Journal of Econometrics 178 (2014) 805–823

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Dynamic binary outcome models with maximal heterogeneity☆

Martin Browning[a], Jesus M. Carro[b,*]

[a] Department of Economics, University of Oxford, United Kingdom
[b] Departamento de Economia, Universidad Carlos III de Madrid, Spain

ARTICLE INFO

Article history:
Received 8 January 2011
Received in revised form
15 October 2013
Accepted 12 November 2013
Available online 21 November 2013

JEL classification:
C23
C24
J64

Keywords:
Discrete choice
Markov processes
Nonparametric identification
Unemployment dynamics

ABSTRACT

Most econometric schemes to allow for heterogeneity in micro behavior have two drawbacks: they do not fit the data and they rule out interesting economic models. In this paper we consider the time homogeneous first order Markov (HFOM) model that allows for maximal heterogeneity. That is, the modeling of the heterogeneity does not impose anything on the data (except the HFOM assumption for each agent) and it allows for any theory model (that gives a HFOM process for an individual observable variable). 'Maximal' means that the joint distribution of initial values and the transition probabilities is unrestricted.

We establish necessary and sufficient conditions for generic local point identification of our heterogeneity structure that are very easy to check, and we show how it depends on the length of the panel.

We apply our techniques to a long panel of Danish workers who are very homogeneous in terms of observables. We show that individual unemployment dynamics are very heterogeneous, even for such a homogeneous group. We also show that the impact of cyclical variables on individual unemployment probabilities differs widely across workers. Some workers have unemployment dynamics that are independent of the cycle whereas others are highly sensitive to macro shocks.

1. Introduction

Models with a binary outcome that depends in part on previous realizations of the outcome – dynamic binary outcome models – are common in applied microeconometrics. Some examples include: labor force participation (Heckman, 1981; Hyslop, 1999); smoking (Becker et al., 1994); firms exporting (Bernard and Jensen, 2004); stock market participation (Alessie et al., 2004) and taking up a welfare program (Gottschalk and Moffitt, 1994; Ham and Shore-Sheppard, 2005). The usual time-homogeneous first order Markov model for unit i (=1, ..., N) in period t ($t = 0, \dots, T$) is:

$$\Pr\left(y_{it} = 1 \mid y_{i,t-1}, x_{it}\right) = F\left(\eta_i + \alpha y_{it-1} + \beta x_{it}\right) \quad (1.1)$$

where $F(.)$ is a probability distribution function and y_{it} is a binary variable indicating, for example, that person i had some unemployment in period t. This 'linear index model' which only allows for a heterogeneous 'intercept' η_i is widely used but it does have problems; Browning and Carro (2007) discuss these but it is worth repeating the objections.

The first problem is that the imposition of common slope parameters (α and β) restricts the class of structural models that are consistent with the reduced form (1.1). For example, consider two people, a and b, with the same value of the x variables (so we can ignore them), and for whom a has a lower probability of being unemployed if they were employed in the previous year:

$$F(\eta_a) < F(\eta_b). \quad (1.2)$$

For example, a might choose a 'safer' job than b. Now suppose we impose the 'same slope' homogeneity assumption $\alpha_a = \alpha_b = \alpha$. This implies:

$$F(\eta_a + \alpha) < F(\eta_b + \alpha). \quad (1.3)$$

This rules out, for example, that a's caution leads her to spend more time looking for a 'safe' job, so that her probability of remaining unemployed is *higher* than b's. Thus the choice of a statistical

☆ For comments and useful suggestions, we thank three referees, Enrique Sentana, Whitney Newey, Ivan Fernandez-Val, Sara Ayo, and participants at seminars at Boston University; MIT/Harvard; Yale University; Nuffield (Oxford); IFS (London); CEMFI; Manchester; Columbia, CAM (Copenhagen) and a conference at the Tinbergen Institute. The second author gratefully acknowledges that this research was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme, by grants ECO2012-31358, ECO2009-11165 and SEJ2006-05710 from the Spanish Minister of Education, MCINN (Consolider-Ingenio2010), Consejería de Educación de la Comunidad de Madrid (Excelecon project); he also thanks the Department of Economics at the MIT at which he conducted part of this research as visiting scholar.

* Corresponding author. Tel.: +34 91 624 9586.
E-mail addresses: Martin.Browning@economics.ox.ac.uk (M. Browning), jcarro@eco.uc3m.es (J.M. Carro).

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.11.005

scheme for dealing with heterogeneity has substantive restrictions on the set of admissible structural models.

The second problem with the conventional approach is that whenever we have long enough panels to estimate the model for each unit individually with minimal bias, we do find substantial heterogeneity in both the 'intercept' and 'slope' parameters in (1.1). A situation where this is the case can be found in Browning and Carro (2010). Additional evidence will be provided in the empirical illustration in this paper.

