{A1}}{=} \Pr\left(\epsilon_{it}^{bp} \leq -x_1\beta^{bp}, \theta^- - x_2\beta^{op} \leq \epsilon_{it}^{op} \leq -x_2\beta^{op}\right)
\end{aligned}
$$

Let $\left(\tilde{\beta}^{bp}, \tilde{\beta}^{op}\right)$ be such that $\left(\beta^{bp}, \beta^{op}\right) \neq \left(\tilde{\beta}^{bp}, \tilde{\beta}^{op}\right)$. Without loss of generality, let x_{1k} be the kth regressor in x_1 and $\beta_k^{bp}, \tilde{\beta}_k^{bp} > 0$ be the associated coefficient of this regressor. As $x_{1k} \to -\infty$ given the other regressors in x_1, we have $-x_{1k}\beta_k^{bp}$, $-x_{1k}\tilde{\beta}_k^{bp} \to \infty$. Since x_2 has full rank by A5, we have x_2 such that $x_2\beta^{op} \neq x_2\tilde{\beta}^{op}$. We now have

$$
\begin{aligned}
&\Pr\left(\epsilon_{it}^{bp} \leq -x_1\beta^{bp}, \theta^- - x_2\beta^{op} \leq \epsilon_{it}^{op} \leq -x_2\beta^{op}\right) \\
&\approx \Pr\left(\theta^- - x_2\beta^{op} \leq \epsilon_{it}^{op} \leq -x_2\beta^{op}\right) \\
&\neq \Pr\left(\theta^- - x_2\tilde{\beta}^{op} \leq \epsilon_{it}^{op} \leq -x_2\tilde{\beta}^{op}\right) \\
&\approx \Pr\left(\epsilon_{it}^{bp} \leq -x_1\beta^{bp}, \theta^- - x_2\tilde{\beta}^{op} \leq \epsilon_{it}^{op} \leq -x_2\tilde{\beta}^{op}\right).
\end{aligned}
$$

As a result, β^{op} is identified. Since x_1 has full rank by A5, we have x_1 such that $x_1\beta^{bp} \neq x_1\tilde{\beta}^{bp}$. Following the same argument as before, we have

$$\begin{aligned}&\Pr\left(\epsilon_{it}^{bp} \leq -x_1\beta^{bp}, \theta^- - x_2\beta^{op} \leq \epsilon_{it}^{op} \leq -x_2\beta^{op}\right)\\ &\neq \Pr\left(\epsilon_{it}^{bp} \leq -x_1\tilde{\beta}^{bp}, \theta^- - x_2\beta^{op} \leq \epsilon_{it}^{op} \leq -x_2\beta^{op}\right).\end{aligned}$$

As a result, β^{bp} is identified. For the case where $\tilde{\beta}_k^{bp} < 0$, we have $-x_{1k}\beta_k^{bp} \to \infty$ but $-x_{1k}\tilde{\beta}_k^{bp} \to -\infty$. Following the same argument as before, we can identify both β^{bp} and β^{op}. Note that the constant terms in β^{bp} and β^{op} are also identified.

Now, we identify θ^-. Without loss of generality, let $\tilde{\theta}^- < \theta^- < 0$. Since β^{bp} and β^{op} are both identified, we take them as fixed in this step. Recall that we have

$$\begin{aligned}&\Pr\left(\epsilon_{it}^{bp} \leq -x_1\beta^{bp}, \theta^- - x_2\beta^{op} \leq \epsilon_{it}^{op} \leq -x_2\beta^{op}\right)\\ &\neq \Pr\left(\epsilon_{it}^{bp} \leq -x_1\beta^{bp}, \tilde{\theta}^- - x_2\beta^{op} \leq \epsilon_{it}^{op} \leq -x_2\beta^{op}\right).\end{aligned}$$

As a result, θ^- is identified.

