A Note on Markov Chain Models to Study the Dynamics of Default Probabilities*

Winfried Pohlmeier†

University of Konstanz, COFE, RCEA

February 5, 2021

Abstract

This note is a rather incomplete collection of approaches on how the dynamics of default probabilities of firms can be modeled by means of dynamic discrete choice models and aggregate time series approaches.

Keywords: Markov chain models, transition probabilities, aggregate data, panel data.

*.

†Department of Economics, Universitätsstraße 1, D-78462 Konstanz, Germany. Phone: +49-7531-88-2660, fax: -4450, email: winfried.pohlmeier@uni-konstanz.de.

1 Introduction

The following note tries to give a first quick overview on which data are needed to estimate the dynamics of default probabilities. The goal is to provide the empirical researcher with information on the transition probabilities of a Markov chain, i.e. on the probability that a firm i in default class j at time $t-1$ switches to default class k in period t. At the end of the day such models may yield valuable information on how the default risks of individual firms or sectors have changed during the course of Corona crisis and how they can be used to assess future default risks. Basic references are Amemiya (1985), Chapter 11, and MacRae (1977). Warning: The note is extremely incomplete!

2 Markov Chain Models at the Micro Level

2.1 Basic Notation

Assume the default risk of a firm is defined by a finite number of discrete states called default classes defined over a grid of discrete time points. Let there be N firms, M discrete default states (risk classes) and T time periods. Define a sequence of binary random variables $y_j^i(t)$:

$$\begin{aligned} y_j^i(t) &= 1 \quad \text{if} \quad i\text{-th firm is in state } j \text{ at time } t \\ &= 0 \quad \text{otherwise,} \end{aligned} \tag{1}$$

with $i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T, \quad j = 1, 2, \ldots, M$. First-order Markov models are completely characterized if we specify the *transition probabilities* defined by

$$P_{jk}^i(t) = \Pr\left[i\text{-th firm is in state } k \text{ at time } t \text{ given that it was in state } j \text{ at time } t-1\right] \tag{2}$$

and the distribution of $y_j^i(0)$, the initial conditions. The following notation will be helpful:

$y^i(t)-\quad M$-vector the jth element of which is $y^i_j(t)$

$$n_j(t) = \sum_{i=1}^{N} y^i_j(t) \quad \text{number of firms in state } j \text{ at time } t$$

$$n_{jk}(t) = \sum_{i=1}^{N} y^i_j(t-1)y^i_k(t) \quad \text{number of firms in state } j \text{ at time } t-1 \text{ and state } k \text{ in } t$$

$$n^i_{jk} = \sum_{i=1}^{T} y^i_j(t-1)y^i_k(t) \quad \text{number of times firm } i \text{ was in state } j \text{ in } t-1 \text{ and state } k \text{ in } t$$

$$n_{jk} = \sum_{i=1}^{T} n_{jk}(t) = \sum_{i=1}^{N} n^i_{jk} \quad \text{number of firms in state } j \text{ in } t-1 \text{ and state } k \text{ in } t$$

$P^i(t) = \{P^i_{jk}(t)\},\quad$ Markov Matrix, an $M \times M$ matrix of transition probabilities

$p^i_j(t) = \Pr\left[i\text{th firm is in state } j \text{ at time } t\right]$

$p^i(t) =M$-vector the jth element of which is $p^i_j(t)$

Moreover the following special cases are often considered in literature:

- $P^i_{jk}(t) = P^i_{jk}$ is called a *stationary Markov chain* .
- $P^i_{jk}(t) = P_{jk}(t)$ is called a *homogeneous Markov chain* .

For firm-level data there is no reason to deviate from the assumption of heterogeneous and non-stationary Markov chains.

2.2 Estimation of Micro-data based Models

The likelihood function of the first-order Markov model takes the form:

$$L(\theta) = \prod_t \prod_i \prod_k \prod_j P^i_{jk}(t)^{y^i_j(t-1)y^t_k(t)} \cdot \prod_i \prod_j p^i_j(0)^{y^t_j(0)} \tag{3}$$

The parameter $P^i_{jk}(t)$ and $p^i_j(0)$ cannot estimated consistently as there much more unknown parameters than observations. Typically theses probabilities are expressed as a function a finite dimensional unknown parameter vector θ and a set of observable regressors, i.e. $P^i_{jk}(t) = F\left(x'_{it}\theta_{jk}\right)$.

This approach needs typical panel data of firms with their time series of the default categories and (most desirable) covariates at the firm level. It is a rather standard dynamic discrete choice approach which can be enriched by tools from machine learning e.g ridging (ℓ_2-norm penalization) and lassoing (ℓ_1–norm penalization), which can account for a whopping number of predictors.

