The identification of a mixture of first order binary Markov Chains*

Martin Browning[†] and Jesus M. Carro[‡]

January 2011

Keywords: discrete choice, Markov processes, global identification

Abstract

Let S be the number of components in a finite discrete mixing distribution. We prove that the number of waves of panel being greater than or equal to $2S$ is a sufficient condition for global identification of a dynamic binary choice model in which all the parameters are heterogeneous. This model results in a mixture of S binary first order Markov Chains.

1 Introduction

When considering observed persistence in time varying choices, Heckman (1981) emphasized the importance of distinguishing between unobserved heterogeneity and true state dependence. A preeminent example in empirical work is the modelling of dynamic discrete choice models, for example, labour force participation. In allowing for heterogeneity it is

*The second author gratefully acknowledges that this research was supported by grant number ECO2009-11165 from the Spanish Minister of Science.

[†]Department of Economics. University of Oxford. E-mail: Martin.Browning@economics.ox.ac.uk

[‡]Departmento de Economia. Universidad Carlos III de Madrid. E-mail: jcarro@eco.uc3m.es

important to capture unobserved heterogeneity in the state dependence parameter as well as in the 'intercept'; see Browning and Carro (2007). One convenient way to do this is to employ finite mixture models. Here we examine the identification of mixture model of S binary first order Markov Chains. This mixture model corresponds to a dynamic binary choice model in which all the parameters are heterogeneous.

Discrete finite mixtures as a flexible ('nonparametric') way to control for unobserved heterogeneity have been widely used. It was popularized in economics by the work of Heckman and Singer (1982) in duration models, but it is used in many other non-linear models, including discrete choice models. The question of identification of finite mixtures has been studied for many decades in statistics and econometrics. Teicher (1961), Blischke (1962), Blischke (1964) and Teicher (1963) are among the first examples. These studies considered the identification of mixtures of normal, gamma, or binomial distributions, but they did not consider mixtures of first order Markov Chains. A recent study in econometrics is Kasahara and Shimotsu (2009), but for the model of our interest they do not give identification conditions for an arbitrary number of periods. Moreover, their conditions are very difficult to check in actual data. In this paper we derive a explicit (and very simple) sufficient condition for global identification in terms of the number of waves of a panel that is needed to identify a mixture of S binary first order Markov Chains.

2 Sufficient conditions for identification

Let $Y_i = (y_{i0}, y_{i1}, ...y_{iT})$ be a realization of a binary variable y_{it} that follows a time-homogeneous first order Markov process. The transition probabilities that define this process are:

$$G_s = \Pr\left(y_t = 1 \mid y_{t-1} = 0, s\right) \tag{1}$$

$$H_s = \Pr\left(y_t = 1 \mid y_{t-1} = 1, s\right) \tag{2}$$

where s indexes the S distributions we are mixing, and we have $T+1$ realizations of this process. We make all our analysis conditional on the initial observation. The distribution of Y_i conditional on y_{i0} is the following mixture

$$\Pr(Y_i|y_{i0}) = \sum_{s=1}^{S} \theta_{s|y_{i0}} G_s^{n_{01}} \left(1 - G_s\right)^{n_{00}} H_s^{n_{11}} \left(1 - H_s\right)^{n_{10}} \tag{3}$$

where n_{01} is the number of $0 \to 1$ transitions in path Y_i, and similarly for the other three transitions. $\theta_{s|y_{i0}}$ gives the mixing probabilities of each value of (G_s, H_s) conditional on the initial observation. That is, we have one mixing distribution for those Y_i that start with $y_{i0} = 0$, and another one, possibly different, for those with $y_{i0} = 1$. The unconditional mixing proportions can be easily recovered using the observed proportion of $y_{i0} = 1$. Note that $\sum_{s=1}^{S} \theta_{s|y_{i0}} = 1$, and therefore $\theta_{S|y_{i0}} = 1 - \sum_{s=1}^{S-1} \theta_{s|y_{i0}}$. $0 < \theta_{s|y_{i0}}, G_s, H_s < 1$ for $s = 1, .., S$. Also, G_s and H_s take distinct values for different s.

The unknown parameters we want to identify are $\left\{ \left[\theta_{s|y_{i0}=0}, \theta_{s|y_{i0}=1}\right]_{s=1}^{S-1}, \left[G_s, H_s\right]_{s=1}^{S} \right\}$; in all there are $(4S-2)$ parameters. We provide sufficient conditions for global identification of the mixture in (3). We say the mixture is identified if from the realized proportions of the mixed distribution we can recover only one distinct value of $\left\{ \left[\theta_{s|y_{i0}=0}, \theta_{s|y_{i0}=1}\right]_{s=1}^{S-1}, \left[G_s, H_s\right]_{s=1}^{S} \right\}$ that yields that mixed distribution. Also, any sets of values of the unknowns that contain the same values but in different order (e.g. (G_1, H_1) in one set is (G_3, H_3) on another set) are the same solution. If that solution is unique we say the model is identified regardless of the number of ways it could be ordered.