Step 2 uses Manski's (1985; 1988) identification argument to identify the coefficients of the lagged dependent variables. To illustrate, consider the following probabilities:

$$\begin{aligned}&\Pr\left(S_{it} = 0, E_{it} = 0 | Z_i^{t-1} = z, Z_{it} = (0,0,0,0), X_{1i}^T = \left(x_1^-, x_1\right), X_{2i}^T = \left(x_2^-, x_2\right)\right)\\ &= \Pr\left(\epsilon_{it}^{bp} \leq -x_1\beta^{bp}, \theta^- - x_2\beta^{op} \leq \epsilon_{it}^{op} \leq -x_2\beta^{op}\right),\end{aligned} \tag{3.6.1}$$

and

$$\begin{aligned}&\Pr\left(S_{it} = 0, E_{it} = 0 | Z_i^{t-1} = z, Z_{it} = (1,0,0,0), X_{1i}^T = \left(x_1^-, \widetilde{x_1}\right), X_{2i}^T = \left(x_2^-, \widetilde{x_2}\right)\right)\\ &= \Pr\left(\epsilon_{it}^{bp} \leq -\tilde{x}_1\beta^{bp} - \gamma_{11}, \theta^- - \tilde{x}_2\beta^{op} - \kappa_1 \leq \epsilon_{it}^{op} \leq -\tilde{x}_2\beta^{op} - \kappa_1\right).\end{aligned} \tag{3.6.2}$$

The expressions (3.6.1) and (3.6.2) will only be equal if and only if

$$\begin{aligned}-x_1\beta^{bp} &= -\widetilde{x_1}\beta^{bp} - \gamma_{11},\\ -x_2\beta^{op} &= -\widetilde{x_2}\beta^{op} - \kappa_1.\end{aligned}$$

Therefore, both γ_{11} and κ_1 are identified because $\gamma_{11} = (x_1 - \widetilde{x_1})\beta^{bp}$ and $\kappa_1 = (x_2 - \widetilde{x_2})\beta^{op}$ under the support condition in assumption A5. Similar arguments can be used to identify the coefficients of the other lagged dependent variables.

Step 3 follows from recognizing that we have a fully parametric binary choice model in (3.3.5) with only one copula dependence parameter left to identify.

In Step 4, we identify the signs of δ_{01}, δ_{02}, and κ_0. Note that we have

$$
\begin{aligned}
&\Pr\left(S_{it}=0, E_{it}=-1|Z_i^{t-1}=z, Z_{it}=(0,0,0,0), X_{1i}^T=\left(x_1^-, x_1\right), X_{2i}^T=\left(x_2^-, x_2\right)\right) \\
=&\Pr\left(\epsilon_{it}^{bp}\leq -x_1\beta^{bp}-\delta_{01}, \epsilon_{it}^{op}\leq \theta^- - x_2\beta^{op}\right)\Pr\left(d_i=1\right) \\
&+\Pr\left(\epsilon_{it}^{bp}\leq -x_1\beta^{bp}, \epsilon_{it}^{op}\leq \theta^- - x_2\beta^{op}\right)\Pr\left(d_i=0\right).
\end{aligned}
$$

Showing that this conditional probability is greater than

$$\Pr\left(\epsilon_{it}^{bp}\leq -x_1\beta^{bp}, \epsilon_{it}^{op}\leq \theta^- - x_2\beta^{op}\right)$$

allows us to conclude that $\delta_{01}<0$. The other cases follow analogously. Note that to avoid cumbersome notation, I omit the conditioning set in $\Pr\left(d_i=1\right)$ and $\Pr\left(d_i=0\right)$.

For Step 5, there are eight cases to consider. The resulting group assignment rules follows the same intuition as Table 3.2.1 and by sketching figures like Figure (3.3.1). One of the cases is that once we know that $\delta_{01}>0$, $\delta_{02}>0$, and $\kappa_0>0$, we must assign $d_i=0$ if and only if

$$
\begin{aligned}
&\Pr\left(E_{it}=0|Z_i^{t-1}=z, Z_{it}=(0,0,0,0), X_{1i}^T=\left(x_1^-, x_1\right), X_{2i}^T=\left(x_2^-, x_2\right)\right) \\
>\ &\Pr\left(\epsilon_{it}^{op}\geq -x_2\beta^{op}\right)
\end{aligned}
$$

or assign $d_i=1$ if and only if

$$
\begin{aligned}
&\Pr\left(S_{it}=0|Z_i^{t-1}=z, Z_{it}=(0,0,0,0), X_{1i}^T=\left(x_1^-, x_1\right), X_{2i}^T=\left(x_2^-, x_2\right)\right) \\
>\ &\Pr\left(\epsilon_{it}^{bp}\leq -x_1\beta^{bp}\right)
\end{aligned}
$$

Of course, these assignment rules can be altered by changing the conditioning sets. The other cases follow similarly.

