Economics Letters 146 (2016) 17–20

Contents lists available at ScienceDirect

Economics Letters

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

Determinants of transition in artificially discrete Markov chains using microdata

Achilleas Vassilopoulos *, Stathis Klonaris

Department of Agricultural Economics and Rural Development, Agricultural University of Athens, Greece

HIGHLIGHTS

- An econometric procedure to model transitions in Markov chains is proposed.
- The model is applicable when the continuous classification variable is observed.
- Transition probabilities, marginal effects and discrete changes are calculated.
- The model might be useful in a number of situations and in several disciplines.

ARTICLE INFO

Article history:
Received 21 March 2016
Received in revised form
2 June 2016
Accepted 9 July 2016
Available online 15 July 2016

Keywords:
Markov chain
Determinants of transition
Microdata

ABSTRACT

We describe an econometric procedure to model transitions in Markov chains whose state space is finite and classification stems from observed continuous variables. We show how stationary and non-stationary transition probabilities as well as the marginal effects of continuous and dichotomous variables determining transition can be estimated. The model resembles the ordered probit approach used in Epstein et al. (2006) but allows for the differences in the nature of the dependent variable and suggests some very important extensions pertaining to more meaningful representation of parameter estimates and the simultaneous construction of transition matrices.

1. Introduction

A Markov chain is characterized by a set of states S, each by-two permutation of which is associated with a transition probability $p_{i\to j}$ expressing the likelihood that an object (firm, country, person, household etc.) is at the next point in time at state j, given that at the present it is at state i. In what follows, we will assume that S is a finite set and that the system is closed so that every object is in exactly one state in each period and either stays there in the next period or moves to a new state; meaning that the sum of probabilities of movements from one state to all other elements of S is equal to one. If we place $p_{i\to j}$ into an $i \times j$ matrix, we can construct the stochastic matrix, known as the transition or Markov matrix of the system which will have the form:

$$\mathbf{T} = \begin{pmatrix} p_{1\to 1} & p_{1\to 2} & \cdots & p_{1\to j} \\ p_{2\to 1} & p_{2\to 2} & \cdots & p_{2\to j} \\ \vdots & \vdots & \ddots & \vdots \\ p_{i\to 1} & p_{i\to 2} & \cdots & p_{i\to j} \end{pmatrix}. \quad (1)$$

With micro-data available, maximum likelihood estimates of transition probabilities for stationary (or time-homogeneous) Markov chains can be derived by the number of movements from state i to state j during one period over the total number of one-period transitions from i (Bishop et al., 1977). Non-stationary (time-varying) transition probabilities on the other hand, can be estimated in the same manner but for each period separately.

However, in dynamic systems, factors such as the dependence of transition on changes in the objects' macro and micro-economic environment would lead to deviation from the stochastic process assumed. If this is true, the mere construction of a Markov matrix would be uninformative and therefore, a deterministic model is necessary to examine whether the underlying process of transition from one class to another is purely stochastic or a function of several other variables. The following section reads a brief presentation of the econometric model and suggests how it should be applied in several disciplines. The suggested methodology resembles the ordered probit approach used in Epstein et al. (2006) but allows for the differences in the nature of the dependent variable and suggests some very important extensions pertaining to more meaningful representation of parameter estimates and the simultaneous construction of transition matrices which are presented in the final section.

* Corresponding author.
E-mail addresses: avas@aua.gr (A. Vassilopoulos), s.klonaris@aua.gr (S. Klonaris).

http://dx.doi.org/10.1016/j.econlet.2016.07.018

2. The model

At first, assume a variable y_i can be captured sufficiently accurately using a linear model, as:

$$y_i = \beta'\mathbf{x}_i + \varepsilon_i \tag{2}$$

with $\mathbf{x}_i$ being the vector of possible determinants.

Now assume that, the researcher has access to longitudinal data on $(y_{it}, \mathbf{x}_{it})$ but is interested in determining the relationship of the explanatory variables with an ordinal transformation of y_{it}, say Y_{it}. For example, think of y_{it} as being a firm's capital (in USD), or a farm's land size (in hectares) and Y_{it} as a size indicator, ranging from "very small" to "very big" or of y_{it} as a region's immigration status (in percentage points of immigrants over total population) and Y_{it} as an immigration density index ranging from "very sparse" to "very dense". Alternatively, y_{it} may be a country's average hourly ozone/$PM_{2.5}$ (in mg/cm^3) or a food product's pathogenic microorganism/chemical residues/natural poisons concentration (in number of organisms or mg/kg) and Y_{it} an air quality indicator ranging from "Good" to "Extremely Hazardous" or a safety claim from "Completely safe" to "Extremely dangerous", respectively. Taking one step at a time, the researcher classifies observations into J categories based on their y_{it} value and using some meaningful cut-off points (μ) as:

$$\begin{aligned} Y_{it} &= 1 \quad \text{if } y_{it} \le \mu_1 \\ &= 2 \quad \text{if } \mu_1 < y_{it} \le \mu_2 \\ &= 3 \quad \text{if } \mu_2 < y_{it} \le \mu_3 \\ &\dots \\ &= J \quad \text{if } y_{it} > \mu_{J-1} \end{aligned} \tag{3}$$

where Y_{it} indicates the category an object belongs to at time t.