2.3 Helpful Representations of the First Order Markov Chain

$$\mathrm{E}\left[y^i(t)\middle|\, y^i(t-1), y^i(t-2), ...\right] = P^i(t)' y^i(t-1) \tag{4}$$

$$y^i(t) = P^i(t)' y^i(t-1) + u^i(t) \tag{5}$$

Equation (5) defines a system of M equations which are linearly dependent as they sum up to 1. Therefore the whole system is fully represented by $M-1$ equations as

$$\bar{y}^i(t) = \bar{P}^i(t)' y^i(t-1) + \bar{u}^i(t) \tag{6}$$

Taking the expectation on both sides of (5) yields the familiar relationship between the state probabilities and the Markov transition matrix:

$$p^i(t) = P^i(t)' p^i(t-1) \tag{7}$$

With the simple but elegant Markov arithmetic multi-step transition probabilities can be computed, e.g. $p^i(t+1) = P^i(t+1)' P^i(t)' p^i(t-1)$, where $P^i(t+1)' P^i(t)'$ is the transition probability from t to $t+2$.

3 Time Varying Markov Models with Aggregate Data

The following set-up rests on Lee et al. (1970) and MacRae (1977), who show how the transition probabilities of a Markov model can be estimated using aggregate data. The basis are time series data for the the state probabilities.

$$p_j(t) = \sum_{i=1}^{M} P_{ij}(t) p_i(t-1) \qquad (j = 1, 2, ..., M) \tag{8}$$

For the vector of state probabilities we get:

$$p(t) = P'(t)p(t-1) \tag{9}$$

which is given (estimated by) the share f firms in each state at time t.

$$\sum_{j=1}^{M} p_j(t) = 1 \tag{10}$$

$$\sum_{j=1}^{M} P_{ij}(t) = 1 \qquad (i = 1, 2, ..., M) \tag{11}$$

Explanatory variables at the aggregate level can be introduced by expressing $P_{ij}(t)$ as a function of macro variables z_{t-1}:

$$P_{ij}(t) = f_{ij}(z_{t-1}, \beta_{ij}) \qquad (i, j = 1, 2, ..., M) \tag{12}$$

Typically we also use a multinomial probabilities for $f(\cdot)$.

$$\ln(P_{ij}(t)/P_{is}(t)) = F_{ij}(t) = z'_{t-1}\beta_{ij} \tag{13}$$

so that

$$P_{ij}(t) = \frac{\exp\{z'_{t-1}\beta_{ij}\}}{1 + \sum_{j=1}^{M-1} \exp\{z'_{t-1}\beta_{ij}\}} \qquad (j = 1, 2, ..., M-1) \tag{14}$$

The probability for the reaming state M results from the adding up constraint:

$$P_{iM}(t) = \frac{1}{1 + \sum_{j=1}^{M} \exp\{z'_{t-1}\beta_{ij}\}} \tag{15}$$

This leads to the likelihood function

$$L(\beta) = \prod_{t=1}^{T} \sum_{N(t)} \prod_{i=1}^{M} (n_i(t-1)!) \left(\frac{\prod_{j=1}^{M} P_{ij}(t)^{n_{ij}(t)}}{n_{ij}(t)!} \right) \tag{16}$$

Maximum likelihood estimation can be replaced by several alternative estimation approaches such as GMM or empirical likelihood approaches e.g. based on the Cressie-Read power divergence criterion, see Miller and Judge (2015). In any case longer time series are needed to estimate β as there is often little variation in the state probabilities and/or the covariates z_t. Therefore, I am very skeptical concerning a study of the impact of the Corona pandemic on the dynamics of the default probabilities based on quarterly sector data.

3.1 Applications

There is very little empirical work on the estimation of Markov chains based on aggregate time series data. The list of applications include Cate (2014) for the ecological inference model[1], Jones (2005) with an application to credit risk and Kelton (1984) for a the estimation of interregional allocations of firms.

References

Amemiya, T. (1985): *Advanced Econometrics*, Oxford: Basil Blackwell.

Cate, A. T. (2014): "Maximum Likelihood Estimation of the Markov Chain Model with Macro Data and the Ecological Inference Model," *SSRN Electronic Journal*.

Jones, M. T. (2005): "Estimating Markov Transition Matrices Using Proportions Data: An Application to Credit Risk," *IMF Working Paper*, WP/05/219.

Kelton, C. M. L. (1984): "Nonstationary Markov Modeling: An Application to Wage-Influenced Industrial Relocation," *International Regional Science Review*, 9, 75–90.

King, G., O. Rosen, M. Tanner, G. King, O. Rosen, and M. A. Tanner (2004): *Ecological Inference: New Methodological Strategies*, New York: Cambridge University Press.

Lee, T. C., G. G. Judge, and A. Zellner (1970): *Estimating the parameters of the Markov probability model from aggregate time series data*, Contrib. Eco. Anal., Amsterdam: North-Holland.

MacRae, E. C. (1977): "Estimation of Time-Varying Markov Processes with Aggregate Data," *Econometrica*, 45, 183–198.

[1]The term ecological inference model goes back to King et al. (2004) . It is somewhat misleading and has nothing to do with research on ecological issues. It is about reconstructing individual behavior from aggregate data.

Miller, D. J. and G. Judge (2015): "Information Recovery in a Dynamic Statistical Markov Model," *Econometrics*, 3.