OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 75, 3 (2013) 0305-9049
doi: 10.1111/j.1468-0084.2012.00696.x

The Identification of a Mixture of First-Order Binary Markov Chains*

MARTIN BROWNING† and JESUS M. CARRO‡

†*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)

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.

I. 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 modeling of dynamic discrete choice models, for example labour force participation. In allowing for heterogeneity, it is 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 (non-parametric) way to control for unobserved heterogeneity have been widely used. It was popularized in economics by the work of Heckman and Singer (1984) in duration models, but it is used in many other nonlinear 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). They consider a more general problem that includes

*The second author gratefully acknowledges that this research was supported by grant number ECO2009-11165 from the Spanish Minister of Science.
JEL Classification numbers: C1, C23, C25.

© Blackwell Publishing Ltd and the Department of Economics, University of Oxford 2012. Published by Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

the model we consider here, as well as other models. For the model considered here, they do not give identification conditions for an arbitrary number of periods. For the periods their identification conditions can be applied, they require smaller number of periods than our condition. However, their sufficient conditions include not only a condition in terms of the number of periods but additional rank conditions that are very difficult to check in actual data.[1] In this article, 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.

II. Sufficient conditions for identification

Let $Y_i = (y_{i0}, y_{i1}, \ldots y_{iT})$ be a realization of a binary variable y_{it} that follows a time-homogeneous first-order Markov process.[2] 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 conditional on y_0 is the following mixture:

$$\Pr(Y \mid y_0) = \sum_{s=1}^{S} \theta_{s|y_0} G_s^{n_{01}} \left(1 - G_s\right)^{n_{00}} H_s^{n_{11}} \left(1 - H_s\right)^{n_{10}}, \tag{3}$$

where n_{01} is the number of $0 \to 1$ transitions in path Y, and similarly for the other three transitions. $\theta_s \mid y_0$ 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 that start with $y_0 = 0$, and another one, possibly different, for those with $y_0 = 1$. The unconditional mixing proportions can be easily recovered using the observed proportion of $y_0 = 1$. Note that $\sum_{s=1}^{S} \theta_{s|y_0} = 1$, and therefore $\theta_{S|y_0} = 1 - \sum_{s=1}^{S-1} \theta_{s|y_0}$. $0 < \theta_{s|y_0}, G_s, H_s < 1$ for $s = 1, \ldots, S$. Also, both G_s and H_s take distinct values for different s.

The unknown parameters we want to identify are $\left\{ \left[\theta_{s|y_0=0}, \theta_{s|y_0=1}\right]_{s=1}^{S-1}, [G_s, H_s]_{s=1}^{S} \right\}$; 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 population proportions of the mixed distribution we can recover only one distinct value of $\left\{ \left[\theta_{s|y_0=0}, \theta_{s|y_0=1}\right]_{s=1}^{S-1}, [G_s, H_s]_{s=1}^{S} \right\}$ 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 are all the possible combination of zeros and ones in the periods we have. To identify $\left\{\theta_{s|y_0=1}, H_s\right\}_{s=1}^{S}$, we take those realizations with

[1]A more detailed comparison with Kasahara and Shimotsu (2009) is presented in section III.
[2]In the rest of this article, we omit subscript i for readability.

© Blackwell Publishing Ltd and the Department of Economics, University of Oxford 2012

$y_0 = 1$, 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: $S_H(u) = \sum_{s=1}^{S} \theta_{s|y_0=1} H_s^u$. For instance, for a given value H_s, the probability, conditional on $y_0 = 1$, of observing $y_1 = y_2 = 1$ (i.e $u = 2$) 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:

$$f_{H,u} = \sum_{s=1}^{S} \theta_{s|y_0=1} H_s^u; \quad (u = 0, \ldots, T) \tag{4}$$

with $S_H(0)$ and $f_{H,0}$ being trivially equal to one. The value $f_{H,u}$ is the population proportion of realizations Y whose first $u+1$ elements are equal to one. 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), first note that equation (4) is the same equation as equation (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 $\left\{ \left[\theta_{s|y_0=1}\right]_{s=1}^{S-1}, [H_s]_{s=1}^{S} \right\}$ is identified from (4).