In Step 6, we can now point-identify δ_{01}, δ_{02}, and κ_0. One route is to look at the conditional probability of $(S_{it}=1, E_{it}=1)$ given $Z_i^{t-1}=z$, $Z_{it}=(0,0,0,0)$, $X_{1i}^T=\left(x_1^-, x_1\right)$, $X_{2i}^T=\left(x_2^-, x_2\right)$, $d_i=0$. This conditional probability is now a function of κ_0 and can be used to point-identify κ_0. The conditional probability of $(S_{it}=0, E_{it}=-1)$ given $Z_i^{t-1}=z$, $Z_{it}=(0,0,0,0)$, $X_{1i}^T=\left(x_1^-, x_1\right)$, $X_{2i}^T=\left(x_2^-, x_2\right)$, $d_i=1$ can be used to point-identify δ_{01}. Finally, the conditional probability of $(S_{it}=0, E_{it}=1)$ given $Z_i^{t-1}=z$, $Z_{it}=(0,0,0,0)$, $X_{1i}^T=\left(x_1^-, x_1\right)$, $X_{2i}^T=\left(x_2^-, x_2\right)$, $d_i=1$ can be used to point-identify δ_{02}. Alternative routes include changing the vector Z_{it} or using other regions found in Figure 3.3.1.

Empirical Results

There are five tables in this Appendix. Table 3.6.1 contain the distribution of spell lengths in the data. Table 3.6.2 contains a description of the regressors used in the empirical application. Tables 3.6.3, 3.6.4, and 3.6.5 contain the estimation results for Specifications A, B, and C, respectively, across Samples 1 and 2.

Table 3.6.1: Length of spells observed in the data

Number of periods	1	2	3	4	5	6	7	8	9	10
Number of males	10	13	23	30	130	131	93	132	121	116
Number of periods	11	12	13	14	15	16	17	18	19	20
Number of males	103	121	138	124	124	118	125	127	103	528

Table 3.6.2: List of variables in Hajivassiliou and Ioannides (1995)

Model for	Regressors in X_{it}
S_{it}	educational category of head (`edycat`)
	dummies for 1976-79 and 1980-83 periods (`era7679`, `era8083`)
	food needs, growth of food needs (`fneed`, `gfneed`)
	age, age squared, age cubed (`hage`)
	live in north/central, south, west, other regions (`liveinnc`, `liveinso`, `liveinwe`, `liveinot`)
	married, race is black or other (`msm`, `raceb`, `raceo`)
	religion is Christian, Jewish, or Protestant (`religceo`, `religjsh`, `religpro`)
	real rate of interest (`rri`)
E_{it}	county unemployment rate (`cunemp`)
	head disabled, educational category of head (`disab`, `edycat`)
	dummies for 1976-79 and 1980-83 periods (`era7679`, `era8083`)
	food needs, growth of food needs (`fneed`, `gfneed`)
	age, age squared, age cubed (`hage`)
	tenure, tenure squared (`htenure`)
	unemployment insurance received by head (`hunemins`)
	imputed wage (`impwage`)[a]
	tightness of labor market conditions (labmkt)
	live in north/central, south, west, other regions (`liveinnc`, `liveinso`, `liveinwe`, `liveinot`)
	married, number of children between 0-5 (`msm`, `numch05`)
	occupational unemployment rate (`occunemp`)
	race is black or other (`raceb`, `raceo`)
	religion is Christian, Jewish, or Protestant (`religceo`, `religjsh`, `religpro`)
	real rate of interest (`rri`)
	head is union member (`unionmem`)

[a] The authors include a variable representing some measure of the imputed wage (`impwage`). Unfortunately, the JAE data archive did not include this variable.