Model (1.1) with maximal heterogeneity has[1]:

$$\Pr\left(y_{it} = 1 \mid y_{i,t-1}, x_{it}\right) = F\left(\eta_i + \alpha_i y_{it-1} + \beta_i x_{it}\right). \quad (1.4)$$

In addition to the homogeneity restrictions, model (1.1) is imposing two kinds of parametric restrictions: the parametric form implied by the linear index and the probability distribution function $F(.)$. In this paper, we consider not only a semiparametric form but also the nonparametric case as well as having maximal heterogeneity throughout the paper.[2] The nonparametric time-homogeneous first order Markov process (HFOM) with maximal heterogeneity allowing that the transition probabilities can be different for each individual can be written:

$$\begin{aligned}&\Pr\left(y_{it} \mid y_{it-1} = y_{-1}, y_{it-2}, \ldots, y_{i0}, x_{it} = x\right)\\&\quad = \Pr\left(y_{it} = 1 \mid y_{i,t-1} = y_{-1}, x_{it} = x\right) = p_{i,x,y_{-1}}\end{aligned} \quad (1.5)$$

where the first equality for all t is what characterizes a HFOM, and we have one parameter to be estimated for each i and the value of x and the lag of y. This does not impose any restrictions on the structural model (except, of course, for the assumption of time invariance and no effects higher than the first order that define the model considered in this paper) and it will fit any data that is generated by a time-homogeneous first order Markov process. For the simpler case without x variables there is a one to one correspondence between (1.4) and (1.5) and, therefore, any $F(.)$ will give the same transition probabilities. For the general case with x variables, a semiparametric form assuming a function $F(.)$ in (1.4) will impose some parametric restrictions that are not imposed in (1.5).

Identifying and estimating the whole set of transition probabilities in (1.5) – the whole set of parameters if we consider (1.4) – or their distribution over the population, allows us to obtain any parameter of interest in this problem, including the average marginal effects (also known as average partial effects, APE) and the median marginal effect of a explanatory variable over the outcome y_{it}. Furthermore, identifying and estimating the whole HFOM model will allow to obtain the entire distribution in the population of the effect of a variable over the outcome. In a program evaluation context, Heckman et al. (1997) present situations in which the entire distribution, and not only the mean effect, is the policy parameter of interest. In the IO literature it is also of interest to identify the entire distribution of the individual price elasticities when estimating demand functions; see for example Nevo (2001).

Given the difficulties in estimating (1.1) with small and fixed T (see Arellano and Honoré, 2001), tackling (1.5) or (1.4) is a formidable task. In Browning and Carro (2010) we suggested two estimation methods for the simple case without x variables, that rely on reducing the bias or RMSE for estimates based on each unit. This gives estimates for each unit and then the distribution for (η, α) can be taken as the empirical distribution of these estimates (or some smoothed version of it).

In Browning and Carro (2010), identification and estimation of (1.5) without imposing any restriction on the distribution of (η, α) nor on the initial condition, relies on the T dimension; that is, it is only consistent when $T \to \infty$. In this paper we propose an alternative approach that relies on large N. In general the model is not nonparametrically identified from a cross section of observations of fixed length T.[3] This negative result is our starting point in this paper: identification from the cross section is our goal since we typically do not have panels with a very large number of periods. Nevertheless, this negative result on identification does not imply that we cannot learn anything from a cross section of paths with a fixed T. In general, some restrictions will have to be imposed on the distribution of the heterogeneity to achieve point identification. The interesting question is the nature of the restrictions we have to impose, or how much information about our model with maximal heterogeneity we can identify from a cross section of length T. To answer this question we use finite discrete mixture distributions for the joint set of unknown heterogeneous parameters. We refer to this as the *flexible discrete scheme* since no restriction is imposed other than there is a finite and discrete number of points of support on this distribution.

An advantage of this discrete scheme is that it allows us to go from the homogeneous case (one point of support) to the totally unrestricted case (as many points of support as N) within the same scheme. Also, given the discrete nature of problem and the finite number of possible observations, it is clear that we cannot nonparametrically identify a continuous distribution. So, the *flexible discrete scheme* is our route to study nonparametric point identification.[4]

The identification issue in this scheme will be: how many points of support can we take for a given T? A major gain from looking at models identified from a cross section with fixed T is that there is no incidental parameters problem nor finite sample bias problem from not having a large number of periods.

Kasahara and Shimotsu (2009) take a different approach to a more general problem that includes the model we consider here, as well as other models. One of the examples included in their paper to illustrate their results is model (1.4) without x variables. However, for this case they do not give identification conditions for an arbitrary number of periods. For example, their most important result for this model (Proposition 7 in Kasahara and Shimotsu (2009)) requires $T \geq 8$. Also they give stronger sufficient conditions than the conditions derived in this paper, whereas here we derive sufficient and necessary conditions for identification. Moreover, their conditions are nontrivial to check in actual data, whereas our conditions are simple to check.

A different and interesting analysis is to look at set identification for the cases that are not point identified. In particular to derive bounds in the non-identified situation when no restriction or distribution is assumed for the heterogeneous parameters.

[1] Model (1.4) can be seen as part of the larger literature on random coefficients model. In that literature there are some cases whose identification and estimation has been studied. An example is Gautier and Kitamura (2013) that considers the estimation of random coefficient static models with continuous covariates. Also, in contrast with us, they assume that the distribution of the unobserved heterogeneous β coefficients is independent of the covariates.

[2] Notice also that in (1.1) an extra homogeneity assumption is imposed by assuming all i have the same $F(.)$. In our nonparametric approach this homogeneity assumption is not imposed either.

[3] In general, not even the restrictive model (1.1) with only one fixed effect is identified; see Honorè and Tamer (2006).

[4] We note that our use of a discrete distribution to capture heterogeneity is different to that suggested by Heckman and Singer (1984). They show that the distribution of a continuous latent variable is nonparametrically identified for a particular parametric duration model. They then suggest that the continuous distribution can be reasonably approximated by a discrete distribution with a small number of support points. In contrast, in our scheme the continuous distribution is *not* nonparametrically identified, and any continuous distribution can be perfectly approximated by discrete finite mixtures (see Lemma A.1 in Ghosal and van der Vaart (2001)).