To identify $\left\{ \left[\theta_{s|y_0=0}\right]_{s=1}^{S-1}, [G_s]_{s=1}^{S} \right\}$, we do the same analysis taking those realizations with $y_0 = 0$, and use the survivor function with the number of consecutive zeros following y_0. This gives the following equations

$$f_{G,u} = \sum_{s=1}^{S} \theta_{s|y_0=0} \left(1 - G_s\right)^u; \quad \left(u = 0, \ldots, T\right), \tag{5}$$

where $f_{G,u}$ 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 $\left(1 - G_s\right)^u$ here. Therefore, if $T \geq 2S - 1$, there is a unique solution to this system and $\left\{ \left[\theta_{s|y_0=0}\right]_{s=1}^{S-1}, [G_s]_{s=1}^{S} \right\}$ is identified from (5). This complete the identification of all the unknowns.

III. 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

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

Three final remarks are important:

(i) Although 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 $G_s = H_s = p_s$

© Blackwell Publishing Ltd and the Department of Economics, University of Oxford 2012

for all s. This is the static version of our dynamic binary choice model. Kasahara and Shimotsu (2009) also consider this static binary choice model without covariates in their Remark 3 (p. 149) in section 2. Though using a very different procedure, they obtain the same condition as Blischke (1964).

(ii) Kasahara and Shimotsu (2009) in section 3.2 consider dynamic discrete choice models. This includes our first-order binary Markov Chain model in Example 5 and Proposition 7 of Kasahara and Shimotsu (2009, pp. 158–159), for which sufficient, but not necessary, conditions for identification are derived. Three differences between our result and their result should be noted. First, Proposition 7 in Kasahara and Shimotsu (2009) only applies to cases with $T \geq 7$, whereas condition (6) applies to any T. Second, their set of sufficient conditions for identifying (3) requires checking the rank of two matrices constructed from the transition probabilities, which are non-trivial to check in actual data. In contrast with this, our condition is straight forward to check. Last but not least, the main advantage of the sufficient conditions in Kasahara and Shimotsu (2009) is that, for the periods Proposition 7 applies, the required number of periods to identify S points of support is smaller than in (6). Moreover, the number of identifiable components increases exponentially with T, whereas here it only increases linearly. For example, if the number of periods is 10, 12 and 14, then the maximum number of identifiable components by Proposition 7 in Kasahara and Shimotsu (2009) is 10, 15 and 21, respectively, whereas the maximum number of identifiable components by condition (6) is 5, 6 and 7, respectively.

(iii) The condition derived in this article 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 the number of periods is 10, 12 and 14, then the maximum number of identifiable components S according with the latter condition is 23, 33 and 46, as opposed to 5, 6 and 7 identifiable according to sufficient condition (6). 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.

Final Manuscript Received: January 2012

References

Blischke, W. R. (1962). 'Moment estimators for the parameters of a mixture of two binomial distributions', *The Annals of Mathematical Statistics*, Vol. 33, pp. 444–454.

© Blackwell Publishing Ltd and the Department of Economics, University of Oxford 2012

Blischke, W. R. (1964). 'Estimating the parameters of mixtures of binomial distributions', *Journal of the American Statistical Association*, Vol. 59, pp. 510–528.
Browning, M. and Carro, J. M. (2007). 'Heterogeneity and microeconometrics modelling', in Blundell R., Newey W. K. and Persson T. (eds), *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, Vol. 3, New York: Cambridge University Press.
Browning, M. and Carro, J. M. (2011). *'Dynamic binary outcome models with maximal heterogeneity.' Unpublished manuscript*, Departamento de Economía, Universidad Carlos III de Madrid, Spain.
Heckman, J. J. (1981). 'Heterogeneity and state dependence' in Sherwin Rosen (ed.), *Studies in Labor Markets*, Chicago: University of Chicago Press.
Heckman, J. J. and Singer, B. (1984). 'A method for minimizing the impact of distributional assumptions in econometric-models for duration data', *Econometrica*, Vol. 52, pp. 271–320.
Kasahara, H. and Shimotsu, K. (2009). 'Nonparametric identification of finite mixture models of dynamic discrete choices', *Econometrica*, Vol. 77, pp. 135–175.
Teicher, H. (1961). 'Identifiability of mixtures', *The Annals of Mathematical Statistics*, Vol. 32, pp. 244–248.
Teicher, H. (1963). 'Identifiability of finite mixtures', *The Annals of Mathematical Statistics*, Vol. 34, pp. 1265–1269.

© Blackwell Publishing Ltd and the Department of Economics, University of Oxford 2012