Table 3.6.3: Results for Specification A

Sample 1						Sample 2					
Equation for S_{it}			Equation for E_{it}			Equation for S_{it}			Equation for E_{it}		
Variable	Coef	SE	Variable	Coef	SE	Variable	Coef	SE	Variable	Coef	SE
Regressors common to both equations											
Intercept	1.270	0.151	Intercept	-2.366	0.166	Intercept	0.939	0.122	Intercept	-1.572	0.129
era7679	0.141	0.033	era7679	0.002	0.035	era7679	0.134	0.030	era7679	0.035	0.030
era8083	-0.311	0.054	era8083	-0.460	0.054	era8083	-0.091	0.031	era8083	0.099	0.030
edycat	-0.043	0.008	edycat	-0.027	0.008	edycat	-0.054	0.006	edycat	-0.071	0.006
hage	-11.077	0.775	hage	1.925	0.721	hage	-9.399	0.616	hage	-1.118	0.552
hagesq	9.207	0.944	hagesq	-4.754	0.872	hagesq	7.221	0.746	hagesq	-0.471	0.660
liveinnc	-0.071	0.037	liveinnc	-0.086	0.036	liveinnc	-0.068	0.030	liveinnc	-0.108	0.028
liveinot	0.502	0.151	liveinot	0.228	0.161	liveinot	0.365	0.109	liveinot	0.123	0.111
liveinso	0.073	0.038	liveinso	0.123	0.039	liveinso	0.064	0.030	liveinso	-0.046	0.029
liveinwe	0.018	0.043	liveinwe	-0.436	0.042	liveinwe	0.050	0.033	liveinwe	-0.400	0.033
mss	0.577	0.043	mss	-0.209	0.042	mss	0.547	0.034	mss	-0.214	0.031
raceb	0.380	0.056	raceb	0.208	0.051	raceb	0.390	0.045	raceb	0.275	0.040
raceo	-0.372	0.054	raceo	0.140	0.049	raceo	-0.347	0.049	raceo	0.222	0.043
religceo	0.092	0.044	religceo	0.273	0.042	religceo	0.125	0.034	religceo	0.235	0.032
religjsh	0.201	0.081	religjsh	0.198	0.081	religjsh	0.197	0.069	religjsh	0.165	0.067
religpro	0.060	0.033	religpro	0.006	0.033	religpro	0.157	0.025	religpro	0.070	0.025
rri	10.017	1.115	rri	13.997	1.141	rri	6.891	0.616	rri	4.193	0.632
Regressors that are included in one of the equations but excluded in the other											
fneed	0.254	0.374	cunemp	0.874	0.596	fneed	0.306	0.317	cunemp	0.679	0.391
gfneed	-0.400	0.047	disab	0.270	0.044	gfneed	-0.508	0.039	disab	0.136	0.034
			htenure	-3.542	0.412				htenure	-3.938	0.325
			htenursq	8.650	1.519				htenursq	8.051	1.183
			hunemins	0.557	0.032				hunemins	0.354	0.016
			labmkt	0.032	0.014				labmkt	0.055	0.011
			numch05	0.007	0.022				numch05	0.018	0.016
			occunemp	5.865	0.518				occunemp	3.647	0.319
			unionmem	0.215	0.029				unionmem	0.213	0.023

Table 3.6.4: Results for Specification B

Sample 1						Sample 2					
Equation for S_{it}			Equation for E_{it}			Equation for S_{it}			Equation for E_{it}		
Variable	Coef	SE	Variable	Coef	SE	Variable	Coef	SE	Variable	Coef	SE
Regressors common to both equations											
Intercept	1.347	0.153	Intercept	-2.106	0.169	Intercept	0.896	0.123	Intercept	-1.391	0.131
era7679	0.141	0.033	era7679	0.011	0.034	era7679	0.124	0.030	era7679	0.038	0.030
era8083	-0.320	0.054	era8083	-0.454	0.054	era8083	-0.081	0.031	era8083	0.088	0.030
edycat	-0.047	0.008	edycat	-0.034	0.008	edycat	-0.050	0.006	edycat	-0.074	0.006
hage	-11.074	0.781	hage	1.150	0.725	hage	-9.344	0.618	hage	-1.348	0.556
hagesq	9.133	0.950	hagesq	-4.131	0.873	hagesq	7.227	0.748	hagesq	-0.502	0.663
liveinnc	-0.068	0.037	liveinnc	-0.048	0.036	liveinnc	-0.065	0.030	liveinnc	-0.111	0.028
liveinot	0.505	0.153	liveinot	0.327	0.162	liveinot	0.388	0.109	liveinot	0.138	0.112
liveinso	0.074	0.038	liveinso	0.128	0.039	liveinso	0.069	0.030	liveinso	-0.036	0.029
liveinwe	0.012	0.043	liveinwe	-0.401	0.042	liveinwe	0.055	0.034	liveinwe	-0.390	0.032
mss	0.569	0.044	mss	-0.146	0.042	mss	0.542	0.034	mss	-0.185	0.031
raceb	0.405	0.057	raceb	0.213	0.052	raceb	0.381	0.045	raceb	0.280	0.040
raceo	-0.399	0.055	raceo	0.154	0.050	raceo	-0.355	0.049	raceo	0.219	0.043
religceo	0.097	0.044	religceo	0.247	0.042	religceo	0.131	0.034	religceo	0.235	0.032
religjsh	0.209	0.081	religjsh	0.174	0.082	religjsh	0.199	0.070	religjsh	0.137	0.067
religpro	0.053	0.034	religpro	-0.027	0.033	religpro	0.160	0.025	religpro	0.061	0.025
rri	10.481	1.125	rri	13.672	1.137	rri	6.476	0.622	rri	4.247	0.632
Regressors that are included in one of the equations but excluded in the other											
fneed	0.101	0.379	cunemp	1.233	0.597	fneed	0.189	0.319	cunemp	0.338	0.391
gfneed	-0.395	0.048	disab	0.280	0.044	gfneed	-0.512	0.039	disab	0.135	0.034
			htenure	-3.695	0.413				htenure	-4.087	0.329
			htenursq	9.044	1.528				htenursq	8.060	1.208
			hunemins	0.537	0.032				hunemins	0.356	0.016
			labmkt	0.037	0.014				labmkt	0.060	0.010
			numch05	0.013	0.022				numch05	0.019	0.016
			occunemp	5.889	0.519				occunemp	3.717	0.318
			unionmem	0.184	0.029				unionmem	0.204	0.023