Assuming ε_{it} follows a normal distribution with zero mean and variance σ^2, an alternative representation based on (2) is[1]:

$$\begin{aligned} \Pr(Y_{it} = j) &= \Pr(\mu_{j-1} < y_{it} \le \mu_j) \\ &= \Pr(\mu_{j-1} < \beta'\mathbf{x}_i + \varepsilon_{it} \le \mu_j) \\ &= \Pr(\mu_{j-1} - \beta'\mathbf{x}_i < \varepsilon_{it} \le \mu_j - \beta'\mathbf{x}_i) \\ &= \Pr\left(\frac{\mu_{j-1} - \beta'\mathbf{x}_i}{\sigma} < \frac{\varepsilon_i}{\sigma} \le \frac{\mu_j - \beta'\mathbf{x}_i}{\sigma}\right) \\ &= \Phi\left(\frac{\mu_j - \beta'\mathbf{x}_i}{\sigma}\right) - \Phi\left(\frac{\mu_{j-1} - \beta'\mathbf{x}_i}{\sigma}\right). \end{aligned} \tag{4}$$

Although (4) has a familiar discrete choice form, as shown in (2), β and σ can be consistently estimated by a linear regression of y_{it} on $\mathbf{x}_i$ and thus their non-linear functions are also consistent estimates of the probability of being in each state at period t.[2] As a result, the obvious use of such a model would be to study the probability of an object being in each discrete state at period t as well as the influence a set of explanatory variables ($\mathbf{x}$) have on this probability. However, since our interest focuses on transition probabilities ($Y_{i(t-1)} \to Y_{it}$) rather than mere probabilities of being in each state at period t, $\mathbf{x}_i$ should include a variable showing the lagged state of the object while if non-stationary transition probabilities are assumed, period dummies should also be included.[3] Both sets of dummy variables can enter the model either as are or as interactions with the rest of the independents, to allow for differential effects of the regressors. Also, lagged regressors (if the examined time period is relatively short), their first-differences or variables whose values correspond to the interim period are more suitable candidates for $\mathbf{x}_i$ than end-state variables since the latter present a post-transition snapshot of the object in the system and are unlikely to have driven the transition. For example, if Y_{it} is a firm's size at period t then the financial leverage at the same period seem not to be a good predictor for an observed transition but its difference between $t-1$ and t may very well be. Or, if Y_{it} is a food product's safety claim at period t, then the average temperature between $t-1$ and t seems more reasonable than the temperature at t.

3. Transition probabilities and determinants

Even with the above suggestions for the $\mathbf{x}_i$ vector, neither the signs not the magnitudes of the $\hat{\beta}$'s are directly interpretable in an ordered choice model such as the one presented here (see also, Greene and Hensher, 2010). Nevertheless, the calculation of the marginal effects will result in a unique vector of parameter estimates for each year and each transition period as in matrix (1). Specifically, if one is to use N explanatory variables in the above model aiming at explaining a Markov Chain with $s = 1, 2, \dots, S$ discrete states over $k = 1, 2, \dots, K$ periods (i.e. $K-1$ transitions), the empirical analogue of (2) would be:

$$\begin{aligned} \hat{y}_{it} = \hat{\beta}_0 &+ \sum_{s=2}^{S} \hat{\beta}_s Y^s_{i(t-1)} + \sum_{k=3}^{K} \hat{\gamma}_k m^k_{i(t-1)} \\ &+ \sum_{n=1}^{N} \hat{\delta}_n \xi_{ni} + \sum_{s=2}^{S}\sum_{k=3}^{K} \hat{\rho}_{sk} Y^s_{i(t-1)} m^k_{i(t-1)} \\ &+ \sum_{s=2}^{S}\sum_{n=1}^{N} \hat{\theta}_{sn} Y^s_{i(t-1)} \xi_{ni} \\ &+ \sum_{k=3}^{K}\sum_{n=1}^{N} \hat{\psi}_{kn} m^k_{i(t-1)} \xi_{ni} + \hat{\varepsilon}_{it} \end{aligned} \tag{5}$$

where:

$Y^s_{i(t-1)}$: Lagged state dummy variables with $Y^s_{i(t-1)} = 1[Y_{i(t-1)} = s]$
$m^k_{(t-1)}$: Lagged period dummy variables with $m^k_{(t-1)} = 1[t-1 = k]$
ξ_n: Explanatory variables

and $\hat{\beta}_0, \hat{\beta}_s, \hat{\gamma}_k, \hat{\delta}_n, \hat{\rho}_{sk}, \hat{\theta}_{sn}, \hat{\psi}_{kn}$ are the estimates of the parameters of interest.

Notice that Y^1_{t-1} and m^2_{t-1} are dropped to avoid multicollinearity while the first period of each observation is excluded, since there is no information on its previous state. Plugging the estimates of β and σ from (5) into (4), with all regressors held at their means,[4] we can estimate $\Pr(Y = s_j)$ which is the probability of being in state s_j. Still, to construct transition matrices analogous to (1), we need the predicted probabilities of transitions to s_j from any other state s_l; this can be easily calculated if instead of holding all regressors at their respective means, we fix $Y^{s_l}_{i(t-1)}$ to one and all other $Y^s_{i(t-1)}$'s to zero. As a result, (4) will now be:

$$\hat{p}_{s_l \to s_j} = \Phi\left(\frac{\mu_{s_j}}{\hat{\sigma}} - A\right) - \Phi\left(\frac{\mu_{s_j-1}}{\hat{\sigma}} - A\right) \tag{6}$$

with A given in Box I. If non-stationary (period-specific) transition matrices are desirable, the appropriate estimate of the $s_l \to s_j$

[1] $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal CDF and PDF, respectively.

[2] Notice that robust standard errors should be computed, taking into account the clustering of observations coming from the same firm, country, product denoted by i.

[3] The reason for both becomes obvious later when the calculation of predicted transition probabilities and marginal effects is presented.

[4] Below we discuss the difference between Predicted Probabilities at the Average (PPA) and Average Predicted Probabilities (APP). For the point made here, there is no need for such distinction.