The possible realizations of Y_i are all the possible combination of zeros and ones in the periods we have. To identify $\left\{\theta_{s|y_{i0}=1}, H_s\right\}_{s=1}^{S}$, we take those realizations with $y_{i0} =$

1, and construct moment conditions using the survivor function. That is, we take the probability that in the u periods following the initial observation we observe only ones: $\boldsymbol{S}_H(u) = \sum_{s=1}^{S} \theta_{s|y_{i0}=1} H_s^u$. For instance, for a given value H_s, the probability, conditional on $y_{i0} = 1$, of observing $y_{i1} = y_{i2} = 1$ (that is, $u = 2$) is, from equation (2), equal to H_s^2. Each value of u will give a moment condition, so we have the following system of equations:

$$f_{H,u} = \sum_{s=1}^{S} \theta_{s|y_{i0}=1} H_s^u; \ (u = 0, ..., T) \tag{4}$$

with $\boldsymbol{S}_H(0)$ and $f_{H,0}$ being trivially equal to one. The value $f_{H,u}$ is the population proportion of realizations Y_i whose first $u+1$ elements are equal to one. In order to have at least as many (informative) equations as unknowns in system (4) we need $T \geq 2S - 1$.

To show that $T \geq 2S - 1$ is a sufficient condition for global identification of the mixture in (3), firstly note that equations (4) are the same equations as equations (6) on page 513 of Blischke (1964), except for the different notation used. Therefore, from the same arguments used in Blischke (1964), if $T \geq 2S - 1$ there is a unique solution to this system and $\left\{ \left[\theta_{s|y_{i0}=1}\right]_{s=1}^{S-1}, [H_s]_{s=1}^{S} \right\}$ is identified from (4).

To identify $\left\{ \left[\theta_{s|y_{i0}=0}\right]_{s=1}^{S-1}, [G_s]_{s=1}^{S} \right\}$, we do the same analysis taking those realizations with $y_{i0} = 0$, and use the survivor function with the number of consecutive zeros following y_{i0}. This gives the following equations

$$f_{G,u} = \sum_{s=1}^{S} \theta_{s|y_{i0}=0} \left(1 - G_s\right)^u; \ (u = 0, ..., T) \tag{5}$$

$f_{G,u}$ is the population proportion of realizations Y_i whose first $u+1$ elements are equal to zero. Again, this is the same system of equations as equation (6) in Blischke (1964) with p_i^k in Blischke(1964) being $(1 - G_s)^u$ here. Therefore, if $T \geq 2S - 1$, there is a unique solution to this system and $\left\{ \left[\theta_{s|y_{i0}=0}\right]_{s=1}^{S-1}, [G_s]_{s=1}^{S} \right\}$ is identified from (5). This complete the identification of all the unknowns.

3 Concluding Remarks

We have shown that $T \geq 2S - 1$ is a sufficient condition for global identification of a mixture to S binary first order Markov Chains. Since in our notation the first observation of the process is 0 and T is the last observation, in terms of the number of periods observed $(= T + 1)$ this condition is

$$\text{number of periods} \geq 2S \tag{6}$$

Two final remarks are important:

1. Though we have made use of some results in Blischke (1964) the condition (6) for identification of (3) is different than the condition for identification of the binomial mixtures studied in Blischke (1964). Our mixture requires one more period to satisfy the sufficient condition for identification. This comparison is relevant since the binomial mixture is a special case of our model: the case in which $G_s = H_s = p_s$ for all s. This is the static version of our dynamic binary choice model.

2. The condition derived in this paper is sufficient but not necessary. At least it is not necessary for all the values of the parameters. The moments we have constructed are ignoring conditions in which both G_s and H_s are combined. Browning and Carro (2011) exploit all the possible conditions to derive, among other results, a necessary and sufficient condition for (generic local) identification. This condition requires a smaller number of periods than condition (6) for identification. In particular, it requires $T \geq \frac{-1}{2} + \sqrt{\frac{-7}{4} + 4S}$. For example, if $S = 4$ the latter requires at least 5 waves as opposed to the 8 needed for the sufficient condition derived here. However, the advantage of the condition derived here (equation (6)) is that it is for global identification and holds everywhere. The condition in Browning and Carro (2011) holds almost everywhere and, as such, it has some exceptions. Moreover, though it is conjectured that it holds also for global identification, the condition in Browning and Carro (2011) is proved in general only for local identification.

References

[1] Blischke (1962): "Moment Estimators for the Parameters of a Mixture of Two Binomial Distributions". *The Annals of Mathematical Statistics*, 33 (2), pp. 444-454

[2] Blischke (1964): "Estimating the Parameters of Mixtures of Binomial Distributions". *Journal of the American Statistical Association*, Vol. 59, No. 306, pp. 510-528.

[3] Browning, M. and J. M. Carro (2007). "Heterogeneity and Microeconometrics Modelling." *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, vol. 3. Edited by R. Blundell, W. K. Newey, and T. Persson. Cambridge University Press.

[4] Browning, M. and J. M. Carro. (2011): "Dynamic binary outcome models with maximal heterogeneity." *mimeo*.

[5] Heckman, J. J. (1981): "Heterogeneity and State Dependence" in *Studies in Labor Markets*, edited by Sherwin Rosen, University of Chicago Press.

[6] Heckman, J. J. and Singer, B. (1984) "A Method for Minimizing the Impact of Distributional Assumptions in Econometric-Models for Duration Data." *Econometrica*, 52(2), pp. 271-320.

[7] Kasahara, H. and K. Shimotsu (2009): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices" *Econometrica*, 77(1), 135-175

[8] Teicher (1961): "Identifiability of Mixtures". *The Annals of Mathematical Statistics*, 32 (1), pp. 244-248.

[9] Teicher (1963): "Identifiability of Finite Mixtures". *The Annals of Mathematical Statistics*, 34 (4), pp. 1265-1269.