Table 3.6.5: Results for Specification C

	Sample 1				Sample 2			
	Equation for S_{it}		Equation for E_{it}		Equation for S_{it}		Equation for E_{it}	
Variable	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Intercept	1.124	0.170	-1.869	0.171	0.759	0.136	-1.403	0.133
`era7679`	0.083	0.035	0.007	0.034	0.062	0.032	-0.038	0.031
`era8083`	-0.369	0.055	-0.426	0.054	-0.136	0.042	-0.146	0.040
`era8487`					-0.037	0.036	-0.282	0.034
`edycat`	-0.032	0.008	-0.028	0.008	-0.037	0.007	-0.050	0.006
`hage`	-10.555	0.801	-1.679	0.793	-8.819	0.642	-2.965	0.613
`hagesq`	8.625	0.966	-0.819	0.950	6.766	0.767	1.550	0.725
`liveinnc`	-0.093	0.039	-0.046	0.036	-0.074	0.030	-0.099	0.028
`liveinot`	0.492	0.152	0.379	0.163	0.386	0.109	0.197	0.112
`liveinso`	0.046	0.040	0.136	0.039	0.058	0.031	-0.014	0.029
`liveinwe`	-0.010	0.043	-0.393	0.042	0.043	0.034	-0.372	0.033
`mss`	0.570	0.045	-0.030	0.046	0.541	0.035	-0.060	0.035
`raceb`	0.408	0.057	0.194	0.052	0.393	0.045	0.237	0.040
`raceo`	-0.376	0.056	0.159	0.050	-0.344	0.050	0.166	0.043
`religceo`	0.087	0.044	0.218	0.042	0.123	0.034	0.203	0.032
`religjsh`	0.197	0.082	0.144	0.082	0.184	0.070	0.093	0.068
`religpro`	0.038	0.034	-0.036	0.034	0.150	0.026	0.033	0.025
`rri`	8.812	1.168	13.064	1.139	5.351	0.780	7.374	0.733
`fneed`	0.071	0.383	2.637	0.344	0.308	0.325	2.657	0.291
`gfneed`	-0.390	0.048	-0.088	0.050	-0.511	0.039	-0.121	0.039
`cunemp`	-1.819	0.625	1.295	0.597	-1.180	0.427	0.369	0.391
`disab`	0.069	0.048	0.277	0.044	0.022	0.038	0.136	0.034
`htenure`	-1.168	0.429	-3.921	0.414	-1.514	0.348	-4.804	0.326
`htenursq`	2.596	1.624	9.529	1.528	2.586	1.322	10.308	1.180
`hunemins`	0.085	0.034	0.542	0.032	0.044	0.017	0.365	0.016
`labmkt`	0.031	0.015	0.038	0.014	0.017	0.011	0.048	0.011
`numch05`	-0.012	0.022	0.000	0.021	-0.021	0.016	0.010	0.016
`occunemp`	3.381	0.537	5.730	0.519	2.029	0.343	4.045	0.321
`unionmem`	-0.015	0.032	0.183	0.029	-0.017	0.026	0.198	0.023